

# Demands, designs and decisions about evaluation

On the evaluation of postgraduate programmes for school  
leadership development in Norway and England

by

Christopher Wales

A dissertation submitted to BI Norwegian Business School  
for the degree of PhD

PhD specialisation: Leadership and Organisation

Series of Dissertations 6/2011

**BI Norwegian Business School**

Christopher Wales

Demands, designs and decisions about evaluation: On the evaluation of postgraduate programmes for school leadership development in Norway and England

© Christopher Wales  
2011

Series of Dissertations 6/2011

ISBN: 978-82-8247-028-5  
ISSN: 1502-2099

BI Norwegian Business School  
N-0442 Oslo  
Phone: +47 4641 0000  
[www.bi.no](http://www.bi.no)

Printing: Nordberg Trykk

The dissertation may be ordered from our website [www.bi.no](http://www.bi.no)  
(Research – Research Publications)

## **Abstract**

The purpose of this study is to investigate decision making processes related to evaluation. The study attempts to combine research from decision making theory and evaluation theory to further explore the evaluation of postgraduate programmes for school leadership framed within a higher education setting. Focus is placed on investigating how providers of postgraduate programmes for school leadership respond in their decision making about evaluation to what is observed to be an increasingly complex and rich web of demands and pressures imposed upon them to assess programme quality and impact. This study addresses problematic areas for evaluative decision making in the phases leading up to implementation, investigating how the subunits under study respond to the demands placed upon them, what designs are considered and how the decision process functions and develops within such cyclical events.

The context for this study is four subunits operating within higher education institutions (HEIs) offering postgraduate programmes for school leadership development. Two subunits are drawn from England and two from Norway. The context has been characterised by an increasingly more complex policy environment linked to the perception that improved leadership at schools will make a contribution to improved pupil outcomes. The English context is characterised by increasingly centralised policy framework for programme delivery, whilst in Norway a more decentralised framework has developed. Additionally, as part of HEIs, the subunits are also subject to multiple pressures related to demands for evaluation and quality assurance emanating from the Bologna Process. Members of the subunits under study are experienced evaluators, who teach about evaluation on their programmes. As such this study might be additionally thought to be about evaluators evaluating evaluation.

The analytical framework is based on application of decision making models, operating as alternate templates, which are grounded in decision making literature. These are thought to offer alternative perspectives to the evaluation processes under study. This framework is applied to a process investigation based on documentary analysis and semi-structured qualitative interviews. The data is analysed through the application of computer assisted qualitative data analysis software, in this case NVivo 7. Themes emerging from the data and drawn from theory are also thought to extend the theoretical perspectives. By applying such a framework, it is hoped that this study will contribute further to the increasing interest in decision making research within the field of evaluation.

Evaluation was seen to become increasingly more institutionalised and bureaucratised within these educational institutions, appearing to be characterised by more assessment activity but less evaluative in nature. Subunits as groups operating within HEIs were observed to become more loosely coupled and even decoupled from the wider organisation with regard to evaluation, with

their members operating individually and independently of the wider organisational frameworks, where focus remained entrenched at the micro-level, adapting to satisfy organisational standards as they are introduced. These processes were characterised by the term dismissive submission. In one subunit however the members operated collectively, taking decisions about evaluation collegially, attempting to recouple with the wider organisation by presenting alternative models to influence, change and improve frameworks for evaluation. These processes were characterised by the term collegial construction. As a result of such processes, the latter subunit was able to engage at the initial stages of evaluation rather than merely implementing designs framed centrally.

The findings present a framework for further investigating and understanding evaluative processes within organisations, and particularly how groups can influence practice despite increased prescription and standardisation.

## **Acknowledgements**

I would like to take the opportunity to thank those who have contributed to the process of writing this doctoral thesis. First and foremost my heartfelt thanks to my supervisor Professor Anne Welle-Strand who has advised, challenged, encouraged, supported and pushed me on when the going got tough. This was necessary to see the project through and I would not have completed it without your help. Creating an environment within the Centre for Educational Management and Research (CEM) provided the catalyst to undertake this work. I am also deeply appreciative of the help and inspiration of Professor Arild Tjeldvoll. This journey started a long time ago on a master's programme when you encouraged me to direct my interest towards studies in educational leadership. My appreciation and thanks also go to the committee members at the pre-doctoral defence, Professor Tor J. Larsen and Professor Tony Bush, for constructive criticism and direction that were important for the progression of this project.

I am also grateful to so many at BI Norwegian Business School for all your help along the way, and especially to the staff of the Department for Leadership and Organisational Behaviour. I would like to offer thanks to current and former Department Chairs, Professors Astrid Richardson and Tor Larsen. Particular thanks also to Professors Svein Andersen and Bård Kuvaas for input and feedback through the doctoral seminars. I would also like to express my thanks to the administrative staff at the Department, past and present, particularly Tonje Horn Omland and Ann-Christin Johnsgård. I also owe my gratitude to the doctoral students at the department, past and current. There are so many now so I can't name you all here but I particularly want to thank Anders Dysvik for stimulating discussions at critical junctures.

To the others members of the "team" at CEM an extra special thanks: Taran Thune, Jan Merok Paulsen and Dijana Tiplic. Thanks for all your advice, encouraging remarks and helpful critical comments that were all necessary along the way. In completing the race ahead of me each of you, in turn, showed me it was possible, and then gave the belief, helping me too to cross the line, cheering on to the finish!

Special thanks are also due to the staff of BI library services. I have never before experienced such dedication, efficiency, professionalism and helpfulness in a library staff. My thanks also to the Doctoral Administration at BI and to Torill Eide for help with formatting the thesis.

My sincere thanks and appreciation go to all the respondents who have contributed to this project. Thanks for giving up your valuable time and for offering intriguing insights into the processes under study.

I would also like to thank my colleagues at Gimlekollen School of Journalism and Communication for their encouragement during the final phases of this project.

Thank you to my family and friends for all your support, whilst tolerating my frequent absences. I want to particularly thank my mother for all your support over the years. Thanks too to my "in-laws" in offering your help and

encouragement. To my two wonderful children Owen and Leah, great challenges that you both had to face along the way and the courage you have dealt with them have been a true source of inspiration for me, never a distraction. To my dear wife Vibecke, I have lost count of the number of times you have sacrificed your time for mine and shown so much understanding, patience, encouragement and support along the way. I simply say thank you because I could never repay any debt nor regain the time.

April 2011  
Christopher Wales

## Table of contents

Abstract	i
Acknowledgements	iii
Table of contents	v
List of figures	ix
List of tables	x
Abbreviations	xi
1. Introduction	1
1.1 Rationale and overview	1
1.2 The field of investigation	3
1.3 The theoretical framework	6
1.4 Analytical framework for the study	8
1.5 Research questions	12
1.6 Methodology	13
1.7 Contribution	15
1.8 The structure of the thesis	15
2. School leadership development programmes in Norway and England	16
2.1 School leadership training and development across and within the OECD	16
2.2 School leadership training and development in England	21
2.3 School leadership training and development in Norway	33
2.4 Linkage between leadership and improved pupil outcomes	40
2.5 Linking programmes, practice and outcomes: considering impact	44
2.6 Discussion	52
3. Evaluation purposes, processes and practices	56
3.1 The purpose of evaluation	56
3.2 Theoretical basis underlying definition of evaluation	58
3.3 Evaluation designs and models	59
3.4 The important context of evaluation traditions	63
3.5 Evaluation in an age of accountability	65
3.5.1 Evaluation under NPM	70
3.6 Evaluation utilization	75
3.6.1 The ideology of evaluation utilization	75
3.6.2 Types of evaluation use	76
3.6.3 Factors thought to influence or affect evaluation use	78
3.7 Summary of perspectives on evaluation design, implementation and utilization	80
4. Quality assurance and programme evaluation in higher education: Norway and England	84
4.1 Introducing quality assurance to higher education (HE)	84
4.1.1 Quality assurance and evaluation in Norway	90
4.1.2 Quality assurance and evaluation in English higher education	93
4.2 Summarising the policy shift in evaluation and assessment in higher education institutions	96
4.3 Developments in the quality culture	100

4.5 An exemplifying case: evaluating impact of the Quality Reform in Norway	109
4.6 Summary	110
5. Unravelling evaluation processes: focusing on decisions <i>about</i> rather than decisions <i>from</i>	112
5.1 Making decisions about evaluation	113
5.2 The evaluation process: “elements, actors and rationales”	120
5.3 Organisational decision making theory applied to evaluation	125
5.4 Decisions and evaluation in higher education	128
5.5 Models of decision making and decision processes in organisations	132
5.5.1 A rational approach- “action as choice”	133
5.5.2 Organizational behaviour – “action as output”	137
5.5.3 Political model (bargaining) – “action as political resultant”	143
5.5.4 An institutional approach - action as ritual decisions with constitutive effects	146
5.5.4.1 Sensemaking – links to Institutional theory?	158
5.6 Combining the models	161
5.6.1 The basis of decisions and decision processes	165
5.7 Decision makers	169
5.7.1 Occupational communities- link to professional judgement	169
5.8 Discussion	174
5.8.1 Proposed implications of the models	174
5.8.2 Initial implications of the rational actor model for this study	175
5.8.3 Initial implications of the organizational behaviour model for this study	175
5.8.4 Initial implications of the political model (bargaining) model for this study	176
5.8.5 Initial implications of the institutional model for this study	177
5.9 Investigating decisions about evaluation	178
6. Knowledge claims and methodology	182
6.1 Purpose of the study and strategy of research	182
6.2 A pragmatic approach to critical realism	183
6.3 Alternate templates strategy	186
6.4 Methods	188
6.4.1 Sampling	188
6.4.2 Levels and unit of analysis	191
6.4.3 Secondary data collection and contextualisation	192
6.4.4 Interviews	193
6.5 Analysis: reflection, coding and revision	197
6.5.1 Coding	197
6.5.2 Template analysis: basis and procedures used	198
6.6 Quality, delimitations and limitations	202
6.7 Writing up	205
7. Pressures and demands for evaluation?	206
7.1 Perceived pressure from the task environment	207



7.1.1 Perception of pressures and demands from policy makers and agencies _____	209
7.1.2 Perceived commissioner pressure _____	225
7.1.3 Perceived participant expectations and demands _____	229
7.2 Perceived institutional pressure _____	233
7.3 Perceived within-unit demands _____	242
7.4 Summary _____	245
8. Evaluation designs _____	248
8.1 Underlying frameworks _____	249
8.1.1 Attitudes to evaluation _____	249
8.1.2 Reflections over influence from subject field / profession _____	251
8.1.3 Current evaluation designs _____	252
8.2 Structures and approaches _____	254
8.3 Purpose _____	260
8.4 Focus _____	265
8.4.1 Participant perception and deliverer reflection _____	266
8.4.2 Commissioner framework _____	268
8.5 Current design limitations and debates _____	269
8.5.1 Structural problems _____	269
8.5.2 Participant characteristics _____	271
8.5.3 Interpretation problems _____	273
8.6 Designs for programme effects _____	275
8.6.1 Ascertaining the effects of study _____	275
8.6.2 Problems with designs for the study of effects and impact _____	280
8.7 Summary _____	284
9. Evaluation decision making _____	287
9.1 Role issues _____	288
9.1.1 The collective role of academic staff _____	289
9.1.2 Role of administrative staff _____	293
9.1.3 Interaction with programme participants _____	293
9.1.4 Relationship to the organisation _____	295
9.1.5 Role of commissioner and other external bodies _____	299
9.2 Decision processes _____	306
9.2.1 Perceived internal proximity within the subunit _____	307
9.2.1.1 Reflections from subunit NOR1 - <i>collegial construction</i> _____	308
9.2.1.2 Reflections from the other subunits: <i>dismissive submission</i> _____	317
9.3 Decision responses to demands on the subunit _____	327
9.3.1 Responses to internal demands _____	327
9.3.2 Responses related to participant demands and participation _____	333
9.3.3 Decisions about ascertaining programme effects and impact _____	338
9.3.4 Decision responses to commissioners _____	344
9.3.4.1 Academic anchoring _____	346
9.4 Summary _____	348
10. Discussion of decision processes thought to influence the design of evaluations _____	350
10.1 Demands _____	351

10.1.1 Balancing tensions	353
10.1.2 The extra level of commissioner and the influence of government bodies	354
10.1.3 Internal pressures and influences	358
10.1.4 Evaluative information as signal and symbol	359
10.2 Definitions and designs	363
10.2.1 Attitudes to evaluation	364
10.2.2 The QA system: formality and informality; accountability versus improvement	365
10.2.3 Focus and limitations	366
10.2.4 Cause and effect models: debates and problems	367
10.2.5 Summary: Defining and designing evaluations	371
10.3 Decisions and decision makers	372
10.3.1 Decision roles and structures	372
10.3.2 Elements of the evaluation decision making process	374
10.4 Decision responses to demands: <i>issues of organisational coupling</i>	378
10.4.1 Impact of subunit behaviour on the organisation	379
10.4.2 Impact of the evaluation frameworks on the groups: constitutive effects	392
10.5 Discussion; framing evaluation within the wider organisation	396
11. Concluding remarks, limitations and implications	401
11.1 Discussion	401
11.2 Summary of findings	403
11.3 Limitations of the study	416
11.4 Further implications for theory, policy and practice	416
11.5 Further considerations for future research	419
Appendix	420
References	430

## List of figures

Figure 1: The conceptual framework of the study .....	8
Figure 2: The analytical framework for understanding decisions about evaluation design .....	10
Figure 3 Evaluation framework for leadership programmes. Source: Leithwood and Levin, 2005: 36 .....	45
Figure 4 The problematic area of utilisation .....	82
Figure 5 The process under study .....	82
Figure 6 Balancing evaluation on the continuum of accountability and improvement .....	88
Figure 7 Three interlinked levels of decision-making in the University (Hardy et al., 1983: 414) .....	131
Figure 8 Thompson and Tuden's framework of decision strategies .....	166
Figure 9: Evaluative goal agreement across organisational levels .....	178
Figure 10: Evaluative goal agreement within context .....	179
Figure 11: Balancing contextual tensions .....	353
Figure 12 The framework for reflection over participant feedback .....	356
Figure 13 The framework for reflection over participant feedback when involving commissioners .....	357
Figure 14: Demands at subunit and wider organisational level .....	359
Figure 15: The elements of evaluation decision making under investigation (after Dahler-Larsen, 2004) .....	375
Figure 16: The perception of how evaluation data is used within the organisation .....	377
Figure 17: Collegial influence on enablers of sensegiving .....	384
Figure 18: Collegial construction and dismissive submission compared .....	387
Figure 19: The subunits within elements of the decision process .....	388
Figure 20: Revision of Dahler-Larsen's fourth quadrant within a continuum ..	389
Figure 21: Subunits placed within the template analysis .....	399
Figure 22: Hardy <i>et al.</i> 's framework reappraised .....	400

## List of tables

Table 1 Criteria for assessment of the value and impact of leadership development programmes. Adapted from Bush, 2008b .....	49
Table 2 Categories of problem related to evaluation decision making (after Stufflebeam et al. 1971).....	119
Table 3 Elements of the evaluation process (after Dahler-Larsen, 2004a: 41-45)	121
Table 4 Phases of interactive decision making (Hardy et al., 1983: 417-418) .	130
Table 5 Scott's three pillars of Institutions (Scott, 2001: 52).....	147
Table 6 Occupational community reactions to evaluation demands (after Dahler-Larsen, 1998).....	171
Table 7: Occupational community reactions to evaluation demands related to responses.....	173
Table 8 Reapplication of Stufflebeam et al.'s categories of evaluation problems .....	180
Table 9: Framework and timetable of data collection and sources .....	190
Table 10: Overview of respondents .....	191
Table 11 Template revision (after King, 2004: 261ff).....	201
Table 12 Tensions related to evaluation designs .....	286
Table 13: Summary of demands upon subunits .....	362
Table 14 Summary of responses to understanding the situational context .....	371
Table 15: Triggers and enablers of sensegiving in NOR1 (after Maitlis & Lawrence, 2007).....	383
Table 16: Collegial construction and dismissive submission compared.....	386
Table 17: Examples of constitutive effects of evaluation (after Dahler-Larsen and Krogstrup, 2000).....	393
Table 18: Placement within evaluative framework.....	398
Table 19: The problem areas of evaluation (reapplying Stufflebeam et al., 1971).....	403
Table 20: Summary of main research areas .....	415
Table 21 Types of Evaluation use (Weiss, 1998b; Weiss, et al., 2005).....	420
Table 22 Meta-analyses of the factors affecting evaluation utilization (after Hofstetter and Alkin, 2003) .....	420
Table 23 Interview guide: main questions and follow ups: .....	422
Table 24 Initial Coding template .....	425
Table 25 Final coding template .....	427

## Abbreviations

BELMAS	British Educational Leadership Management and Administration Society	MA	Master of Arts
CAQDAS	computer assisted qualitative data analysis software	MBA	Master of Business Administration
CEO	Chief Executive Officer	MBOR	Management by objectives and results
CERI	Centre for Educational Research and Innovation	MOLIS	Miljø og Ledelse i Skolen (Environment and Leadership in Schools)
CPD	Continuing Professional Development	M.Ed	Master of Education
CR	Critical Realism	M.Sc.	Master of Science
DfEE	Department for Education and Employment	NCSL	National College for School Leadership
DfES	Department for Education and Skills	NDC	The National Development Centre for School Management Training
Ed.D.	Doctor of Education	NDPB	Non- Departmental Public Body
EHEA	European Higher Education Area	NFR	Norwegian Research Council
ENQA	European Association for Quality Assurance in Higher Education	NG	Network Governance
EIPP	Evidence Informed Policy and Practice	NOKUT	Nasjonalt organ for kvalitet I utdanningen (The Norwegian Agency for Quality Assurance in Education)
ESG	European Quality Assurance Standards and Guidelines	NPM	New Public Management
EUA	European Universities Association	NPQH	National Professional Qualification for Headship
HEAD	School Management Training for Quality and Accountability	NSD	Norsk Samfunnsvitenskapelig Datatjeneste

HEADLAMP	The Headteachers' Leadership and Management Programme	OECD	(Norwegian Social Science Data Services) Organisation for Economic Co-operation and Development
HE	Higher education	OfSTED	The Office for Standards in Education, Children's Services and Skills
HEFCE	Higher Education Funding Council for England	PIU	Performance and Innovation Unit
HEI	Higher education institution	QA	Quality Assurance
IPA	Interpretive Phenomenological Analysis	QAA	Quality Assurance Agency
HRM	Human Resource Management	SCRELM	Standing Conference for Research on Educational Leadership and Management
ISL	Improving School Leadership	RAE	Research Assessment Exercise
IT	Institutional theory	SOFF	Sentralorganet for fleksibel læring i høgre utdanning (Norway Opening Universities)
KD	Kunnskapsdepartementet (Ministry of Education and Research)	SOPs	Standing operating procedures
KS	Kommunenes Sentralforbund (The Norwegian Association of Local and Regional Authorities)	SMTF	School Management Task Force
KUFD	Kirke- utdannings-og forskningsdepartementet (Ministry of Church Affairs, Education and Research)	TA	Template Analysis
LEA	Local education authority	TDA	Training and Development Agency
LEVIS	Ledelse i Videregående skolen (Leadership in Upper Secondary School)	TTA	Teacher Training Agency

LIS	Ledelse i skolen (Leadership in School)	UCET	Training and Development Agency
LPSH	Leadership Programme for Serving Headteachers	UFD	Utdannings-og forskningsdepartementet (Ministry of Education and Research)
LUIS	Ledelsesutvikling i skolen (Leadership development in school)	UFG	Universities Funding Council
		UK	United Kingdom
		USA	United States of America





# 1. Introduction

## 1.1 Rationale and overview

In recent years there has been greater interest in the training and development of school leaders with the aim to improve the quality of schooling and output of educational activity. Internationally there have been varying policy solutions of how such enterprises should best take shape, with governments funding different measures, while at the same time becoming increasingly concerned to see return on their investment, understanding the impact of their policy provision or at least ascertain evidence of value for money. School leadership training, development and other associated educational initiatives are therefore, in concert with other public policy measures, subject to increasingly more stringent assessment and evaluation. This focus on evaluation is considered to be part of an evaluation wave, within a growing “evaluation culture” (Dahler-Larsen, 2006b). Whilst demands for evaluation have increased, less is known about how these demands are operationalized, particularly at micro-level. The programmes of interest in this study are implemented within higher education institutions (HEIs), which are also subject to greater focus upon quality assurance and evaluation. This study aims to investigate further how decisions are made about evaluation.

The subject of evaluation is not, however, uncomplicated. In recent years there has been greater focus on how evaluation findings can contribute as evidence for use in decision-making (Weiss, 1979). This has increased under policy making based on implementing New Public Management (Norris & Kushner, 2007) and additionally under Modernisation. At the same time research suggests that more frequently evaluations are not utilized (Henry & Mark, 2003; Hofstetter & Alkin, 2003; Russ-Eft, Atwood, & Egberman, 2002), that the quality of evaluations vary (Palumbo & Nachamias, 1983; Schwartz, & Mayne, 2005) and the type of use of findings may only be symbolic or aimed at legitimating a programme and its theory (Alkin & Taut, 2003). It has further been suggested that even where high quality evaluations are implemented there is no guarantee that they will result in the findings being utilised (Dahler-Larsen, 1998; Abma, & Noordegraaf, 2003). There are also conceptual difficulties with ascertaining evidence of impact through evaluation (Alliger & Janak, 1989; Holten III, 1996). The field of school leadership has also received some interest regarding these deliberations and investigations (Barker, 2007; Bush, 2008b; Guskey, 2000; Leithwood & Levin, 2005).

The major foci of research on improving evaluation models have on one side addressed the technical quality of implementation (Weiss, 1972; 1982, 1998b) and on the other on increasing participation and relevance to stakeholders (Cousins & Earl, 1995; Patton, 2003). Some evaluation research assumes that evaluations are designed through a rational form of decision-making; where evaluators interpret the demands placed on them, search for an optimum model

to investigate goals, inputs, process and outcomes, looking for causal relationships between them (Weiss, 1987). Other research has focused more closely upon the stakeholders and other interested parties involved in the process itself and how this affects outcomes (Greene, 1988). Whilst the attempt to improve the quality of evaluation models and their output is important, as is study of interested parties and impact of stakeholder voice, this study considers that deeper investigation is also required into the decision making processes that contribute to a particular choice of evaluation model, and how and why this choice is made by the evaluating group, and in what way it is a response to different pressures and demands. This is an area considered to require more study (Holton III & Naquin, 2005; Rogers & Hough, 1995), even though there have been attempts to develop frameworks for investigating these processes (Dahler-Larsen, 1998).

The purpose of this thesis is to explore decision making processes related to the demand for evaluation. The context is considered interesting for such research as the respondents are involved in evaluation on many levels, with wide experience as internal and external evaluators. Vedung (2003) notes there to be a special evaluation tradition within HEIs. In addition significant focus is placed upon evaluation within their programme content. The responses given in this study could therefore be described as evaluators evaluating evaluation. The investigation is built upon trying to understand the perceptions of subunit members of HEIs responsible for implementing evaluations of their postgraduate programmes in school leadership. While this study is not focused upon the implementation of the evaluation and how evaluation findings are utilized per se, it is concerned with how the perception of the purpose and intent to utilize the results of evaluation might impact upon responses to demands and pressures, whether internal or external to the organisations under study. It is felt that increased understanding of these processes will further contribute to research into the factors that influence utilization and understanding of how organisations respond to the demand to evaluate and be accountable. This study will therefore also involve investigation of programme providers' values and ideologies concerning evaluation, especially in relation to the goals and rationale of their programme(s). In order to investigate such processes decision making theory is combined with evaluation research. The area of focus is developed from Stufflebeam et al.'s (1971) problematic areas for evaluative decision making. The analytical tool is formed from 4 decision making models that draw mainly on the research of Allison (1971; & Zelikow 1999), Peterson (1976), Thompson (1967, 2003) related and extended by Dahler-Larsen (1998, 2001, 2006b) to evaluation theory, as well as the work of Hardy et al. (1983) into decision making within HEIs. These models are applied in combination functioning as alternate templates, described by Langley (1999) to be alternative theoretical interpretations of the same event, which are thought to provide a more detailed explanation of processes under investigation. These are employed in the empirical part of the study to analyse organisational behaviour. These

elements are outlined briefly below before more detailed discussion in the ensuing chapters.

## **1.2 The field of investigation**

Discussion with regard to how programmes for training and developing for school leadership are evaluated has become more clearly evident in recent years. Such programmes have become a more common part of national public policy reforms aimed at developing the quality of educational provision (Hallinger, 2003). Educational legislation and statutory guidance has become more focused upon the necessity for “high quality professional development” that should improve school leadership (Guskey, 2003). This was visible in policy documentation in England, for example ‘Every Child Matters’, (UK Treasury, 2003); and in Norway, ‘Culture for Learning’, (UFD, 2004b). Such development is often linked to the contested belief that pupil learning outcomes will improve as a result of a better leadership and management skills base (Bell, Bolam, & Cubillo, 2003; Bush, 2005c; Leithwood & Levin, 2005). Research into how such programmes are designed and developed was also studied as a part of a comparative research project, HEAD, which investigated training and development across five countries<sup>1</sup> which ran from 2004 to 2008. The findings of the research noted that England has developed a more formal structure of programmes, including the mandatory NPQH<sup>2</sup>, whilst Norway has focused on investing in Master degrees and locally organised programmes<sup>3</sup>, of which many are included in generic local authority management training structures. While it is still possible to study for a master degree in England on a programme related to educational leadership and management, many of the HEI departments have come under increasing pressure as a result of competition from nationally sanctioned programmes, under the responsibility of the National College for School Leadership (NCSL)<sup>4</sup>. This study was developed during the period of the HEAD project.

The emphasis in this current study is upon the evaluations designed within subunits in HEIs offering postgraduate programmes for school leadership. These programmes are often awarded as master degree programmes, although individual programme modules may also be offered to external commissioning bodies as training and development programmes. This study will consider these

---

<sup>1</sup> The HEAD Project 2004 – 2008 (School Management Training for Quality and Accountability) was a 4 years research project on school management training and development in Norway, in cooperation between Norwegian school of Management and the University of Oslo. It was an action research project on curriculum, organisation and achievements of school manager training programmes in Norway. Norwegian training and achievements were compared to international “good practice” in Finland, France, UK and USA. The HEAD project was funded by the Norwegian Research Council (NFR) within the FIFOS programme (Research on Innovation and Renewal of the Norwegian Public Sector).

<sup>2</sup> National Professional Qualification for Headship

<sup>3</sup> Which may often also be included as a part of a Master programmes.

<sup>4</sup> Now the National College for Leadership of Schools and Children’s Services.

programmes in relation to the changes toward a policy era of wider demands for results and new about effects, offering an evidence-based perspective. Research from the HEAD project indicated the increasing discussion over how effects of these programmes can be discovered through evaluation, particularly in Norway and England (Wales & Welle-Strand, 2005). This discussion was based on the academics' perception of evaluation theory and practice, noted to be shaped by important evaluation traditions with in the field of study, as well as more widely with regard to policy making traditions. In Norway the varied nature of programme evaluation was noted, where the purpose is often unclear and the audience uncertain. Norway has more generally been described as a 'latecomer' to the concept of systematic evaluation, (Baklien, 1993; Sverdrup, 2002) especially related to public expenditure (Ovrelid & Bastoe, 2002). This may reflect the small, 'egalitarian' nature of its society, where there is a general tradition for framing policy by consensus, participation, pragmatism and incrementalism; much of which may result from relative financial stability (Ibid.). However, within a system that has more traditionally focused on accountability and assessment, evaluation in the United Kingdom<sup>5</sup> has been characterised by even greater visibility in recent times, with a general shift in public policy focus from the evaluation of management of policy and resources to the management of outcomes (Gray & Jenkins, 2002). All of these macro level developments are thought to impact the micro level.

Despite the relatively different approaches to school leader training and development and traditions of evaluation, there is an experience of greater demands and pressures from both national and local mandators to evaluate more effectively and allow future policy decisions to be 'evidence based' (Anderson & Bennett, 2003; Simkins, 2005a) offering a degree of understanding of impact. There is a growing trend towards so-called evidence based decision making (Weiss, 2004) but the application of such evidence is noted to be the exception rather than the rule (Pfeffer & Sutton, 2006). With greater demands placed upon providers of publicly funded programmes to account for and evaluate their activity, there arise threats to the credibility of evaluation information (Schwartz & Mayne, 2005). Political and organizational pressure can lead to a-priori bias, whilst there is also the more *pervasive* threat of 'shoddy practice'(2005: 7). Organizations need, therefore, to attend to their "blind spots" and find and follow the "best data and logic" (Pfeffer & Sutton, 2006). For many this movement will suggest an over-emphasis upon experimental and positivist data collection. However, it is also recognized that there is weakness in information collected from programme evaluations, whatever the evaluation tradition from country to country. These wider trends are considered useful for the analysis of the evaluation models used to assess the programmes under investigation in this study.

---

<sup>5</sup> Which the authors apply to the English context.

It has been claimed that rather than ascertaining the impact of programmes on participants and changed behaviour at their schools, the majority of evaluation models discover little more than how satisfied participants have been with the courses they have attended (Guskey, 2000; Leithwood & Levin, 2005). This is also reflected in wider research of evaluation of training (Alliger & Janak, 1989; Holton III & Naquin, 2005; Kraiger, 2002). This background frames the field of investigation for research into the decision processes that guide the choice of model adopted for the evaluation of school leadership programmes. It has been considered unclear as to how findings from the evaluation of training and development can be utilized, as adjustments made to programmes resulting from such types of evaluation often appear small, incremental and self-reinforcing and are more likely based upon perceptions of learning rather than actual changes in performance (Holton III, 1996). Whatever processes are set in motion, they offer little to the aid the discovery of whether a programme is 'good or bad', and are said to require greater 'effort' (Goldstein & Ford, 2002). Most evaluations focus on trainee / participant reactions, saying little about learning or improved outputs / performance (Goldstein & Ford, 2002; Guskey, 2000, 2002). Criticism, in particular, of a widely used model developed by Kirkpatrick (1998), is that it is little more than taxonomy of outcomes, where the implicit causal relationships remain 'unoperationalised' (Alliger & Janak, 1989; Bates, 2004), and too many intervening variables that are ignored (Holton III, 1996). Perhaps Holton's strongest criticism is the model's reliance upon 'participant reaction' as a "primary outcome of training", supporting Alliger and Janak's point reflections that reactions are not linearly related to learning, but may moderate or mediate it.

Therefore, the information that organisations often claim to base decisions about impact on is considered to be flawed. The greatest problem appears to be the evaluation models applied to programmes, and the conceptualisation of what the organisation is attempting to achieve. Related questions have formed on-going dialogue between the wider research field and England's NCSL, raising questions of control over the process (Bush, 2005a, 2005b), assumptions and purpose behind the act (Simkins, 2005b), and types of model applied (Earley, 2005; Earley & Evans, 2004). The NCSL, at the time, in responding to these criticisms were seen to accept the challenges (Conner, 2005), noting that design is an important area of focus, particularly being concerned about the impact of available time and resources (Southworth, 2004). However, perhaps analysis should also be directed more widely, in particular towards the decisions that guide the choice of approach and development of evaluation model with utilization in mind.

While there is a great deal of new emphasis on training and development programmes, the role of the HEIs providing postgraduate programmes is still an interesting area of study. They have come under greater pressure; where the very basis of academic development appears to be challenged. In summary, the chosen area of investigation of school leadership development programme

providers is considered salient for three key reasons. Firstly, there is an increasing desire amongst mandators to use evaluation to discover what the impact of programmes has on pupil outcomes. This raises the issue of what focus on evaluation is meant to achieve and how it is believed that improving the quality and extent of evaluation will give greater information about the quality of the programmes. Secondly, questions are raised as to the normative models and values that underlie these programmes, particularly what impact this has on programme content and how it is evaluated in the light of them. It is therefore important to discuss the decision processes related to choice of evaluation model based on the subunits' ethos for their programme and for evaluation. Choice of model is thought to affect the utilization of the information, which is considered a major purpose of the evaluation process. It is proposed that while the same evaluation models might be applied across different contexts, the organisational values and decision-making processes underlying them might differ. Finally, the context is interesting due to the complexity of demands subunit members face with regard to evaluation, combined with the fact that programme content often focuses upon programme evaluation and the academic staffs are often experienced evaluators. Decisions about evaluation are not simplified by the context within which they take place. The subunits in this study are all situated within Higher Education Institutions. It is recognised that HEIs evaluate their programmes within multiple, overlapping frameworks and contexts, for example as part of their institutional quality assurance systems, also with regard to the 'professional' field within which they operate and in relation to any external programme mandators and funders. Although such frameworks are formally stated within the organisation, it is considered that they are not necessarily followed by linear implementation. These are aside from the various dynamics within the subunits. Therefore study of these processes will require investigation into the perceived impact of these factors, as well as attempting to uncover other influences and demands. This should enable greater insight into the processes under investigation.

### **1.3 The theoretical framework**

With regard to these assumptions, analysing the underlying decision-making processes, that have often been ignored, should offer a more in depth understanding of the evaluation design process (Holton III & Naquin, 2005). In this way, even if better measures of programme impact are developed, and greater understanding is achieved concerning utilisation, it is still considered important to investigate the attitudes and underlying values programme providers have to the purpose of the evaluation process.

Issues such as these were raised by Stufflebeam et al. (1971). The authors framed 5 questions or problematic areas related to evaluation decision making: *definitions of evaluation, decision making, values and criteria, administrative levels, and evaluation and research compared*. While these areas are considered important, the focus of this study is on part of the evaluation process; responses to demands for evaluation and evaluation design. Perception of the interaction

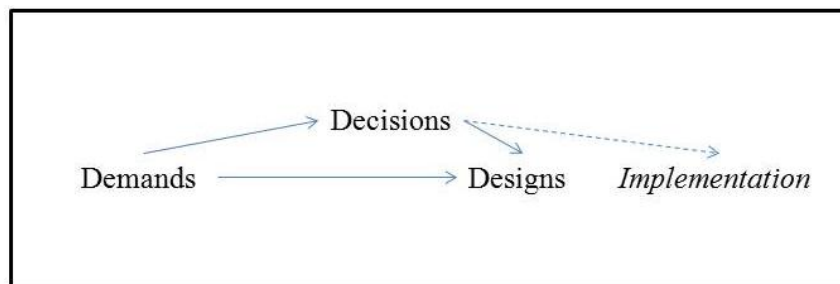
between evaluators and mandators / commissioners is under investigation more than focus upon how to meet demands, and thus upon decision making surrounding the adoption and implementation of evaluation models. The evaluation process is considered to be a complex interweaving pattern of events, which also challenges the proposition that the decision process is hierarchical and linear (Dahler-Larsen, 2004a; Dornbusch & Scott, 1975). Focus is therefore placed upon who makes the decisions about evaluation design within the implementing organisation. Stufflebeam et al.'s problematic areas have therefore been reapplied for this study. Thus questions are raised concerning the *demands* placed upon organisations and within organisations with regard to evaluation and accountability; programme providers *definitions* of evaluation; the *designs* in use, which may include those chosen to meet these demands; and the *decision making processes* that takes place, which will involve investigation of decisions made as well as consideration of *decision making roles* and those *decision makers* responsible for taking these decisions.

In order to investigate these concepts, focus is placed on *elements* of the evaluation decision making process (Dahler-Larsen, 2004a). The elements chosen are those leading up to the implementation of an evaluation, namely: *initiation, agenda, knowledge management and organisation, and design*. Additionally, it is also considered important to understand the influence of the context on these decisions as well as how respondents view the possibility to ascertain programme impact in relation to it (Stake, 1990).

Whilst this work does not investigate evaluation utilisation, perception of it is considered an important concept for decision makers. Evaluation findings are used in varying ways, from instrumental and conceptual use aimed at improving programme delivery, to symbolic and legitimative use focused on gaining support for programme survival (Greene & Walker, 2001; Hofstetter & Alkin, 2003; Weiss, 1998b). Despite recognition of these forms of use, it is generally suggested that research into the purpose, framing and implementing of evaluations has struggled to isolate the factors that influence the way results and findings are utilised, both internally and externally (Alkin & Taut, 2003; Caracelli, 2000; Cousins, 2003; Cousins & Leithwood, 1986; Johnson, 1998; Preskill et al.2003). Research has attempted to create an overall framework of factors that influence utilization, resulting in the higher order categories of decision / policy setting and quality of evaluation implementation (Cousins & Leithwood, 1986; Johnson, 1998; Leithwood & Levin, 2005). Interest continues to focus on the internal and environmental factors that shape utilization (Kirkhart, 2000), as well as the level to which evaluations are operationalised and investigate programme effects (Guskey, 2000). However, the evaluation field has generally continued to adopt a fairly rational view of the assessment and evaluation process, while appearing to fail from fully applying findings from organisational decision-making research (Holton III & Naquin, 2005; Palumbo 1987; Palumbo & Nachmias 1983; Rogers & Hough, 1995; Shapiro, 1984). While interest in the processes of utilisation is valid, so is greater interest

in the decision processes considered additionally necessary. If evaluation results are not used then understanding is required of how designs meet demands and how decisions are made and by whom. The parts of the organisation where these are made may appear to vary and the context is also considered to play an important role.

It is proposed therefore that such investigation of organisational members may help reveal the values and ideologies underlying the decision-making process concerning the design and how an evaluation should be implemented. In particular, understanding of the organizational decision-making function needs to be developed. It will therefore be important to attempt to illuminate the decision process in terms of describing what happens, and the perception of why programme groups think it happens like it does. What kinds of demands are placed upon them and how do they come up with a design for their programmes. This focus is outlined in the conceptual framework presented in figure 1 below. It should be noted that the dotted line leading implementation highlights that research is focused on decisions concerning implementation but not the implementation of the evaluation itself.



**Figure 1: The conceptual framework of the study**

#### **1.4 Analytical framework for the study**

An important focus of this study is the attempt to explore why and how evaluations develop, while investigating the designed models. In order to understand these processes, this study draws on decision-making theory. While it is necessary to outline the major developments in decision-making research, it must be specified that the focus is on decision-making in action rather than on prescriptive decision modelling or experiments. Consequentially, it is considered vital to illuminate evaluation research with naturalistic decision-making theory (NDM) which attempts to discover the underlying attitudes and ideologies of those evaluating programmes, an area which has been suggested to have been under investigated (Holton III & Naquin, 2005). NDM research is considered useful as it focuses on how decisions are made, drawing on the perceptions of those involved in decision-making processes.



At the same time it is recognised that such organisational decision-making processes are complex, often unnoticed and not open to reductive descriptions. Therefore a framework of decision-making models is applied, operating as alternate templates, which are grounded in decision making literature that can offer alternative perspectives to the process at hand. Such a framework is based on the application of multi-faceted models incorporating different strands of organisational decision making research, exemplified by Allison in his analysis of the handling of the Cuban Missile Crisis (1969, 1971; & Zelikow 1999). However, such an approach has also been adapted and further applied within various educational settings (Ellstrom, 1983; Hardy, 1990b, 1991; Hardy, Langley, Mintzberg, & Rose, 1983; Peterson, 1976; Sergiovanni, 1979)<sup>6</sup> based on the recognition that no one decision model will satisfactorily help analyse and explain all decision behaviour. As outlined earlier these models are supplemented by a template drawn from more recent developments in Institutional theory (Dahler-Larsen, 1998; W. R. Scott, 2003; Thompson, 1967, 2003), where the work of the former author has been applied generally to the field of evaluation.

Allison's three conceptual decision making models, are: "rational actor, organizational behavior, and Government politics". The rational model paints the broader picture of a decision, including the search for an optimal choice. The organisational behaviour model focuses on the organisational rules and routines that produce information, options and action. Allison's third model investigates individual action and how perceptions and preferences are combined to influence decisions (1999: 392ff). These models were further adapted by Peterson (1976), where in particular, the final model is further divided into "ideological bargaining, and pluralist bargaining". Allison recognised that a combination of these models or lenses should enable broader analysis of decision processes. As a result the third model is defined here as political bargaining. Research into decision processes since Allison's models were outlined allows for further development. A model based on developments in institutional theory allows for focus more on how environmental influence constitutively forms, develops and changes organisational identity. The models, then, rather than competing, combine to offer a more in depth understanding (Pfeffer, 1981b), helping to illuminate different ways that decisions are made. As a result one could anticipate, like Dahler-Larsen (1998), that political and institutional models will overlap but that the impact over a longer period might be different.

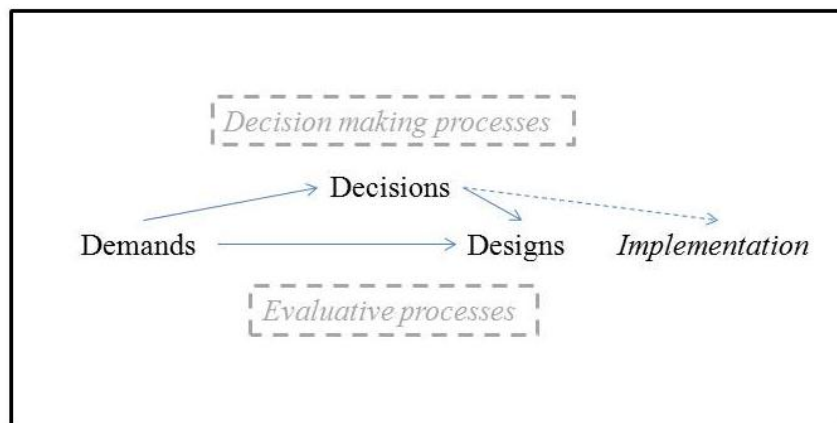
These models are used to analyse the subunit decision-making concerning the adoption and implementation of a particular evaluation model for a programme.

---

<sup>6</sup> Valovirta (2002) can also be interpreted of conceiving these decision processes in a similar way, questioning whether evaluations are found to be more 'academic', attempting to produce instrumental changes; bureaucratic, producing conceptual changes; or political, producing symbolic or legitimative utilization.

As has already been stated, focus on utilisation is delimited to how the intent to utilize information influences the type of evaluation model that is applied. This is considered to be part of the rationale or purpose for evaluation. This reiterates that the subsequent utilisation of the results of a particular evaluation is not under study in this project. This indicates that this concept it is important for its influence on and contribution to the purpose of the evaluation and how a particular model is chosen and implemented, but will not be fully investigated. Hence the figure is not a causal model, but rather a framework to discover how the decision making process influences the resulting models chosen to evaluate school leadership training.

This model, outlined in figure 2, is developed to further direct focus to the organisational decision making process and related to the elements of evaluation decisions. Application of such a model in this study is at the micro level, where focus is placed upon the decision making concerning evaluation within HEI subunits responsible for postgraduate programmes for school leadership development. While the study is focused upon the micro level, it is also important to recognise that such decisions are taken within wider institutional and environmental contexts. With this in mind it is important to recognise how such models have been observed within a macro-perspective, for example Thompson (1967, 2003) and Scott's (2001, 2003) combined organisational models and their influence on decision-making. Figure 2 below therefore takes into the account the influence of actual and perceived environment demands on the decision makers. Understanding of the impact of these demands will be drawn from self-reports of interviewees as well as secondary data, including circulars, letters and planning documents etc. It is proposed that these demands can both influence evaluation design directly and indirectly and may depend upon the extent of the demands placed upon and perceived by decision makers. This perception will be investigated through the study.



**Figure 2: The analytical framework for understanding decisions about evaluation design**

Figures 1 and 2 highlight the role of the decision makers in developing a design for an evaluation model that is considered to have an ultimate utilization purpose. It is reiterated that the final use of the evaluation findings is not under investigation per se; rather that investigation of the design process and model chosen is thought to give a clearer picture for analysis of how organisations respond to the demand to evaluate and plan to implement an evaluation. In the figure above the external and internal demands to evaluate are thought to influence the purpose of the evaluation (where external demands can notably influence internal demands as well). How these demands are interpreted within the subunit are part of the decision process to design a subsequent evaluation. This is however not merely a political model, as is reflected in the use of alternate and inter-related templates to give a richer picture of the process. It is also recognised that the demand to evaluate might also be associated with a demand to evaluate in a particular way or with particular emphasis on indicators in mind. This too is thought to affect the process of decision-making and will be further investigated in the empirical research. This framework will be adapted and developed as the study progresses in line with the methodology outlined briefly in sections 1.7 and in more detail in chapter 5.

The basic idea behind the decision framework, then, is drawn from Allison (1999) and Peterson (1976), Dahler-Larsen (1998) and Thompson (2003) where information is drawn from decision and organisation theory and compared to current developments within the evaluation field. This also involves discussing forms of evaluation use, which are thought to influence decisions about evaluation purpose and design. The implications of such a coupling require investigation of the decision process in an organisation. The early design phases will be important as much as a discussion over who will be responsible and how they will carry it out. Therefore, although it is considered correct to “distinguish between internal and external responsibility” for an evaluation at the arrangement, production and utilization phases (Vedung, 1997 in Dahler-Larsen, 2000), I also agree that this should be extended to look at the initiation of an evaluation and the broad influence across the phases towards use (Dahler-Larsen, 2000). This view appears even more necessary as evaluation is considered to take a much stronger role under NPM implemented throughout the public sector (Dahler-Larsen, 1998, 2005a, 2005b). This has seen evaluation developing from a typically one-off approach to a more institutionalised part of organisational routine (Hellstern, 1986 in Dahler-Larsen, 1998). This will require greater understanding of relationships and decision-making processes within organisations.

While organisational and decision-making theories are thought to illuminate the activity of evaluation there has been relatively little application of such research within the field of evaluation (Dahler-Larsen, 1998; Holton III & Naquin, 2005). In recent years, however, there has been a growing interest in research towards these areas, particularly within the Danish research programme (Albaek, 1996; Dahler-Larsen, 1998, 2004b, 2006b; Hansen, 2001, 2005a, 2005b; Krogstrup,

2006). While research elsewhere has continued to focus on the mechanics of improving evaluation and particularly utilisation, “[i]t is recognised that it is living social, political and organisational processes that form evaluations and decide whether evaluation results will be used”<sup>7</sup> (Dahler-Larsen, 1998: 25).

Many writers hypothesise the link from evaluation purpose and ultimate use, although one would particularly consider from rational approaches this link to be tightly coupled. But despite evaluation’s relative rational basis, as a process it has not always matched up to general expectations. Therefore, evaluation is recognised to be both the “child of rationalism and of rationalism’s limitations” (Vedung, 1991 in, Dahler-Larsen, 1998: 23). Such reflections would appear to require deeper understanding of the framework within which evaluation develops. While this is recognised to be multifaceted and affected by many factors, certain areas will provide useful starting points for further study. With this in mind it is considered helpful to focus on one aspect of the evaluation process.

### **1.5 Research questions**

This study will continue to explore decision-making activity surrounding the initial phases of evaluations of postgraduate programmes for school leadership development. Further delimited, this applies to understanding the subunit decision making process that results from the demand to evaluate and leading to the design of the subsequent evaluation implemented to meet this demand<sup>8</sup>. The overall focus of this study is related to the question:

- What influences the decision of how postgraduate programmes for school leadership are evaluated?

It is recognised that this is a complex area, influenced by many different factors and variables. Therefore attention is delimited to 3 important and interlinked sub-questions related to their decision making about evaluations:

- What pressures and demands do subunits face?
- What design frameworks are available to them?
- What decision processes take place within subunits about the choice of evaluation model?

As has been stated, in order to answer this overall question it will be required to develop an understanding of the basic purposes and rationale of evaluation, as well as intentions for future utilization of findings and existing knowledge of factors thought to influence this process. When observing the response to this,

---

<sup>7</sup> My translation from Danish

<sup>8</sup> Based on decision theory outlined above (Allison & Zelikow, 1999; Peterson, 1976; Pfeffer, 1981b; Thompson, 2003) (Dahler-Larsen, 2006a) (Guskey, 2000), (March & Heath, 1994)

an interpretive design combined with a pragmatic approach is required, investigating members' understanding of their decision making with regard to the process of evaluation. In doing this understanding of the relationship to programme goals, content and underlying rationale will support analysis. This is outlined in the section below.

## **1.6 Methodology**

In this research, analysis is made of the decision-making process through which designs for programme evaluation are adopted within subunits offering postgraduate programmes for school leadership. The unit of analysis is the organisational decision making process. The unit of observation will be the individual actors as members of subunits involved in the decision making process. With a lack of research in this area (Holton III, 2005) a pragmatic framework is constructed. At the same time it is recognised that theory has addressed this topic earlier even if has not been applied fully into the field. Therefore an a priori theoretical and analytical framework is applied in the study. In this case it is an alternate templates strategy building on process rather than variance research (Langley, 1999). An alternate templates strategy involves analysing a process from a number of different perspectives and can involve both deductive and inductive approaches (Ibid.). The alternate templates strategy, is based on analysing and interpreting the same events through "different but internally coherent sets of a priori theoretical premises", which are then assessed to the extent "to which each theoretical template contributes to a satisfactory explanation" (Langley, 1999: 698). Each alone will, however, be *insufficient* despite its relevance. The explanatory power and accuracy of the models chosen here are considered to be increased when they are applied in tandem. According to Langley, this application of the different, but complimentary models can lead to data interpretation that may reveal "contributions and gaps in each". Langley sees this strategy as similar to Allison's multi model approach and drawing also on Weick, Langley describes it as a process of sensemaking. This has particular relevance for this study. Weick (1976) recognised that critical analysis is required of language and communication that facilitates the decision process, and in order to do this different theoretical perspectives should be held. Such research opens for a combined strategy of deductive use of theory and inductive use of data (Langley, 1999), which appears similar to Ragin's *retroduction* (1994), as well as the interactive research process described by Maxwell (1996). Such a strategy also appears close to that applied by Peterson (1976), which led to his nuanced view of Allison's third model. Instead of attempting to generalise, the intention is to develop propositions and limited theory by "[r]efining partial paradigms, and specifying the classes of actions for which they are relevant", (Allison, 1971 in Langley, 1999: 699). This method also has a similar rationale to that of

research revealing toxic decisions by Maitlis and Ozcelik (2004)<sup>9</sup>. Langley notes, however, that this approach can provide difficulties when attempting to combine the models again. To combat this Thompson's (2003) combined approach is seen as a useful framework against which to analyse the findings of this research. This approach is outlined in more detail in subsequent chapters.

The strategy appears similar to the transformative method that employs a theoretical lens to analyse data (Creswell, 2003), but does not in the case adopt mixed method approaches. This will instead involve considering decision making processes through a permutation of combined models. In this study I combine the models developed through the research of Allison (1999), Peterson (1976), Dahler-Larsen (1998) and Hardy (1990b, 1991, Hardy et al. 1983), which will inform the alternate templates. There is mainly use of qualitative methods, in line with naturalistic decision-making research, which focuses on the actual activity of decision makers rather than "the decision event". Although such underlying values are often difficult for respondents to reconstruct (Beach & Connolly, 2005) interviewing is considered to be a useful strategy to investigate such processes (Lipshitz, Klein, Orasanu, & Salas, 2001). It is therefore felt appropriate to interview 'providers' involved in decision making about postgraduate programme evaluation, sampling purposively and theoretically. The intention of sampling purposively is to capture a semblance of heterogeneity in the population (Maxwell, 1996). Qualitative interview techniques are used in order to gain as rich a description as possible of the process and the actors' interpretation of it. In using qualitative interviewing as the main choice of methods, emphasis is placed upon discovering, recording and analysing participants attitudes to what guides the decision making process concerning evaluation models to be enacted within their organisation. The interviews are supported by analysis of secondary data, including documentary analysis from programme materials, national policy documents and other terms of reference, combined with literature review framing the problem within the fields of evaluation and naturalistic decision theory. The interview responses are then framed against the alternate templates (Langley, 1999) of organisational decision-making types for use within template analysis of the data (King, 1998, 2004). Template analysis is not a "single, clearly delineated method" but rather a thematic organisation and analysis of textual data, based a list of codes (the templates) that represent themes in the data (King, 2004: 256). The initial template of theoretically determined *nodes*<sup>10</sup> on which the analysis is based is outlined in the appendix. The final coding template is also included.

---

<sup>9</sup> The authors analysed aspects of three different organisational contexts that shaped the decision processes, noting that the decision process itself, and not just the issue under discussion, affected a certain outcome (Maitlis & Ozcelik, 2004: 377).

<sup>10</sup> A node refers to a place where an analytically determined category of data is assigned and stored in the place of a qualitative software programme (Richards, 2005).

## **1.7 Contribution**

While much has been written on the subject of evaluation with regard to utilization, thought to be the resultant of good evaluation design, and additionally the factors that are believed to influence it, there has been much less focus on the underlying decision-making process and mechanisms that inform it. With this in mind, the aim is to tie together research from the fields of programme evaluation, evaluation research and decision-making. The intention is also to develop existing frameworks and templates of decision-making, leading to greater understanding of the processes involved. This research is also an exploration of the impact of ideological positioning on the evaluation process. This is considered important with regard to how organisations can understand the processes that they develop and address any major issues arising from their values related to evaluation.

## **1.8 The structure of the thesis**

Chapter 2 focuses upon the context leading up to and including the data collection process relating the research to the field of study related to the school leadership development, discussing policy developments and responses. This is followed by discussion in chapter 3 of theoretical development from the field of evaluation, with particular regard to processes, purpose and utilization focus, and their relevance to this study. Chapter 4 returns to the context looking at the development of quality assurance and programme evaluation in higher education in Norway and England. This chapter is followed by an outline of theory regarding decision-making processes, as well as links made to evaluation. The chapter draws together chapter 2-5 to form the analytical framework. Chapter 6 deals with methodology. Chapters 7-9 present the empirical interview data. Chapter 10 presents an analysis of the data. Chapter 11 offers a summary and concluding remarks.

## **2. School leadership development programmes in Norway and England**

The purpose of this chapter is to present the context for this study, namely the policy and practice of school leadership training and development in England and Wales leading up to 2008. In attempting to understand the context for these initiatives, focus is initially placed on the Improving School Leadership project, undertaken by the OECD with contributions from both England and Norway. This project was a pertinent issue during the period of data collection. Responses from the country background reports prepared for this project provide a framework for discussing further the history and context of the programmes under study in both countries. As will be seen, policy making in England towards leadership of schools has become more focused, particularly since the formation of the National College for School Leadership (NCSL) in 2000. Aligned to these developments is the increasing perception of linkage between leadership and improved pupil outcomes. One of the underlying arguments for setting up the National College was the perceived linkage between leadership and improved pupil outcomes. Accepting or rejecting the strength of this linkage has great implications for approaching the assessment and evaluation of programmes, as well as consideration about the search for indicators of what training and development functions best. As a result greater space is given in this chapter to discuss these issues and their wider impact. The focus of this study is upon decision making related to evaluation of postgraduate programmes for school leadership development within Higher Education Institutions. It is considered that outlining the above mentioned issues will help provide a framework for data collection with the programme groups.

### **2.1 School leadership training and development across and within the OECD**

Within the area of education policy the OECD is considered to be an actor with great political “influence”<sup>11</sup> (Møller, 2006b: 39). In addition to the focus upon the economic benefits that improved educational results are thought to bring, the OECD, with its link to academic research, has in recent years turned attention to the role of leadership in developing more effective schools (Møller, 2006b: 40 - 41). I briefly consider some of these initiatives and their impact upon England and Norway policy and practice in the period leading up to and including 2008.

The OECD “what works” series of studies on educational innovation highlighted developments and challenges to the field of education, developing during the 1990s. In 2001 focus was placed on the changing demands for the management of schools, with examples of policy responses including case studies drawn from

---

<sup>11</sup> My translation.



9 countries<sup>12</sup> (Centre for Educational Research and Innovation, 2001). The report referred to *school management* as “essentially a twentieth century invention” resulting from abrupt challenges at the close of the last century arising from a supposed need to “download” managerial responsibility to the individual school leader (Centre for Educational Research and Innovation, 2001: 17). The report highlighted the “triple challenge” of educational and school redefinition, service and performance focus across the public sector, and creation of learning organisations focused on knowledge management (ibid.). As Shuttleworth later recognised the 2001 OECD /CERI report also highlighted “the tension that exists between the “top-down” approaches to reform, based on an industrial-age scientific managerial style, and those seeking renewal from the “bottom-up” through knowledge leadership in the 21<sup>st</sup> century learning organizations” (Glatter, Mulford, & Shuttleworth, 2003: 79). Shuttleworth saw this as developing a “loose/tight system”, considering it to be an effective approach to ensuring accountability at national and local levels. Also evident were increases in testing and inspection, more complex levels of decentralisation and a wider role for the school leader within the community. This was part of the protracted development from seeing school leadership as an extended teacher role to a “full-time professional manager of human, financial and other resources”, adding responsibilities such as instructional leadership, evaluation and assessment of staff and wider school performance (Glatter et al., 2003: 81). However, the perception of recreating the head teacher within a transformational role of “motivational leader” and “knowledge manager” was considered to require a new approach to leadership preparation.

These developments have placed the role of the school leader in particular, but also school leadership further under the spotlight. In this era of focus upon increased accountability and quality, the school leader’s role is changing, as the school is observed to change from institution to organisation (S. G. Huber, 2004). The question is raised as to how well leaders are prepared for this new, ‘patchwork’ role (*Ibid*). Writing in 2003, Shuttleworth considered this to be a “neglected” area, requiring increased investment to renew “self-esteem, learning capacities and leadership skills” (Glatter et al., 2003: 82-3). Interestingly, however, when referring to this area of neglect the author uses the term “training” of school leaders, whereas the 2001 CERI report referred to leader development (2001: 32). The implication appears to be more upon training, as competencies and skills of the ‘principal’ are highlighted as central for school improvement based on the development of learning communities. The CERI report appears, however, to juxtapose training and development, distinguishing between different content, delivery mode and timing and coverage of the initiatives. This will be further discussed in section below with regard to a brief overview of the development of policy approaches and programmes in England and Norway.

---

<sup>12</sup> Of which England was one.

In 2005 the OECD reported on the project “Teachers matter: attracting, developing and retaining effective teachers” (OECD, 2005b). The report highlighted the importance of implementing strategies that would, in turn, improve school leadership quality in order to improve the conditions for teaching staff that would improve quality of schooling (OECD, 2006: 3). This combined with the CERI research outlined above, developed into the Improving School Leadership Activity, aimed at gathering “information and analysis” to “assist [policy makers] in formulating and implementing policies to support the development of school leaders who can systematically guide the improvement of teaching and learning” (OECD, 2005a: 5). Of particular interest to this study, the rationale for the activity raised the importance of evaluating professional development and training programmes (OECD, 2005a: 9), noted to be often based upon “standards of professional performance” (OECD, 2005a: 11). Reflections from this activity are outlined next.

### ***Improving School Leadership***

The Improving School Leadership (ISL) activity involved a comparative study across 22 countries, including background reports, case studies, country workshops, international conferences, publications and a final report and website<sup>13</sup>. With recognition of the changing role of school leaders, the aim of the study was to: “synthesise research on issues related to improving leadership in schools; identify innovative and successful policy initiatives and practices; facilitate exchanges of lessons and policy options among countries; [and] identify policy options for governments to consider”. One of the key questions focused on how to ensure leaders would “develop the right skills for effective leadership”, recognised to be distributed and not merely based on formal position.

The executive summary of the OECD report on the Improving School Leadership project opens with the statement:

“School leadership has become a priority in education policy agendas internationally. It plays a key role in improving school outcomes by influencing the motivations and capacities of teachers, as well as the school climate and environment. Effective school leadership is essential to improve the efficiency and equity of schooling” (Pont, Nusche, & Moorman, 2008: 9).

Within this short paragraph the key concepts regarding many of the current debates over schooling are raised. The authors recognise a focus upon “leadership” centred upon within public policy. Additionally, the “key role” of leadership is linked indirectly to improved “outcomes”, mediated through the development of the school climate and teaching staff. While not a positional focus, the old prospect of the school principal or headteacher being challenged,

---

<sup>13</sup> [www.oecd.org/edu/schoolleadership](http://www.oecd.org/edu/schoolleadership)

the assumption is when leadership is “effective” “efficiency and equity” will improve. There is continued focus upon the issue of understanding how school outcomes can be improved, most recently with regard to ascertaining the “value-added” impact of schools (OECD, 2008). The report goes on to recognise the greater demands on leaders. While they are awarded increased autonomy they also face greater accountability for output. As will be seen later, increased autonomy of decision making is not, however, always symptomatic of increased control over resources (Aarebrot, 2006).

***ISL reports: a view from the environment.***

Both England and Norway participated in this project and delivered background reports. Most pertinent to this study, the framework for the background reports outlined the structure for inclusion of a chapter response on the training and development of school leaders. The sixth chapter of the response to the OECD was to include six major sections: policy concerns; preparation of school leaders; professional development of school leaders; relevant research studies; policy initiatives and innovative approaches (OECD, 2006). While the report from England appears to aim to broadly cover the questions associated with each section, the Norwegian report deals with them chronologically. Key points from the sections are outlined below. This overview is not essentially intended to be a direct comparison of the policies and practice in both countries. Neither are the arguments raised supported or challenged from other sources. The purpose is rather to provide an overview of policy current at the time of empirical investigation and practice as reported to the OECD from the respective governments. Pointers with regard to evaluation are thought to offer especially helpful background data. This information will be outlined further with regard to development of school leadership training and development programmes in both countries.

The England report links the developing policy concerns for the “preparation, development and certification of school leaders” with research on the impact of leadership on student outcomes (Higham, Hopkins, & Ahtaridou, 2007: 57). These concerns were once again tied to the drive for improved standards of schooling, “preparing pupils to achieve economic and social well-being in fast changing world” (ibid). Training programmes are designed to produce “effective leaders to meet these demands”. The resulting leadership programmes were thought to “expand” and develop instructional leaders. Focus is then placed upon the pathways to leadership and their considered effectiveness, followed by qualifying requirements for the school leader role. The English response is centred upon policy development since 1998, and particularly the NCSL and its Leadership Development Framework (Higham et al., 2007: 58). The report localises responsibility for programme evaluation upon the NCSL and outlines their framework of participant feedback and external evaluation of new programmes as well as provider based internal assessments. With regard to qualifying requirements the England report presents a long section, outlining in more detail the National Professional Qualification for Headship (NPQH) as

qualification for the school leadership role, as well as its revision based upon various evaluation exercises. The report reiterates that the NPQH will be mandatory for newly appointed head teachers. The report also links research findings to these evaluations (2007: 67) and highlights the underlying national policy focus upon evaluation and impact assessment of all initiatives. Evidence based research is referred to, but it is not clear how this has “informed policy development and to what extent” (OECD, 2006: 21-22). This important area of evidence informed policy and practice is returned to below. Evaluations and research are claimed, however, to reveal the “need for a more contextualised, personalised and innovative approach” to school leader training (Higham et al., 2007: 69). Discussion of alternative pathways and the difference between the two is taken up in relation to the Continuing Professional Development (CPD) of heads. The section recognises the decreasing number of participants taking HEI programmes. A University Partnership group is, however, linked to the NCSL “to support progression between the national programmes and higher degrees as MAs and MBAs” (2007: 66).

Chapter 6 in the Norway report focuses more upon outlining the development of what was seen to be a fragmentary framework of programme initiatives. Unlike the England report there is virtually no discussion of the linkage between leadership and pupil outcomes. What is interesting, however, is presentation of the various actors involved the debate over school leadership training and development, especially highlighting the role of the county and municipal authorities as school owners, and their jurisdiction over their employees. In this sense “best practice” is accepted as the basis for such activities but as decided by school owners and therefore is presented as quite a fragmented understanding. They are responsible for ensuring leaders are competent and equipped for their roles, as well as for “evaluating, developing and implementing” the programmes (Utdanningsdirektoratet, 2007: 56), a point also recognised elsewhere. While the national and regional authorities have previously offered training and in-service programmes, initiatives ranging from modular courses to Master programmes are now offered by HEIs. There is, however, noted to be great variation in competence over the subject field, the ability to evaluate and the “profiles and standards” of the programme providers. The report suggests criticism of the providers when they do not operate in a supportive role and meeting mandator needs (2007: 57). In the Norwegian document the sections related to leadership pathways and requirements are placed together and outlined within one page (compared to the 6 pages given to the sections in the England report). This section recognises how the lack of a national mandate combined with the fragmentary approach at local and county level means that less than one fifth of school leaders have formal training that has lasted more than a year. Supporting school leaders or potential leaders to attend Master programmes, or parts of Master programmes that can lead to future formal qualification, has been one part of the initiatives followed by the employers. Head teachers are though required to have formal educational qualifications on appointment. The section on frameworks and legislation in two short points, merely confirms the lack of

state regulations and the resulting lack of coordination (2007: 58). The section on development and evaluation offers interesting data. It highlights the lack of a standard based approach to evaluation and absence of impact assessments. Connection is also made between evaluative activity surrounding the programmes to the introduction and development of quality assurance initiatives following the 2002 higher education reform. It is recognised that these “self-evaluation[s] and surveys of participant satisfaction” might provide information on programme quality but do not provide impact data on learning (2007: 59) and it is “not known whether Norway has in general “good” school leaders” (2007: 61) even though some local authorities attempt to ascertain this (2007: 64). The report recognises that there has also been little research into programme effectiveness, particularly from amongst programme providers (2007: 60-1). Involvement in the ISL project is considered to signal a new national initiative in introducing new policy for school leadership training and development. A major part of this is considered to be aided by attending to the paucity of data that are available with regard to this area (2007: 64-5). There was a perception from within the Directorate for Education that this process would contribute to concrete policy change (Hegtnun, 2007).

I now consider these reflections in more detail within the context of school leadership training and development programming in England and Norway at the time of study.

## **2.2 School leadership training and development in England**

In this section I provide a brief overview of school leadership training and development in England. Leadership, as Harris put it, is “currently in vogue” (2003: 9) and there has been a significant shift in focus away from management (Bush, 2008a). Drawing on MacBeath (1998), Harris notes that efficiency at the micro level is thought to be a solution to “macro-problems” in society. At pupil level and across the school, leadership is considered to have an “unequivocal” potential to influence performance, in both the effectiveness and improvement research paradigms. Understanding of what makes leadership effective has therefore become a critical concern. In England the shift to management on site, post 1988 Education Reform Act, has required once again a new leadership role, which at the same time strengthened the position of school leader (Gunter, 2005: 181). In more recent times there has been a public policy shift towards approaches based upon New Public Management and later Modernization (Coupland, Currie, & Boyett, 2008). Despite the current agenda of Modernisation with its emphasis on social goals, Coupland et al (2008) note the economic focus associated with New Public Management, dealt with more fully in the next chapter, remains in focus. This dichotomy causes difficulties when attempting to interpret the role of the head teacher, for example “tightly controlled bureaucratic systems in a rigid hierarchy” are balanced against the emphasis on operating as transformational leaders (2008: 1080).

### ***A brief history of initiatives and programmes***

The organisation and provision of training in the 1960s and 70s has been described “ad hoc”, delivered by universities, Local Education Authorities, Her Majesty’s Inspectorate and other professional bodies (Bolam, 2004; Bush, 2008c). The introduction of the comprehensive system in the 1960s to early 1970s, was considered to demand a different role from the school leader, and focus was placed upon what skills and training were required (Bolam, 2004: 252). This period also saw the relatively late introduction of University based Master programmes, which were later developed through the establishment of the first chairs in educational administration during the 1970s (Brundrett, 2001). The national framework was considered, however, to have remained “unsatisfactory” (Bolam, 2004).

Brundrett notes that from the 1980s more substantial, practically based training courses were arranged under greater Governmental direction that included “visits to schools and other institutions, seminars, private study and encounters with managers from other fields of education, commerce and industry” (2001: 236). The National Development Centre for School Management Training (NDC) was established at Bristol University and ran for five years from 1983 – 1988 (Bush, 2008c). The Centre coordinated short term programmes for heads and deputies delivered from regional centres based in HEIs, as well as the promotion of “management development to schools and LEAs” (Bolam, 2004: 253). The introduction, however, of “centrally determined and accredited training for those seeking to move into headship” was seen as “an attempt to break with the past” (Gunter, 1999: 251). Interestingly, drawing on reflections from within the field, Bush commenting on the reflections of Gunter and Hughes et al. noted that “[u]niversity courses on school and college management became increasingly popular” (2008b: 74). By the 1980s taught higher degrees in educational management were becoming more important parts university courses, but demonstrated “a patchwork of provision including certificate, diploma, MA, MBA, M.Ed, M.Sc. and Ed.D. courses” which, despite “confusing variety”, provided a “comparatively structured provision of progressive academic qualifications grounded in both theory and practice” (Brundrett, 2001: 235). It was in this period that there developed a greater “coherence and coordination” of initiatives (Bolam, 2004). The initial emphasis was upon “voluntaristic and pluralistic provision” (Gunter, 1999: 251). However, most MBA courses continued to have more of an academic focus than their counterparts in the USA and mainland Europe. LEAs also provided ‘in house’ training but this provision was inconsistent too (Bolam, Dunning, & Karstedt, 2002); the first MBA education was set up at Keele University in the mid-1990s (Gunter, personal correspondence). The rationale was rather to provide reflection than training, to be studied over a longer period of time, thus explaining the focus upon Master programmes. Similarly the ‘professional doctorate’ had emerged, aimed at “mid-career education professionals”, the first at Bristol University in 1992, then further courses at 8 other HEIs (Gregory, 1995 and Myers, 1996 quoted in Brundrett *Ibid.*).

Policy developments from the later 1980s brought what Bolam described as “significant changes” (Bolam, 2004: 253). School management training was designated a priority within national funding from 1987, which Bolam recognised it to have remained “in one guise or another, ever since” (2004: 253). The Education Reform Act of 1988 also placed greater responsibilities at school level and upon the head teacher and senior members of staff in particular (Bush, 2008b: 74). In 1989 the School Management Task Force (SMTF) was set up by the government to assist the execution of reforms (Bolam, 2004: 253). It operated until 1992 in collaboration with LEAs to tighten control and coordination of training and improve access. It also introduced biennial appraisal for heads and deputies as well as mentoring programmes for new Heads (ibid.), considered “it’s most important legacy” (Bush, 2008b: 74). Bush notes that the SMTF, and particularly its report on the way forward for school management training, “set the agenda” for the ensuing period (ibid.), despite the National Professional Qualification for Headship appearing as a shift from the more supportive mentoring scheme (Bush, 1998).

Throughout the 1990s increasingly greater moves were made towards centralisation of control over training and development issues. The Teacher Training Agency (TTA), which would later become the Training and Development Agency (TDA), was established after the 1994 Education Act. Its purpose was to assist the improvement of teaching quality through better training, education and professional development of staff from recruitment to headship in order to “improve the standards of pupil’s achievements”<sup>14</sup>. Part of this remit saw the framing of a leadership development structure, based upon preparation, induction and in-service training which led to the respective introduction of the National Professional Qualification for Headship (NPQH), Headteacher Leadership and Management Programme (HEADLAMP) and the Leadership Programme for Serving Headteachers (LPSH) (Bolam, 2004: 253). These programmes were later transferred under the responsibility of The National College for School Leadership (NCSL). In the next section I will outline some critical issues related to the formation and development of the NCSL. This is not an exhaustive account of the College but rather provides the background for understanding current climate and approaches to school leadership training and development programmes in England.

***The National College for School Leadership*** After plans had been announced in 1998, The National College for School Leadership (NCSL) was officially launched in November 2000 as a Non- Departmental Public Body<sup>15</sup>. The NCSL is government funded, receiving its remit from the Secretary of State, currently from the Department for Children, Schools and Families. In September 2000 the

---

<sup>14</sup> Source: <http://www.archive.official-documents.co.uk/document/caboff/pubbod97/tta.htm>

<sup>15</sup> Now an executive agency

former Secretary of State, for the then Department for Education and Employment (DfEE), David Blunkett, transferred responsibility for the administration of the three national headship training programmes to NCSL, which should operate, amongst other things, as the “single national focus for school leadership development, research and innovation” (DfEE, 2000). A result of the College’s think tank was the formation of the five stage Leadership Development Framework, framing the provision of programmes from emergent through to consultant leadership, currently though under revision (Bush, 2008b: 75). A key part of this was the NPQH, described by NCSL as the “flagship programme” and “designed to establish leaders’ suitability for headship” based upon national standards (Bush, 2008b: 77). The NPQH, in its revised form is from 1<sup>st</sup> April 2009 a mandatory qualification for applicants to become a head teacher for the first time, and according to the College focused “solely upon those who can demonstrate that they are 12 to 18 months from headship and are committed to applying for posts immediately after graduating”<sup>16</sup>. Good overviews of the NCSL’s programmes can be found in Bush (2007, 2008b).

The other initial aims of the College were to “be a driving force for world-class leadership in our schools and the wider community; provide support to and be a major resource for school leaders; [and to] stimulate national and international debate on leadership issues”<sup>17</sup>. From 2006 four goals were outlined for the College, “to transform children’s achievement and well-being through excellent school leadership; to develop leadership within and beyond the school; to identify and grow tomorrow’s leaders; [and] to create a ‘fit for purpose’ national college that is more strategic and offers school leaders even more leadership support”<sup>18</sup>. The goals are more specifically focused upon the work of the College, reiterating the connection between excellent leadership and improve pupil achievement. The importance is declared of “tailoring... services to individual and local needs” combined with drawing inspiration from international “best practice” so as to “remain an authoritative national voice”. Since opening, College programmes have provided over 230,000 places. The remit of the College, in line with Government policy, is now being extended to include provision for wider children’s services, with the proposed name change to the National College for School and Children’s Leadership<sup>19</sup>. This continues the “mission creep” associated with the College (Riley & Mulford, 2007).

According to Bolam, in “one generation” a new model of school leadership development had gradually been framed, linked to the holistic restructuring of in-service provision and marketization of HEIs and developments in the

---

<sup>16</sup> Source: [http://www.ncsl.org.uk/aboutus-index/press\\_office-index/pressoffice-latestreleases.htm?id=31619](http://www.ncsl.org.uk/aboutus-index/press_office-index/pressoffice-latestreleases.htm?id=31619)

<sup>17</sup> Source: <http://www.ncsl.org.uk/aboutus-index/pressreleases-index/pressreleases-2002.htm?id=13920>

<sup>18</sup> Source: <http://www.ncsl.org.uk/aboutus-index/about-role-index.htm>

<sup>19</sup> Source: <http://www.ncsl.org.uk/about-role-remitextension> . It actually became the National College for Leadership of Schools and Children’s Services.



“professional infrastructure” (Bolam, 2004: 255). In addition, Bolam goes on to recognise that at this point the government returned to an idea mooted from the 1970s and 1980s for a national college. The idea resurfaced under the Conservative government of John Major, but was followed through by the first Labour government of Tony Blair (Bolam, 2004: 256). Bolam argues that it had again become a viable alternative due to political linkage to the perceived potential to raise school standards, technological advances reduced the scale of residential provision necessary less important along with the increase in political will to invest. Bush notes that the scale and reach of NCSL has been “impressive”. Additionally the NCSL has raised the national focus, developed a strategy for career stages, emphasising practice and basing programmes on research much of which is produced by practitioners (Bush, 2008b: 75). Despite these momentous advances and increased funding during this period there has also been, over a period of years, a developing ambiguity of purpose for school leaders (Bush, 2004) who have been challenged to develop skills closer to that of the CEO than the traditional head teacher. This is similar to Møller’s reflection over moral dilemmas (Møller, 1997).

These aims are quite different to the focus of the traditional HEI approach to school leader development, and even though the pathways through NPQH are considered to be becoming more open, allowing for a different role for HEIs, the basic difference in aims and objectives appears to create a continued discontinuity. Glatter noted the relative popularity and success of the master degree in derivatives of educational leadership management, but recognised that this was challenged by the introduction of the national programmes “closely tailored to assumed career and professional needs” (2004: 213-14). However, the impact of the NCSL upon HEI Master Programmes has been dramatic, where the latter are unable to ‘compete’ with the statutory requirements and funding support the college has received. HEIs refocusing upon research and international students may only be delaying the demise of their “specialist centres” of educational administration and leadership (Bush, 2006: 510).

The moves made towards the NPQH becoming mandatory for first time head teachers, locked school leaders further into the nationally sanctioned system. Content of the national programmes varies considerably from that on offer at HEIs, and as yet further research is still necessary to discover what impact these developments have had on the preparation of good school leaders, despite the assertion by then Schools Minister Jim Knight that the “NPQH is the best preparation for headship”<sup>20</sup>. In addition NCSL domination has been considered to be “unhealthy” (Bush, 2008a: 85). Even though Glatter opened for the possibility “in the long term” that engagement with national programmes might lead to increased interest in postgraduate education (2004: 214) this has not yet been evident with more HEI departments downsizing or closing as a result.

---

<sup>20</sup> Source: [http://www.ncsl.org.uk/about-us/index/press\\_office-index/pressoffice-latestreleases.htm?id=31619](http://www.ncsl.org.uk/about-us/index/press_office-index/pressoffice-latestreleases.htm?id=31619)

Gunter (1999) argued that the processes were far from linear and developmental than they might at first seem. Even before the NCSL was formed, the introduction of the NPQH under the auspices of the Teacher Training Agency (TTA) had seen challenges to HEIs as providers of school leadership programmes. This was part of a public sector restructuring, or what Bolam referred to as “the latest stage of an evolving policy innovation” (Bolam, 2004: 251). For example, the idea of leadership development stages was already part of the NDC and SMTF initiatives (Bolam, 2004: 256). Bolam argues that the scale and execution of the NCSL have however been wider, signifying a major policy change. Gunter quotes Ouston (1998), who noted that greater political control of these processes and subsequent changes in funding reduced the role of HEIs, with Ouston correctly predicting that many would see their role reduced to contractors of “centrally approved programmes” (Ouston, 1998 in Gunter, 1999: 252). The challenge to produce a dual system of programming allowing for central training programmes of “professional qualification” alongside Master programmes offering a more “academic qualification” had initially been ignored (Bush, 1998). Despite cooperation between a Universities Partnership Group and the NCSL there appears to remain some tension and difficulty coordinating the two approaches. Bush recognises the significantly negative impact of the NCSL generally upon HEIs (2008a: 85). The NPQH, as noted above being implemented in its revised form, operates in monopoly as the route to headship (Bush, 2008b: 79). Discussion over the transferability between Master’s degree and the NPQH has been on-going, with debate over the number of credits to be awarded as well as the additional requirements that candidates might be expected to achieve. Additionally, the NPQH has, Bush observes, “always been more concerned with what leaders can do, than with what they know and understand” (Bush, 2008b: 77). Gunter concurs, noting that the assessment focus on “completing tasks sends out the message that headship is about getting things done” (1999: 260). While, as Bush also notes, the current model also requires “master’s-level work” to be produced by candidates the tension still appears to remain between producing “a sufficient supply of ‘qualified’ candidates” and providing a programme structure that will demanding enough to “contribute to improving standards of headship” (2008b: 79). Bush notes that very few taking a College programme make a linkage to a postgraduate degree (2008a: 85).

Commenting on the introduction of the NPQH, Gunter considered that the “normative models of leadership promoted by government agencies and their collaborators present certainty in the cause and effect connection between effective leadership and effective schools” (1999: 255). This approach was perceived to remove debate and lead to a greater degree of “contractualism” backed up by the introduction of national standards aligning more closely the causal connection between what the head does and outcomes” (1999: 257). As a result a qualification was introduced that evaluated the ability to ‘get things done’ rather than dealing with processes that “cannot be assessed” central to the headship role (1999: 260). Gunter warned at this stage that “trainer and candidate satisfaction with the training should not be interpreted as an

endorsement that the NPQH is meeting its objectives” (1999: 262). Gunter revisited these arguments reiterating the contrast between the contingent nature of the head teacher’s everyday experience of work and the “normative” central policy focused upon effectiveness within a system focused on and driven by performance (Gunter, 2001: 158). She noted that “[o]utcomes are targeted, prioritised, and maximised, with a strong utility imperative towards value for money through direct and measurable impact” (Gunter, 2001: 157). Improving attainment through enhanced capacity for leadership has been NCSL’s “primary purpose”, even though there have been weak levels of empirical support to demonstrate this, which has been “an undue burden on the NCSL and the schools” (Riley & Mulford, 2007). This lack of empirical support is interesting when considered against the focus upon building policy decisions upon supporting evidence. This subject is considered in the next section, first in relation to the policy making process in England and then implications for the NCSL. These processes are considered to impact choices made about evaluation.

### **Approaches to the Policy making process**

The policy innovations outlined by Bolam are considered related to the developments outlined earlier within and across the OECD. Ball discussing Labour education policies in England post 1997, claims that they are “not specific to Labour” but are “local manifestations of global policy paradigms – policyscapes” (2001: 46). Any differences are claimed to be of “emphasis rather than distinctiveness” (2001: 47). Ball calls this “paradigm convergence”, characterised by a commonality in principles, operations and mechanisms which focus on impacting the profession (first order) and wider social justice (second order) (2001: 48). Part of this more general paradigm convergence is the increased focus upon student outcomes to meet future economic demands. As will be seen in section 2.3 these developments have also been considered to apply increasingly to Norway (Karlsen, 2006; Møller, 2006a).

This focus upon outcomes requires policies that will result in provision and programmes that lead to improved standards. Such policies should be informed by evidence. While commenting on between school discrepancies and the quality of schooling Christine Gilbert, HM Chief Inspector of education, children’s services and skills, declared there to be “no quick fix but providers should learn from what works” (Lipsett, 2007). This perception of evidence based and informed policy has become a critical underpinning of educational, as well as more general public policy initiatives. The current policy making process is focused upon the concepts of ‘what can be measured’ and ‘what works’. Taking the former, Broadfoot, in adopting Lyotard’s conception of ‘*performativity*’, explains that ‘educational assessment’ has been “the defining principle of education policy in the late twentieth century” and “[r]ooted in the rationalistic assumption that it is possible and, indeed, desirable – to ‘measure’ performance” (2001: 136-7). Ball describes performativity as “a technology, a culture and a mode of regulation that employs judgements, comparisons and displays as means of incentive, control, attrition and change – based on rewards

and sanctions (both material and symbolic) (2003: 216). This is a move from “professionalism and bureaucracy” (ibid) to “an existence of calculation” based on “targets, indicators and evaluation” (2003: 215). This focus is built upon a developing belief that assessment raises standards, especially since 1988<sup>21</sup>, and competition provides a “valuable spur to improvement” (Broadfoot, 2001: 138). The suggestion that quality can be measured<sup>22</sup> creates the cornerstone of policy and in turn reflects upon the policy of encouraging focus upon leadership for the improvement of results. The “new hero of educational reform” was seen to be the manager, given the task of transforming co-workers to feel concurrently accountable and committed to the organisation (Ball, 2003: 219), although this already suggests a greater shift towards the focus on the leader. Ball suggests a link between a performativity based approach and symbolic and constructed behaviour rather than the expected increased openness and greater transparency that were anticipated by policy makers. Organisations will be reshaped by these monitoring processes and act according to what is expected rather than what they believe in (Ball, 2003: 220). Organisations ultimately produce “paradoxical fabrications” of themselves for different audiences related to demands placed upon them, as they attempt to deflect attention while submitting to the frameworks (ibid: 224ff). I return to the implications for this consideration with regard to evaluation and decision making in subsequent chapters.

Denham (2003) remarks how the Labour Party manifesto from the 2001 general election focused on two main themes: “maintaining a strong economy and reforming the public services”. During the second term education was a special focus of the public sector improvements. Denham clearly points out that “to achieve this, ministers envisaged an increasing role for the private sector, a new approach that would refashion public services and the welfare state in line with the needs of the twenty-first century, redrawing what were deemed to be old-fashioned, out-dated boundaries between the public and private sectors” (2003: 282). These developments would be more clearly connected to revised economical ideals, and ‘depoliticised’, that is more rules-based, with less room for ‘political discretion’ and greater treasury control over management of the economy (Grant, 2003). This is a model of governmental choice over means and ends (after Burnham, 2001, in Grant, 2003).

The policy approach of the recent Labour governments was declared to be one of “high challenge, high support” (Barber, 2001: 19). This approach focused upon “continuous improvement” and setting high standards for pupil achievement, increasing accountability and measures of performance, and investment in quality, partly by modernising the profession and creating a framework of professional development. Barber linked this to “standards based reform” in the USA (Barber, 2001: 21). Elliott and Doherty considered this part

---

<sup>21</sup> The Education Reform Act, the purpose of which was to raise standards and pupil achievement and create an educational market (Broadfoot, 2001: 142).

<sup>22</sup> By applying assessment findings to criteria expressed in *categoric* form (Broadfoot, 2001: 138).

of a wider neo-liberal reform agenda, where educational policy is subordinate to economic imperatives (2001: 209). These standards should be “world class” and hinge upon the provision of good candidates to leadership and teaching positions. This requires schools to take “responsibility for their own destiny [which] puts a high premium on leadership” (Barber, 2001: 32). Barber goes on to claim that “[i]t may be a simplification to say that the difference between success and failure is the quality of the headteacher, but it is not far from the truth”. Headteachers move from “smooth administration” to an “unrelenting focus on pupil outcomes” (ibid: 33). Commenting on Barber’s (2001) outline of national education policy, Fielding notes that “the only acceptable arbiter is a rigorous and undeviating insistence on what works” (2001: 3). Hargreaves, however, was “optimistic” that this “pragmatic approach” could disseminate “what works” as “good practice” through the system considering it to be controversial only because educational researchers found it to be so (Hargreaves, 2001: 203). More direct intervention into professional practice was considered appropriate if “informed” by evidence rather than ideology or political preference (2001: 204).

This idea of evidence informed<sup>23</sup> policy and practice (EIPP) is considered to improve decision making processes, where the evidence should help “inform” but not necessarily “define” (2001: 205). EIPP consists of a five-fold policy development process whereby “decision-making at every level can be done in the knowledge of the best possible evidence” (Sebba, in Levačić and Glatter, 2003: 56). Levačić and Glatter note how this process links to the “knowledge chain”, consisting of, “knowledge systems, knowledge creation, dissemination, absorption and application in decision-making and practice” (*Ibid*) that involve intensified interaction processes. In this context knowledge, evidence and research appear juxtaposed, mixed and interchangeable. Such approaches are considered to favour “evaluative and instrumental” knowledge (Ribbins, Bates, & Gunter, 2003). The kind of ‘evidence’ that was now considered not to make it into the syntheses of research was notably “‘grey skies’ research – independent of government, critical of it and driven by an ideology resting on values antithetical to government...” (Wallace, 2001: 27). EIPP was adopted as one of the processes to extend educational and health policy in England, signalling a shift from experiment based evidence to evidence applied within social settings (Kemmm, 2006). Kemmm and Sebba (2003) recognise the early rationale was thought to be to enable swifter acceptance of up to date research thought to be ignored by practitioners and thus improve policy decisions. The rationale is that solutions developed from one situation can be applied to other settings, which has evoked criticism of ignorance to context (Kemmm, 2006: 321-2). As a result Kemmm, along with others, portrays a suggestion of reductionism within evidence-based policy, which indicates a prevalence of search for linear causal

---

<sup>23</sup> As opposed to evidence *based*, this definition offers an understanding of ‘the conjectural and changing nature of knowledge and the often dichotomous relationship between knowledge production and its application (see Levacic and Glatter, 2004).

models. One indicator of this problem has been shown to be limited time-scale available for making new policy decisions (Schwartz & Rosen, 2004). Hargreaves recognises that this approach is still part of a wider “act of faith” (2001: 206).

The move towards evidence informed policy and practice was bolstered by the opinion that educational research was “second rate” (Ribbins & Gunter, 2003). This was summed up in the view increasingly adopted by government and policy makers that the research was “not cumulative or coherent and was too often inward looking, irrelevant and lacking in impact” (Hargreaves, 1996 in *Ibid*: 169). In so much as the general premise of EIPP comes under criticism, one should question perhaps its basis further. Gunter, however, considers EIPP a policy ahead of its own data<sup>24</sup>. Wallace (2001) recognised that EIPP was a new governmental policy agenda, and that policy agendas frequently tend to serve policy makers interests. Any partnership would be based on predisposition of the central actor in the partnership, the initiator, in this case the Government.

This is considered to have taken place alongside the narrowing in understanding of the leadership role within English schools vis-à-vis the developing concept of improved public sector leadership. Simkins considered that a policy environment was developing placing ‘intolerable’ expectations on public sector leadership, both in “range... complexity and [in subsequent] internal tensions” (2005a: 15). Simkins recounts the complications of implementing a ‘what works’ policy programme<sup>25</sup>. If establishing a “powerful and engaging vision” is what works, then Simkins questions whose vision it is and for what purpose it is implemented. This creates new demands upon leadership, which is equated with sense making rather than implementing prescribed solutions, even if the process is evidence *informed*. As a result school leaders must see themselves first as public service employees; the policy of implementing ‘what works’ calling for greater cross sector appraisal in the market-led economy (Levačić & Glatter, 2003; Performance and Innovation Unit, 2003). Suggestion is made that “leadership development initiatives and new leadership colleges”, of which NCSL has been a key part, can contribute to better public sector leadership. Internal sector reports are deemed necessary to chart the impact and effectiveness of these initiatives, and the end-to-end review executed by NCSL and DfES (2004) is an example of this with regard to ascertaining and understanding the link between leadership and output and outcomes. As suggested above, these directives were not without their critics, with Government emphasis claimed to present school leaders as “senior managers in medium-sized business enterprise[s]” with their “schools as businesses in a market-led economy for education” (Thody, in Caldwell, 2004: 2).

---

<sup>24</sup> A point outlined in plenum during a lecture, 10.09.2007 at the University of Oslo

<sup>25</sup> See also Desforges (2003).

There is also a challenge for HEI evaluation of programmes. Such programmes are inevitably difficult to evaluate in terms of impact. The longitudinal connection, in particular, is not naturally there without an institution initiating some research processes. Neither are the input and outcomes controlled to the same extent as training programmes. As will be seen in Chapter 4, there is also a sense in which quality assurance has become the main focus of programme evaluation at universities, with the primary focus often upon the experience of those undertaking a programme and reflections over its delivery rather than appraisal of programme content and some measure of efficacy.

As will be seen below, Norway has only begun in more recent times to outline a similar policy making approach. For example, Karlsen (2004) in responding to the Norwegian Government's White paper, "Culture for Learning" from 2004, noted a growing tendency in both domestic and international education policy towards greater proximity between financial and educational politics. This was, for example, recognised in Norway in regard to the appointment to key positions within educational advisory boards of those with business administration qualification (in Norwegian 'siviløkonom'), rather than professional qualifications in pedagogy, considered to break with political tradition.

### **Evaluation and assessing impact**

The process of assessment and evaluation of school leadership programmes has increasingly come to the forefront in recent years. In this section I will discuss further how the changing focus has framed approaches at the NCSL. These processes have also affected HEI programmes in a more indirect way through, for example, the funding frameworks for HEI programmes but as also through engagement of academics evaluating programmes for NCSL.

The PIU<sup>26</sup> report into public sector leadership (2003) highlighted desire for a modernised public service, including greater acquiescence to customer demands, increased public-private partnership and demands for greater use of information technology. Subsequent demands upon leaders include deeper generic training alongside this "freedom to lead". As a result school leadership is bureaucratized and drawn more closely into the role of civil servant, where an improved evidence base is developed. This does however conflict slightly with the focus on evidence informed policy outlined above even though the terms appear interchangeable in government documents from the time. The PIU report focuses quite clearly on generic skills<sup>27</sup> that can be adopted and adapted across the public sector, which includes schools, despite the caveat that contexts be respected. However we see that in one section schooling is compared with defence, raising the question of how 'generic' and transferable leadership skills should be (PIU: 2003, 20-21).

---

<sup>26</sup> Performance and Innovation Unit

<sup>27</sup> See also Wales, 2004.

As was recognised earlier, NCSL functioned as a Non- Departmental Public Body (NDPB), a public sector organization required by government and therefore required to be “transparent, responsive and accountable” bound by public service agreements and performance targets, reflecting good business planning and reasonable public expectation (Office of Public Services Reforms, 2002: 31). What is clear from this definition and similar Government documents is the increasing requirement for public bodies to set individual goals as well as adopt cross-cutting objectives<sup>28</sup> and collective expertise in public service provision. This was positioned within the climate of overall focus upon the then Labour Government’s four principles of public service reform, “high national standards<sup>29</sup> devolution and delegation to the local level to and clear accountability; encourage diversity and creativity; flexibility and incentives to encourage excellent performance at the frontline; and expanding choice for the customer” (*Ibid*: 9). NDPBs were given a statutory statement of aims and are overseen by an independent board, but are also challenged to maintain a clear departmental focal point at senior level. The intention with ‘departmental sponsors’ was not strictly to direct but to help monitor performance and set agenda within the broader policy context (*Ibid*).

Brundrett (2006) notes how one of the NCSL’s five key objectives from its 2003 corporate plan was to demonstrate its own impact on leadership in schools, which was further reinforced by the then Secretary of State for Education, Ruth Kelly, who outlined that the College would be liable to “rigorous evaluation and impact assessment, as part of a strong research and evidence-based approach” (Kelly, 2004:4, quoted in Brundrett, 2006: 474). While the extent to which the NCSL arranges for the evaluation of its programmes is considered positive there are also many challenges (Bush, 2008b; Earley & Evans, 2004). Earley and Evans were, for example, unsure as to how some of the indicators set up to demonstrate the impact of the College were directly attributable (2004: 327). The authors were sceptical as to whether further research would reveal any stronger findings with regard to impact of leadership on pupil outcomes, and likewise the connection between training and leader development (2004: 335).

In 2004 the NCSL underwent an end-to-end review which addressed these ideas further. An end-to-end review is an: “in-depth study of specific outcomes from the point of policy making through to service delivery, leading to focused and evidence-based change programmes” (DfES & NCSL, 2004: 3). Under the terms of the initial NCSL remit letter (DfEE, 2000) stress was quite clearly made upon a desire to move towards greater steering and accountability alongside a development of what would be EIPP principles, where the direction would come from Government. Part of the end-to end review notes how NCSL was considered in a position to:

---

<sup>28</sup> matching PIU objectives outlined above.

<sup>29</sup> And these are also matched in the National Standards for Headteachers



“develop a clearer and stronger contribution to this policy making process, based on its robust knowledge, experience and interpretation of the system’s diverse needs and views, which is relevant to the government’s priorities and needs, well timed, clearly presented and precise” (DfES & NCSL, 2004: 13).

This report also “clearly presents” the intention to evaluate and assess programme impact and as part of the “ten key issues” raised by the report focus was to be placed upon “[g]reater understanding of the linkages and mediators between leadership and educational attainment and social outcomes” (DfES & NCSL, 2004: 7). Under the key issue of “maximising the articulation between research, practice and delivery” the NCSL would seek to apply “stronger effectiveness measures and a more developed analysis of participants and impact” despite what was considered to be “a lack of consensus” and “difficulty” regarding these issues (DfES & NCSL, 2004: 9). This should also impact “goal clarity”, noting that:

“[w]here there are significant uncertainties about linkages between aspects of leadership and educational attainment, NCSL research must continue to feed into the commissioning and design of high leverage initiatives, to enable controlled experimentation on different approaches. There should be robust, external evaluation of all programmes and initiatives, including evaluation of the relative efficiency, quality and effectiveness of different providers”(DfES & NCSL, 2004: 13).

This interesting move was combined with the introduction of a balanced scorecard of effectiveness for the College as part of a “relentless focus upon the achievement of national targets and priorities”, where they should press ahead despite the disagreement over the linkage between leadership programmes and pupil outcomes (ibid.). This approach was also based on a review of Leithwood and Levin’s report to the DfES, “Approaches to the Evaluation of Leadership Programmes and Leadership Effects” (Leithwood & Levin, 2004), revised and followed up as “Assessing school leader and leadership programme effects on pupil learning : conceptual and methodological challenges” (Leithwood & Levin, 2005). The content and findings of this report will be discussed further with regard to the issue of linking leadership and pupil outcomes, outlined in section 2.4 below.

### **2.3 School leadership training and development in Norway**

There is a sense in which the background account for Norway is less detailed than that of England. Clearly the advent of the NCSL, while building upon changes already underway in education policy has been a “paradigm shift” in the field of school leadership training and development (Hallinger in Bolam, 2004: 260) and had a great impact on other programme providers, especially those within HEIs. As will be seen in this section moves toward such approaches in Norway are in their infancy, however as was seen in the ISL reports outlined above they are very much beginning to appear upon the political agenda. The

basis for the broad overview in this section is taken from Country reports prepared for the HEAD project (Wales & Welle-Strand, 2005, 2008). These reports also outline these developments in more detail, especially with regard to details about specific Institutions and their programme content.

The importance of tradition and historical context is noticeably made by Norwegian commentators, but at the same time there is recognition of cross border convergence with regard to education policy and approaches to school leadership. Møller, for example, has similar reflections to those of Bolam noted above, considering that “[c]ountries’ approaches to school leadership are culturally and historically distinct, but at the same time they are currently drawn together by common economic and political forces” (2008: 1). A market approach to education adopted across Scandinavian countries during the 1990s was considered a shift away from the more “distinguishing features” of “equity, participation, and welfare state” under the influence of social democracy (Møller, 2008: 2). Whilst the structure of schools was traditionally flatter and more collegially based, where the leader was considered first among equals, the shift towards “managerial practice and external accountability” has required a greater degree of training and a stronger role (2008: 4-6). Møller considers that the local and regional authorities as employers responsible for training their school leaders have been more influenced “business management approaches”. The fragmentary nature of this control has led to great variation in terms of level of autonomy and managerial approach, where the focus on NPM can lead to reduced interest in leadership for teaching and learning (2008: 7). Attempting to balance these varied demands has created “moral dilemmas” for school leaders (Møller, 1997).

Due to the educational history of Norway, Headteachers have often found themselves under crossfire with pressure from above and below (Dahl, Klewe, & Skov, 2004a, 2004b; Møller, 2004). Telhaug et al have mapped a shift in education policy during the last fifty years across the Nordic countries, from socially focused objectives of schooling to cognitive-instrumental objectives combined with emphasis upon freedom of choice and decentralisation of power, where control of employment has moved from State to Local Authority (2006: 277). While generally developing at a slower pace of change the model operates as a “composite” of Anglo Saxon and Continental European approaches, combining “economic liberalism and competition” and “a large public sector, social welfare and security” (Telhaug et al., 2006: 278). The result of the change is “emphasis on equality, inclusion and adaptive learning” giving way to greater competition and increasing focus upon standards. This may, however, not always be to the detriment, because as change is slow in Norway old ideas might successfully be reappraised before the new ones are fully introduced (Olsen, 1996). Levels of governance appear to function on their own terms, despite overlapping interests, where each layer is given more responsibility. Each layer to a degree develops its own goals and framework. In addition, each

has an overlapping but differential stakeholder group (Afsar, Skedsmo, & Sivesind, 2006).

In Norway there has been a more diverse approach to school leadership training nationally, which became more necessary following greater decentralization of powers to school leaders in the 1970s and 80s (Karlsen, 1993, 2003). Although national programmes were attempted during this period<sup>30</sup> there appears to have been little success in operationalising their main aims and objectives (Johansen & Tjeldvoll, 1989). The University of Oslo offered “school leadership” as a semester module from 1992, building up programmes in concert with local and county authorities, which were partly based upon self-financing through to 2003 (Hegtun, 2007). Similar initiatives developed in other parts of the HEI sector in Norway. There has in recent times been more focus upon funding HEI programmes, which have developed into the current crop of Master programmes in educational leadership across the country, with the providers knitted in a loosely coupled National Network (Tjeldvoll, Wales, & Welle-Strand, 2005). These developed from the nationally funded “SOFF” project, focused upon digital and distance learning to form in-service and further education programmes that would lead to Master Degrees (Hegtun, 2007; Wales, 2004). The SOFF project also required the developments to be evaluated. The programmes were run as cooperation between HEIs. The introduction of the Quality Reform for Higher Education also placed demands upon the form of Master programmes, which had impact upon that which had been developed under SOFF.

The White Paper ‘Culture for Learning’<sup>31</sup> (Utdannings og forskningsdepartementet, 2004b) saw the Government announce plans for new national programmes in cooperation with LEAs via Kommunenes Sentralforbund (KS)<sup>32</sup>. KS has gained greater influence since decentralization, culminating in the municipalities receiving employer status in 2004. These municipalities became more focused on New Public Management as a governing principle during this period (Karlsen, 2006; Møller, 2004), where school leaders received increased responsibility and authority, as the municipal levels were reduced in number and flattened with more simplified political management (Finstad & Kvåle, 2004). At the same time State influence upon the local level has also shown elements of discontinuity, noted to travel via two channels, the practical (via the Ministry of Local Government and Regional Development<sup>33</sup>) and the theoretical (Ministry of Education and Research<sup>34</sup>) implying a theory/practice divide at State mandator level, requiring greater dialogue at the local level. The research noted, however, diversity of local approach, mainly between ‘bridgebuilding’ / unifying school development and a more fragmentary ‘contract-dialogical’

---

<sup>30</sup> Under the titles LIS, MOLIS, LUIS, LEVIS etc.

<sup>31</sup> Kultur for Læring

<sup>32</sup> The Norwegian Association of Local and Regional Authorities.

<sup>33</sup> Kommunal og Regional Departement

<sup>34</sup> Utdanning og Forsknings Departement

approach (Finstad & Kvåle, 2004). This appears to be the basis for the more common, generic form of leadership training outlined in ‘Culture for Learning’, with its focus on increased quality and “clear and strong leadership” (Utdannings og forskningsdepartementet, 2004b), noted to be missing from previous attempts to produce quality results (Søgnen, 2003). The current state of school leadership was also criticised, referring to Ekholm’s discussion, as “compliant” (føyelige ledere)<sup>35</sup> (Utdannings og forskningsdepartementet, 2004b: 28) implied to be reinforced by existing school leadership programmes. The pointers outline the critical attitude to the development of school leaders. As will be seen in later chapters, the relationship to local authorities as mandators has become increasingly more important for school leadership programme providers.

The document “Strategy for Competence Development” ( Utdannings og forskningsdepartementet, 2004a) suggested directions for how the new reforms should be implemented, as well as describing the areas of responsibility for their realization. Whilst recognizing that a network of providers who offer further education was already in place, the report suggests that there should also be, “further development of the programmes, so that they cover both the competence required for leading knowledge organizations in a process of change and development, and the more reform specific requirements”<sup>36</sup> (Utdannings og forskningsdepartementet, 2004a: 7). These developments were to be enacted in conjunction with a wider group of stakeholders than merely the programme providers embodied in higher education institutions, described as a priority. The *Strategy for Competence Development* also recognised the need for a restructuring of schools, moves towards individualised education plans and focus on building learning organisations, required significant capacity building for school staff, particularly school leaders and teachers. In the strategy document a decentralised initiative was outlined, in which the local/regional authorities as school owners and their respective schools would cooperate with HEIs and training institutions to make local plans meeting local needs in response to national demands. Hagen et al (2006: 45) in evaluation of the strategy suggested that HEIs could be allowed to play a much greater role in advising, implementing and assisting future development. However, they also note that questions of capacity and competence to deliver were being raised. For further discussion see Wales and Welle-Strand (2008).

With the advent of a new centre-left Government, questions were raised as to whether the policies of the previous government would be continued, paused or reversed. Following this in terms of school leadership training and development,

---

<sup>35</sup> Leaders who transfer responsibility for teaching to teachers, and engage little in dialogue about how the task of schooling should be achieved. It is said that these leaders do not develop collective practice and thus discourage true development (*Ibid.*).

<sup>36</sup> My translation

there appears broad agreement across the parties sharing Government, but also the wider parliament too, as to its importance and emphasis. The ruling coalition's governing "Soria Moria" declaration continues to outline training and development of school leaders as a priority<sup>37</sup>, particularly the focus on developing the school as a learning organization has remained a focus. The role of HEIs in providing suitable initiatives is of vital importance despite the fact that control and choice of content remains at school owner level.

Perhaps the most influential of the co-drivers of school leadership reform over recent years has been the local and county authorities, represented by their National Association (KS / NALRA). A report commissioned by KS, "Schooling and education on the agenda?"<sup>38</sup> (Bæck & Ringholm, 2004), was an investigation into the attitudes of top leadership of municipal and county authorities towards educational focus in relation to the factors they felt were of most importance. This report emphasised the new role these authorities received when taking over as employer<sup>39</sup>, particularly concerning their approach to the content and quality of schooling. A discrepancy between political will and practice was noted, with a disproportionate focus upon budgets and buildings in relation to discussion over the content and quality of education, which they had described to be most important. The report outlined 4 major factors for reflection: this discrepancy is probably due to a lack of knowledge; practical issues always appear take precedence over esoteric ones; focus for quality improvement is placed on the role of the school leader as key position for change<sup>40</sup>; and greater cooperation between school and employer is necessary for improvement. An unclear basis of competency with regard to programme design, implementation and evaluation between commissioning part and programme provider has been recognised (Hagen et al., 2006). This means that the role of HEIs in providing suitable initiatives is of vital importance despite the fact that control and choice of content remains at school owner level. Hagen et al (2006: 45) suggested that HEIs could be allowed to play a much greater role in advising, implementing and assisting future development. Møller, however, noted a diminishing respect for academia, as the local school owner focuses more upon developing more informal learning scenarios, in addition to drawing expertise more widely than the traditional HEIs, eg. consultancy firms (Møller, 2006c). This appears to inhibit the important factors of a formal education as the HEIs see them, while also refocusing the basis of what the school owner is looking for and will evaluate for. Møller called for a closer relationship between academia and the practice field rather than a withdrawal from it.

---

<sup>37</sup> [http://www.uhr.no/documents/Kunnskapsl\\_fet\\_januar\\_07\\_1.doc](http://www.uhr.no/documents/Kunnskapsl_fet_januar_07_1.doc)

<sup>38</sup> Skole og utdanning på dagsordenen?

<sup>39</sup> From 1<sup>st</sup> May 2004 the municipal and county authorities also became the teacher employer as well as having structural upkeep responsibilities.

<sup>40</sup> Especially in the 2 level authority structure.

## Later Initiatives

This section contains events and information that occurred post data collection. The reason for their inclusion is that respondents were aware of and involved in the preliminary and developing stages of discussion surrounding the issue of introducing a national programme for school leadership development and training. The information provided here therefore sets these debates in a wider context. Such a move had been increasingly discussed during 2007, both as a proposal for a mandatory programmes by Members of the Parliamentary Opposition, as well as being further drafted by the then Minister of Education, Øystein Djupedal, at education conferences and seminars (e.g. Djupedal, 2007; Smedstad, 2007). In addition the focus of such programmes was to be addressed. At a School Leaders Conference at the University of Oslo the Minister declared himself “surprised that the knowledge base concerning management and leadership in Norwegian schools is so limited”, noting that research in the field of school leadership had been little focused upon “uncovering possible links between leadership and pupils academic and social improvement and learning” but should be part of future research initiatives (Djupedal, 2007). In the Government White Paper 31 (2007-2008): “Quality in school”, the Ministry for Education and Research announced that a national school leadership programme would be developed. The programme would be for newly appointed head teachers as well as being made available to others without such an education (Kunnskapsdepartementet, 2008b). This initiative was claimed to be based on a broad desire within the profession, in the parliament, across the public sector and within the wider society. There was recognition of the increased responsibility and authority given to school leaders, particularly the administrative role which needs to be addressed within development and training (2008b: 64). In addition, the link between the leader role and “pupil outcomes” also began to be discussed a little more clearly along with developing vision and goals, local implementation of the national curriculum and teaching practice (ibid.). There is a clear desire to move towards the propositions from the OECD Improving School Leadership project (2008b: 66) as well greater focus on discovering “what works” in relation to school practice (2008b: 67). The proposal was to develop a programme for newly appointed heads and others “lacking such training”, as well as bringing greater clarity to the role of school leadership. The question of whether the programmes should be mandatory was not finalised, but interestingly the point was made that the Ministry did not wish “to introduce increased competence requirements for those to be employed as a school leader<sup>41</sup>” (ibid.).

The Ministry of Education and Research retained regulation of the content of the programmes, through which they would “make clearer the expectations and demands on school leaders” (Kunnskapsdepartementet, 2008a). The proposed providers were to be chosen from consortiums, where the leader organisation

---

<sup>41</sup> My translation from Norwegian.

should be found amongst higher education institutions, but the specifications demanded that they work in cooperation with other institutions / groups, and at least one of these groups should be from an environment other than “teacher training institutions”. The Norwegian Directorate for Education and Training was given the task of defining the demands and expectations for a head teacher, as well as initiating a competition for tenders for a national educational programme<sup>42</sup>. The development and implementation costs of the programme were to be financed by the State. The original framework was a 30 study point programme to be completed over an 18-24 month period, and should be compatible with a Master degree in educational leadership. The initial foci of the programme were to be “academic and pedagogical leadership, supervision of teachers and knowledge about change leadership within schools” (Kunnskapsdepartementet, 2008a). Additionally it was later made clear that the programme should be “controlled and goal oriented”, “needs focused” and “practically aimed”(Utdanningsdirektoratet, 2008b: 5-6). In line with OECD definitions, presentations of expected “competencies” were set out, covering required knowledge, skills and attitudes. The outline recognised the generic nature of leadership functions across different sectors. Within the document “Competence for a head teacher - demands and expectations”, the foci are outlined in 4 main areas: pupils’ learning results and environment; governance and administration; cooperation and organisation building; and development and change (Utdanningsdirektoratet, 2008a). The first area is given most weighting, while a fifth area is added that is based on self-reflection over the “leader role”.

This process signalled a move towards greater focus on pupil outcomes and a stronger leader role, while at the same time affirming many of the traditional approaches to school leadership in Norway. While the ISL background report did little to connect leadership to pupil outcomes, there was a change in emphasis as discussions turned to the necessity for a national programme. The Directorate announced that the studies would be built upon a definition of leadership as “taking responsibility to ensure that good results are achieved... in a good way” as well as ensuring staff have a “good and development work environment” in an “employer role”<sup>43</sup>. It appeared also to be the closest step yet towards introducing some kind of national “standards”, even though they are presented as “competencies”. The introduction and involvement of academic and consultancy based groups from outside of the traditional educational arena is interesting. Additionally, and of prime interest for this study, a key area for the tendered bids to address is evaluation, where the one mandatory area to be included is focused upon ascertaining the impact of the programme after completion upon the head and her/his school. Interestingly when discussing the values underpinning and processes driving the NPQH from its early stages, Gunter outlined the framework of what “aspiring head teachers are expected to know, understand and do” (1999: 252), which were the same categories set out

---

<sup>42</sup> [http://www.udanningsdirektoratet.no/templates/udir/TM\\_Artikkel.aspx?id=3720](http://www.udanningsdirektoratet.no/templates/udir/TM_Artikkel.aspx?id=3720)

<sup>43</sup> [http://udir.no/templates/udir/TM\\_Artikkel.aspx?id=3768](http://udir.no/templates/udir/TM_Artikkel.aspx?id=3768)

for the newly proposed national programme in Norway (Utdanningsdirektoratet, 2008a).

A key issue in both the English and Norwegian summaries has been that of evaluation, and in particular the growing interest in investigating and ascertaining the impact of programmes. As has been considered a central part of this demand builds on an assumption that there is a link between school leadership and pupil outcomes. It is to this important subject that I turn to next.

## **2.4 Linkage between leadership and improved pupil outcomes**

The purpose of this section is to offer an overview of research regarding the link between school leadership and pupil outcomes with the intention of illuminating the context for the data collection process. First I summarise some of the major initiatives and understanding that have contributed to this discussion, and then continue to consider how these hypotheses have also challenged the connection between leadership development and training through to pupil outcomes. There has been increasing interest in the effect, or impact, that development and training programmes for school leaders have upon the outcomes of schooling. While this accounts for interest in improved pupil outcomes, focus is also placed upon other variables like pupil motivation and engagement (Mulford & Silins, 2003). This perception underlies the forming of “specific” programmes thought “necessary if leaders are to operate effectively” (Bush, 2008b: 25). This raises many questions, but in relation to this study there is recognition that in order to evaluate for such an impact, some kind of causal chain needs to be constructed between school leadership and pupil outcomes. Then the next task is to connect particular programme impact with processes and outcomes at school. All of this is challenged by the point that leadership influence upon pupil outcomes, even though it is significant, is considered to be mainly indirect (Hallinger & Heck, 1998, 2003).

According to Day et al the English policy context is “dominated by concerns for external accountability and increases in the academic performance of pupils”, where standard focused governmental agency is considered to be interventionist, operating through “sustained and persistent initiatives” (2008: 5). The authors see improvement of school leadership as a central part of these initiatives, where focus on recruitment, selection, training and development is based upon the assumption of linkage between school leadership and student learning and achievement. The NCSL is considered “a highly visible manifestation of this attention” (Leithwood & Levin, 2005: 10). As Bush (2004) recognises, “[e]ffective school leadership and management are widely regarded as essential dimensions of successful schooling across the world today”. It is upon this “belief” that the NCSL has been developed (Bush, 2008b: 7). Identifying factors for school effectiveness have often been based on eliciting the views and behaviour of school leaders, constructing leadership as the independent variable (Møller, 2006b: 31). Research into the effects of leadership upon pupil outcomes suggests a small but statistically significant impact of the school leader, where



the greatest effect from schooling per se is noted via the mediating variable of teacher input (Hallinger & Heck, 1998; Leithwood, Louis, Anderson, & Wahlstrom, 2004; Mulford, Silins, & Leithwood, 2004). Hallinger and Heck report the effects of school leadership to account for 3 to 5 % of variation discovered in student achievement (Hallinger & Heck, 1998), although this has more recently been proposed to be somewhat higher (Bush, Bell, & Middlewood, 2010: 6)<sup>44</sup>. When taken into account with the total effect of variables that are attributable to the school, leadership amounts to about a quarter of it. The independent variable is, however, considered heavily moderated by contextual variables such as pupils' family background (Leithwood & Levin, 2004; Møller, 2006b), even though Leithwood et al. (2010) have more recently suggested that leaders might additionally be able exercise greater influence over this area.

Despite its “murky nature” and “limited evidence” from these research findings (Riley & Mulford, 2007), there is continued interest in the link between school leadership and pupil achievement especially to policy makers in England. A good example is the Department for Education and Science and NCSL sponsored large-scale study of successful leadership, undertaken as part of their ongoing interest in accumulating evidence on this subject. As part of this study Leithwood et al produced a “wide-ranging review of theory and evidence about the nature, causes and consequences for schools and students of successful school leadership” (Leithwood, Day, Sammons, Harris, & Hopkins, 2006b: 5). The authors consider there to be a fairly universal acceptance of the “significance of school leadership” but rather more debate with regard to its influence upon “organisational effectiveness” (Leithwood et al., 2006b: 12). As part of the report they drew initially on 5 types of empirical evidence addressing this issue. Qualitative case study evidence from “exceptional school settings” has uncovered “very large leadership effects... on an array of school conditions”, but this evidence is considered by the authors to lack external validity. Secondly, the authors noted that meta-analysis of “large-scale quantitative studies of overall leader effects” revealed evidence of small but significant effects on pupil outcomes when the direct and indirect leadership effects are combined (Leithwood et al., 2006b: 13). Thirdly, meta-analysis of such quantitative research but into “specific leadership practices” attempted to identify “leadership responsibilities” correlated to pupil achievement. Fourthly, the authors outlined research that has focused on leadership effects upon “pupil engagement”, especially how “transformational leadership” might impact this. Finally, the authors considered “leadership succession research”, which extended understanding how planned succession had positive effects upon pupil achievement just as high turnover showed negative effects (2006b: 14). The authors concluded from meta-review of the varied evidence that “leadership has very significant effects on the quality of the school organization and on pupil learning” (ibid.).

---

<sup>44</sup> The authors refer to research by Leithwood et al. (2006a) and Robinson (2007)

In assessing this research Leithwood et al (2006b) outlined a “State-of-the-Confusion” with regard to leadership impact upon outcomes. In particular they observed that research into successful school leadership had focused upon “values, beliefs, skills and knowledge” rather than upon “leadership practices” (2006b: 8). This had led to “unwarranted assumptions or links between internal states and overt leadership practices” that had become codified into “leadership standards”, which were exemplified by the National Standards for Headteachers in England (ibid.). In this way appeared to be an intermingling of research and evaluation where the standards for leadership practice underpinned perceptions of what was required for “effective leadership practice” and therefore how to assess its degree of presence or absence. The authors argued against such a position, preferring to focus upon “effective leadership practices”. The summary of their literature review led to what the authors call “the 4 core sets of leadership qualities and practices” which “almost all” successful leaders draw on (Day et al., 2008: 112). The 4 core sets are “building vision and setting directions; understanding and developing people; redesigning the organization; and managing (directly or through others) the teaching and learning programmes”, (ibid.). These were also outlined originally as part of “seven strong claims about successful school leadership” (Leithwood, Day, Sammons, Harris, & Hopkins, 2006a) drawn confirmed and supported by from the ongoing research findings and “comprehensive” review of literature (Day et al., 2008) that were outlined above. This research appeared to further strengthen the notion of important indirect effects of leadership on pupil outcomes which were stronger than direct effects. The authors also formulated a nuanced hypothesis with regard to the direct influence of the head, noting greater impact within disadvantaged contexts that offered greater challenge (Day et al., 2008: 111).

On consideration of these findings Day et al regard the underlying research support for the position of leadership impact as “particularly robust” (2008: 5) but also recognise that there are many areas requiring more sophisticated investigation. On reflection Bush considered that the impact of leadership and management as “significant factors determining school outcomes” is “not well supported by hard evidence” (Bush, 2008b: 7). Research into direct effects, and especially indirect effects, of leadership impact upon pupil achievement is relatively sparse. The EPPI review could only isolate 8 reports of evidence of leader effects on pupil outcomes (Bell et al., 2003). Witziers et al (2003) recognised weaknesses in the direct effects approach but considered that further indirect effects models may only confirm the link as “weak”. Although there has been an increased interest and more studies in this area in more recent times, many issues remain to be resolved. Bush therefore questions whether the difficulty of detection of leadership effects can support such assuredness, even though these effects are “by no means negligible” (2008b: 8). The ‘problem’ appears to remain the same for researchers, with no agreed ‘conceptualization’ of educational leadership or its location (after Hallinger and Heck, 1998) and the multiple interpretations of student outcomes, both cognitive and non-cognitive. Operationalising the concept of leadership and drawing data from the natural

setting continue to be great challenges (Levačić, 2005). Aside from perceived issues of methodological competence within the educational research field, Levačić notes, problems arise when ascertaining how many factors and mechanisms are needed for causal modelling, which she lists (but not exhaustively) as background factors, which will inevitably be hard to declare in closed model form. Levačić's causal field is open; it is not possible to identify and control for all causal factors. The problem remains consistent; within this field one cannot measure the counterfactual causal effects. Recent meta-analysis of research on the impact of different types of leadership on student outcomes suggested that "the closer educational leaders get to the core business of teaching and learning, the more likely they are to have a positive impact upon on student outcomes" (Robinson, Lloyd, & Rowe, 2008: 664). The authors recognise the contingent nature of leadership and open for more research, particularly that which connects leadership to classroom practice and "how leaders attempt to influence *the teaching practices that matter* ", arguing that focus on teacher impact upon students should provide the "source of our leadership indicators rather than various theories of leader-follower relations" (Robinson et al., 2008: 669). They note, however, that research into different leadership dimensions and their impact upon processes is still at a "level of abstraction". Fuller reflection over recent initiatives (e.g.: Gu, Sammons, & Mehta, 2008; Leithwood et al., 2006b), which include greater emphasis on longitudinal studies will be required. Following up this research Leithwood et al. (2010) have recently outlined 4 paths to improved student outcomes and attempt to isolate leadership practices that when combined will enable such improvement by influence.

These views are not however accepted fully across the field. There continues to be great disagreement over the linkage between leadership impact and improved pupil outcomes. Barker suggests that the "great majority of schools seem to be performing at levels that could be predicted from the knowledge of their intakes", taking into account socioeconomic status and student background (2007: 25). Additionally, Barker notes that while focus has been placed upon developing "strong" leaders to transform schools and improve performance, the amount and size of improvement has often remained undefined or related purely to test and examination performance (2007: 39). Case evidence suggested that positive changes in intake variables offered greater explanatory power for a case of school improvement than the transformational leadership qualities of the leadership, perceived to be a "critical organizational variable" (ibid.). Leithwood et al. (2010) suggest that leaders may be able to influence these variables as well.

While the issue of linking leadership practice to improved pupil outcomes remains controversial it also creates difficulty for the next area of study, namely the attempt at ascertaining the impact of training and development programmes for school leaders upon pupil outcomes. It is to this subject that I continue in the next section.

## **2.5 Linking programmes, practice and outcomes: considering impact**

Discussion with regard to the quality of evaluation has been clearly evident in discussions surrounding training and development programmes for school leaders. School leadership training and development programmes are, as has already been referred to, becoming a more common part of national public policy reforms aimed at developing the quality of educational provision (Hallinger, 2003). Educational legislation and statutory guidance has been more widely focused upon the necessity for “high quality professional development” that should improve school leadership (Guskey, 2003), which is visible in recent policy documentation in England, for example ‘Every Child Matters’, (UK Treasury, 2003) and as was noted above beginning to appear in rudimentary form in Norway. As was also noted this development is often linked to the contested belief that pupil learning outcomes will improve as a result of a better leadership and management skills base particularly in the formal leader of the school (Bell et al., 2003; Bush, 2005c; Leithwood & Levin, 2005). Within a system that has traditionally focused on accountability and assessment, evaluation in the United Kingdom<sup>45</sup> has been characterised by even greater visibility in recent times (Gray & Jenkins, 2002). As has already been noted, there has been a general shift in public policy focus from the evaluation of management of policy and resources to the management of outcomes.

A ‘tension’ exists over the fact that little is really known about the impact of these programmes. Even amongst training providers there is great variation in the acceptance of causal linkage between training and improved pupil outcomes. This to a great extent appears to reflect the way the programmes are evaluated. As Leithwood and Levin noted,

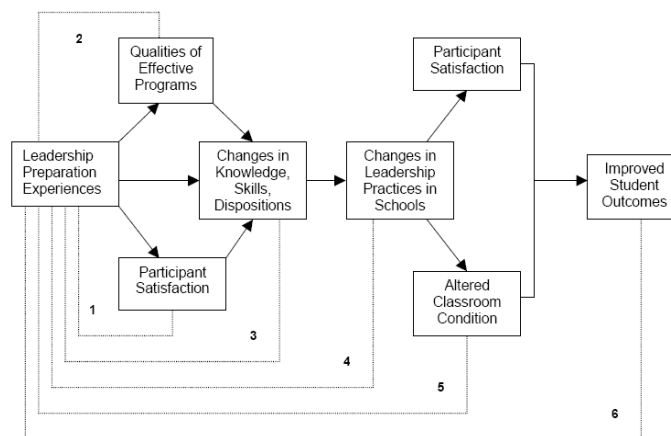
“In a recent analysis of leadership preparation programmes across the United States, McCarthy (1999) concluded that we do not actually know whether, or the extent to which, such programmes actually achieve the goal of “...producing effective leaders who create school environments that enhance pupil learning?” (p. 133). This gap in our knowledge is not because leadership preparation programmes are never evaluated; rather, the vast majority of such evaluations do not provide the type and quality of evidence required to confidently answer questions about their organizational or pupil effects. Most evaluations are limited to assessing participants’ satisfaction with their programmes and sometimes their perception of how such programmes have contributed to participants’ work in schools (McCarthy, 2002)” (Leithwood & Levin, 2005: 10-11).

The commissioned reports by Leithwood and Levin (2004, 2005) furthered developed a model to clarify the linkage between leadership and student outcomes, despite recognising the complexity of such a process and the

---

<sup>45</sup> Which the authors apply to the English context.

“methodological challenges”. They note that “highly sophisticated frameworks... potentially include all of the variables at the school and classroom level that are themselves the focus of independent lines of active research with the usual debates and uncertainties about their effects on pupil learning” (Leithwood & Levin, 2005: 10). The authors recognised weaknesses in current evaluative approaches and research, which would require greater comprehensiveness in measurement of leadership practices, formation of “an expanded set of dependent (outcome) variables”, description of how leadership influences “the condition of variables mediating their effects on pupils”, understand moderators of leadership effects, while using more varied methodological approaches (Leithwood & Levin, 2005: 4-5). Their *six-step* “framework to guide evaluations of leadership programmes” is included below, where within a hierarchy of increasing complexity the models building upon previous levels working towards level 6 where the evaluative criterion is improved student outcomes:



**Figure 3: Evaluation framework for leadership programmes. Source: Leithwood and Levin, 2005: 36**

The models face additional challenges from variation in school type, transferability of data, changes of measures in longitudinal data, missing data in addition to complexity with defining the unit of analysis (2005: 40-2).

There are, however, recognised problems with such models which were also dealt with to some extent by Leithwood and Levin. Whatever processes are set in motion, they offer little to the aid the discovery of whether a programme is ‘good or bad’, and are said to require greater ‘effort’ (Goldstein & Ford, 2002). Most evaluations focus upon trainee / participant reactions, but appear to say little about learning or improved outputs / performance (Goldstein & Ford, 2002; Guskey, 2000, 2002). Additionally, as will be seen below, these are considered short term (Bush: 2010). A vast majority of evaluation models for

formal training are observed to adopt rational perspectives (Holton III & Naquin, 2005). For example, Kirkpatrick's (1998) 4 level model continues to be widely adapted, despite considerable criticism (Alliger & Janak, 1989). Kirkpatrick's (1998) four level model is a hierarchical model ascertaining a programme's impact on participants in terms of their reactions to it, their learning, transfer of behaviour and the impact upon results in the workplace. Criticism of Kirkpatrick's model in particular, is that it is little more than taxonomy of outcomes, where the implicit causal relationships remain 'unoperationalised' (Alliger & Janak, 1989; Bates, 2004), while too many intervening variables that are ignored (Holton III, 1996). Perhaps Holton's strongest criticism is that the model relies upon 'participant reaction' as a "primary outcome of training", supporting Alliger and Janak's reflections that reactions are not linearly related to learning, but may moderate or mediate it (Alliger & Janak, 1989). Holton III followed up these research findings that demonstrate "little correlation between reactions and learning" (1996: 10), and therefore no direct link, but recognised that reactions have been shown to reinforce interest and enhance motivation acting as a moderator function (after Patrick, 1992, in *Ibid.*), whilst mediating other relationships (after Mathieu, Tannenbaum, & Salas, 1992). Holton's conclusion, like that made to some extent in the educational field by Guskey (2000) and Leithwood and Levin (2005), is that less focus should be placed on reactions to the process and more on performance outcomes. Reactions should be considered as an evaluation measure of the learning environment instead, moderating motivation to learn and learning<sup>46</sup>. It is often experienced that evaluations do not provide the information necessary to support evidence of their effects. Such findings further add to the necessity for research into underlying attitudes amongst decision makers responsible for evaluation, requiring models that moderate or adjust the more prescriptive rational models.

Kirkpatrick's 4 level model, is still a strong influence within educational evaluation activity (Guskey, 2000), and even more widely in Human Resource Management (HRM) (Holton III & Naquin, 2005). Guskey modified Kirkpatrick's model in response to criticism that it did not "reflect training's ultimate value in terms of organization success criteria" (2000: 55). This led to the inclusion of a level focused upon "organization support and change", to investigate which factors might moderate the impact of any development initiative (2000: 83). This is the level where the school can "support or sabotage" a professional development initiative (Bubb & Earley, 2007: 69). The information that organisations are claimed to base decisions on is thus declared to be flawed. The greatest problem appears to be the evaluation models applied to programmes, and the conceptualisation of what the organisation is attempting to achieve. It has been recognised as fairly straightforward to analyse programme structure, "potential utility" and participant perceptions of

---

<sup>46</sup> In meta-analysis Alliger and Janak (1989) found an overall correlation of .07 between reactions to training program and learning outcome.

effectiveness without ascertaining whether they “produce effective school principals” and without good measures of effectiveness (Cowie & Crawford, 2007: 133). Guskey’s model did appear to offer a wider view of student outcomes to include the cognitive, affective and psychomotor, where the data should be gathered by mixed and multiple methods (2000: 212ff). Commenting on this Bubb and Earley offer an interesting note to focus in England when they partly describe the cognitive as being understood in this context as “the most obvious – pupil attainment (the dreaded performance tables!)” (2007: 69).

Bush affirms the complexity of this model in ascertaining transference from programme to school (2008b: 123). He outlines two of the major problems with programme evaluations, firstly that they rely “mainly or exclusively on self-reported evidence”, seldom with an a priori element and secondly that they focus on “short term” impact (2008b: 114). This latter observation recognises that most changes will take place over time. These weaknesses are significant, as are Bush’s further comments with regard to difficulties of attribution. He refers to Bush et al (2006) adaptation of Leithwood and Levin’s model when interpreting the findings of their evaluation of the NCSL New Visions programme, noting the “diminishing influence of the programme as the model moves through each phase” (2008b: 120). The authors had concluded that “[p]roving a straightforward link” between a programme and evidence of school improvement “is fraught with difficulty” (Bush et al., 2006: 197). Simkins et al (2007, 2009) also adopted a model influenced by Leithwood and Levin’s, when evaluating 3 NCSL programmes. The authors also agreed that Kirkpatrick’s model is too linearly focused and omits key variables, particularly contextual (Simkins et al., 2009: 34). As a result they focused on factors influencing participant learning in relation to “in-school components” (2009: 29) developing from programme input through intermediate to final outcomes dependent on antecedents and moderators (2009: 35-7). Surveys were used to ascertain evidence of longer term impact, but in line with Leithwood and Levin’s outlined challenges data quality was weakened by poor response rates (2009: 31). The “poverty of theory in the evaluation of learning and development interventions” also creates the problem of investigating how “individual development might translate into *organisational transformation*” (Bush, Glover, & Harris, 2007: 15).

Møller considers another weakness of this model to be that it is grounded within the “rationalistic paradigm” and therefore ignores critical, institutional and political theories (Møller, 2006b: 35). Additionally it is mainly based upon Anglo-American studies and therefore limited in terms of generalizability, particularly as it ignores outcomes such as democratic and social development and outcomes, key areas in Norwegian education. Nordic research has focused much more upon educational frameworks, particularly from cultural and micro political perspectives and considering leadership from a relational perspective rather than based upon role but also perception of identity amongst leaders (Møller, 2006b: 37-8). As a result there is little research on school effectiveness

and scepticism to attempting to link school leadership and pupil outcomes, even though these areas are being increasingly referred to by politicians and within government documents (Møller, 2006b: 40).

The perceived link between training and development and improved outcomes is evident across the educational spectrum, where the terminology in governmental documents highlights the necessity of ascertaining “impact” of initiatives (e.g.: OfSTED, 2004; TDA, 2007). With particular regard to educational leadership Bush notes the importance of connecting investigations of impact of professional development with the nature, purpose and intended outcomes for the initiative in focus (2008b: 107). Bush recognises the importance of linking programme impact with the intended outcomes of the initiatives, but is an area that he claims has received only limited discussion (Ibid.). Bush goes on to outline how this discussion has been limited; mostly surrounding “student outcomes” and “school improvement”, which he considers to be a “vaguer notion”. For example, Flecknoe’s study of a CPD programme for teachers confirmed the difficulty of ascertaining a link to direct effects upon pupil outcomes and subsequently whether “all teachers” could be enabled to raise achievement on completion, “once the importance of the easily measurable has been exposed for its inadequacy” (Flecknoe, 2000: 455). Flecknoe further highlighted the challenges of controlling for halo effects (Flecknoe, 2002). It is an area still supported by “belief” rather than “evidence” (Bush, 2008b). Almost writing in terms of faith, González et al declared that “[h]owever ludicrous to some and uncomfortable to others it may seem, we believe in the existence of a linkage between principal preparation programs and student achievement in schools” (2002: 265-6). The authors’ purpose in studying such linkage was to refocus research onto outcome-based standards that would in turn help develop a model that could “adjust preparation programs with the intent of improving student achievement” building upon the development of internal activities (ibid.). Research in the learning and skills sector in England revealed a relationship between type of leadership development experienced and espoused views of leadership (Muijs, Harris, Lumby, Morrison, & Sood, 2006). Although the authors recognise that this offered no proof of causality, they observed different development forms related to different styles of leadership (2006: 103). The challenge of which models were most effective, still, however, remained. Outcomes from leadership programmes can include “sustained” change in leadership behaviour, school conditions, processes of teaching and learning and pupil outcomes (Simkins et al., 2009: 34). The authors affirm the protracted nature of the processes required for to uncover evidence of effects. Leadership development is a long term course of action requiring time for change to take root in others. There are additionally many other variables that will mediate and moderate the quality and timing of change. While this area has been of significant interest to the NCSL, meta-analysis of NCSL evaluations, however, revealed that there was little evidence of how the impact of programmes was understood and measured (Bush et al., 2007).



*Achieving outcomes, ascertaining impact of programmes*

Bush (2008b) interestingly outlines discussion over the “significance of leadership and management development”. The general “purpose of leadership development is to produce more effective leaders”, which implies the achievement of intended outcomes (Bush, 2008b: 108). The main reported criteria utilised in assessing the value and impact of initiatives include improvement of *pupil learning, attitudes and engagement*, improved staff *motivation, capability and performance*, and promotion of *equity and diversity, democracy and participation*. Bush outlines a series of alternatives and challenges with regard to the design components and focus of programmes which influences their assessment and evaluation. These are presented in the table below.

**Table 1: Criteria for assessment of the value and impact of leadership development programmes. Adapted from Bush, 2008b**

<b>Design components</b>	<b>Alternatives for programme design</b>	
Main purpose	Developing leaders	Leadership development
Underpinning	Succession planning	Meeting individual needs
Focus	Standards-based	Holistic development
Implementation style	Content-led	Process-rich
Aims	Specific repertoire	Contingency
Implementation context	Campus-based	Field-based
Participation and ethics	Generic <sup>47</sup>	Equity and diversity focused

Bush raises important issues with regards to these different components. He notes, for example, how the overall purpose of NCSL programmes has predominantly been directed towards training and developing leaders, particularly the role of the Head teacher. There has been a movement towards more generic programmes in line with the policy initiative of systematised “succession planning”, rather than those based upon “individual needs” of participants. The NCSL programmes are tied to the national policy initiatives, and their underlying values, that were outlined above. That these programmes are based on standards appears to constitutively highlight the importance of technical aspects of leadership and management of schools. These points are highly significant for this study as policy attention appears to shift in terms of content, and subsequently assessment, away from development based programmes, more usually associated with Master degrees. Bush notes that the

---

<sup>47</sup> Bush raises the issue of whether leadership learning should “address issues of equity and diversity”, the alternative position could therefore possibly be seen as ‘generic’.

apparent purpose of standards is to measure performance against an *articulated* “clear set of expectations for leaders”, whereby successful programme completion is seen to provide “baseline competence in the leadership role” (2008b: 110-1). But, as Bush further points out, such an approach is in danger of reducing the complexity and contextualised nature of this role. This raises issues for the evaluative approach, where constitutively the “measurable” takes precedence over the less quantifiable. While these approaches apply more strictly to the nationally mandated programmes under the control of the NCSL, the wider implications of the public policy initiatives outlined earlier mean that HEIs are also increasingly expected to focus in this way. This might be due to particular programmes being funded or part funded by national bodies, or as will be seen in Chapter 4, because the higher education field also faces parallel demands for evidence of impact, which is increasingly linked to funding. In the next section I consider examples of how evaluation of NCSL programmes has developed along with the ensuing debate.

### **Evaluation, school leadership and the NCSL**

While this study only focuses indirectly upon the NCSL, I consider how it has impacted the wider field of school leadership and particularly with regard to evaluation. The rationale behind this focus is that the evaluations of NCSL programmes are often undertaken by HEI academics under tendered contract and questions related to these exercises have formed on-going dialogue between the College and the wider academic and research field. An example of these debates was the BELMAS/SCRELM symposium in 2004. One issue raised was that of ownership and control of the evaluation process, where the “terms of engagement” were considered “determined largely by the College” (Bush, 2005a: 35). While evaluation aims were described as “absolute”, there was noted to be some flexibility over methodology and the possibility to suggest alternative methods or designs (2005a: 34). Such a process was described as a partnership but with controlling interests over some areas (2005a: 35). Additionally, questions are raised over the assumptions and purpose behind the evaluation act (Simkins, 2005b). Simkins reiterates the centrality of evaluation with NCSL programmes, but also considers how the underlying “expectations and constraints” influence choices about methodology and approach. He states, “[u]nderlying these choices will be assumptions about the kinds of knowledge that evaluations can or should generate”, particularly within a wider framework of increased accountability and focus upon improving educational outcomes (Simkins, 2005b: 35). The amount of resources and time given influences the process greatly. The issue of impact, in moving beyond participant reactions, is noted again to have been a difficult area within evaluations mandated by the College. This involves taking into account “contextual complexity” in specific evaluations and “joining up” the different activities for improving design and development (Wright & Colquhoun, 2005). This complexity issue with regard to impact is reiterated by Earley (2005), who isolated some of the shortfalls in models applied, particularly with regard to time and use of mapping techniques

and use of more qualitative approaches. He notes that baseline data and follow up studies are often of limited value (Earley, 2005; Earley & Evans, 2004), but that longitudinal designs fit poorly against demands for “immediate evidence” that policy is working (2005: 37).

A response from the NCSL recognised that evaluations had focused more particularly on programme effectiveness rather than on a “wider critical understanding of the evaluation of school leadership development” (Conner, 2005). This issue was also related to that of control, where the College hoped that good relationships with the wider academic field could be maintained. There was however a sense in which the College has increased its distance from the HEI field and less of a feeling of a partnership. NCSL was also concerned about the impact upon the quality of evaluations due to available time and resources, which had also been raised elsewhere (Southworth, 2004), but there was some suggestion of disagreement over how long this time scale should be. The College therefore appears to be represented as having a different purpose and agenda for its programmes, and subsequently their evaluation. This suggests a complex field and meeting of minds that needs unravelling. Within a climate of greater control and demand for evidence, it is suggested that analysis should also be directed to the whole development of evaluation models, in particular towards the decisions that guide the choice of approach and development of evaluation model with utilization in mind. Such engagement is considered interesting to explore further with subunit members.

What is interesting is the way the evaluative approach and decisions appeared to be reframed. Bush for example noted that this standardising approach provided “the potential for significant control over the nature and outcomes of evaluation” (Bush, 2005a: 35). There is a sense from these examples of how the shift in the framework of school leadership training and development in England has affected the nature of knowledge production and dissemination. The underlying desire to assess impact has shaped the evaluative process much more directly as well as the tighter linkage to programme effectiveness assessments which have been used for internal control mechanisms as much as for contributing to the debate about how programmes might improve school outcomes. There continues to be variation in the data gathered. This has also increased in complexity as leaders engage with more programmes as part of the leadership development framework. Brundrett outlined how research revealed that “many school leaders see value in national programmes” but there was still an extent to which “the effects of such training and the quality of such development are seen to vary across programmes and over time” (Brundrett, 2006: 484). However, the author did note a “tentative indication” that the “combined effect” of programmes has made a difference despite the continued difficulty in attributing causality.

Similar findings were revealed through more recent evaluation of the LftM<sup>48</sup>, NPQH and LPSH programmes (Simkins et al., 2009).

## 2.6 Discussion

The previous sections have in varying degrees of detail attempted to provide an overview of the field of school leadership training and development programmes in England and Norway, in relation to national policies as well as wider policy demands and goals within the OECD. The focus was then narrowed to consider how interest in improved pupil outcomes and raising educational standards has challenged evaluation frameworks. The increased search for a link between leadership and pupil outcomes has also driven the search for measures of effectiveness of leadership training and development programmes. Programmes are now increasingly challenged in terms of accountability, impact and degree of value for money. However, as Bush et al point out, “the impact of leadership development programmes on pupil outcomes... is by no means the only argument for the effective preparation of school principals and other leaders” (Bush et al., 2006: 198). Nevertheless, much focus continues to be placed on this area.

Impact models are noted to focus upon the degree to which a programme might work and why, investigating transferability and heavily linked to accountability frameworks (J. M. Owen, 2004: 362). There has been a developing pattern in public policy making towards national programmes which has also influenced programme providers of school leadership in Higher Education Institutions. Many of these academics perform external evaluations of the nationally mandated programmes. In England the frameworks for these evaluations are tied closely to these demands for “rigorous” assessment are to ensure the implementation of programmes that are most “effective” in raising standards while providing “strong” evidence to be made available to policy makers. While there has not been the same drive in terms of policies and demands in Norway it was seen that the issue of raising outcome standards is beginning to gain greater credence across the political spectrum and moves towards a national programme are developing more rapidly.

These specific developments are considered to be part of a general trend towards so-called evidence based decision making that is changing the approach of the evaluation field (Weiss, 2004) but its application is noted to be the exception rather than the rule (Pfeffer & Sutton, 2006). This suggests that it is a problematic to link evaluation, accountability, evidence finding and the discovery of best practice. At the same time research suggests that more frequently evaluations are not utilized, and that if they are may vary in terms of the original purpose. With greater demands placed upon providers of publicly mandated programmes to account for and evaluate their activity, there arise threats to the credibility of evaluation information (Schwartz & Mayne, 2005). It

---

<sup>48</sup> Leading from the middle

is recognized that political and organizational pressure can lead to a-priori bias, whilst there is also the more *pervasive* threat of ‘shoddy practice’. Organizations need, therefore, to attend to their “blind spots” and find and follow the “best data and logic” (Pfeffer & Sutton, 2006). For many this movement will suggest an over-emphasis upon experimental and positivist data collection, but this move to be considered as an attempt to address the weakness in information collected from programme evaluations, whatever the evaluation tradition from country to country. On many occasions this may equally be a problem of defining the purpose and intention of a particular evaluation, while attempting to regulate the possible future use of conclusions that might be drawn from its findings. The challenge of evaluation utilization is generally considered to be problematic (Alkin & Taut, 2003; Vedung, 2006b). As has been discussed, EIPP as a process is not unproblematic with demands to discover what works.

There is also evidence of more central control in the system, where the impact of the NCSL has on a wider level curtailed the role of the more traditional HEIs. Providers then are experiencing greater demands and pressures from both national and local mandators to evaluate more effectively and allow, as mentioned earlier, future policy decisions to be ‘evidence based’ while being accountable for activities undertaken (Anderson & Bennett, 2003; Simkins, 2005a). However, it is claimed that rather than ascertaining the impact of programmes on participants and changed behaviour at their schools, the majority of evaluation models focus only on how satisfied participants have been with the courses they have attended (Guskey, 2000; Leithwood & Levin, 2005). Overemphasis on ascertaining the connection between leadership and pupil outcomes might cause other more important factors to be ignored, contributing to ‘dangerously simple prescriptions’ about leadership impact (Simkins, 2005a). It might be said that this leads to an assessment of skills taught under training but less understanding of individual development. This is also reflected in wider research into the evaluation of training (Alliger & Janak, 1989; Holton III & Naquin, 2005; Kraiger, 2002). Organisations assigned to be evaluated or to self-evaluate their activities are forced to face the question of how they perform the task, what methods are employed and who is involved in the process. This too has been the traditional grounds for division within the evaluation field. Pressure to produce results of effects and provide evidence of impact appears to underlie the rationale for programme evaluations. One might then question the extent to which information gained can add evidence of successful programming and impact on educational outcomes, which is the wider purpose of the public policy initiatives. As was seen in the previous section, it has been considered unclear as to how findings from the evaluation of training can be utilized, as adjustments made to programmes resulting from such types of evaluation often appear small, incremental and self-reinforcing and are more likely based upon perceptions of learning rather than actual changes in performance (Holton III, 1996)<sup>49</sup>.

### **Focus on decisions about evaluation**

The reflections outlined above create an opening for greater understanding of how evaluation is designed and undertaken. But it also raises another question. What decision processes take place in response to these demands? How are the demands felt and how will they be responded to? When such demands for evaluation exist, along with differing purposes for use of the findings, it is thought useful to consider how such responses develop. In this context it is therefore considered to investigate the decision processes that guide the choice of model adopted for the evaluation of postgraduate programmes for school leadership.

The chosen area of investigation is considered salient for three key reasons. First there is a desire amongst mandators to use evaluation to extend the focus of the understanding of impact of training and development particularly on organisational change and pupil outcomes. In this way we need more focus on what evaluations are meant to achieve if it is believed that improving the quality and extent of evaluation will give greater information about the quality of the programme. Questions might therefore be raised as to the normative models and values that underlie these training programmes, particularly what impact this has on programme content and how they will be evaluated in the light of them. Second, choice of model is assumed to affect the utilization of the information, which is considered a major purpose of the evaluation process. It is proposed that while the same evaluation models might be applied across different contexts, the organisational values and decision-making processes underlying them might differ. In addition, the context surrounding school leadership training and development programmes is considered particularly interesting as programme content often relates to the use of evaluation in schools. That is, those evaluating their own programmes are often teaching about evaluation techniques to others and are therefore considered to be more aware of different rationale and practices. Research in the late 1980s had shown there to many educationalists within the evaluation field (Shadish & Epstein, 1987). This area of study bestrides two interesting organisational situations, programme participants are from schools and their development is based in HEIs, two arenas that according to Brunsson suffer the problem of knowing what they are doing and likewise suffer the same ignominy from their environment (Brunsson, 2002: 4). The third area that is considered salient concerns who makes the ultimate decision about evaluation design within the organisation that is implementing it. Thus questions are raised concerning the *demands* placed upon organisations to evaluate and be accountable, the *designs* chosen to meet these demands, and the *decision making processes* by which these decisions are made, that is, who is involved in making these decisions and how they respond to the demands?

With regard to these assumptions, analysing the underlying decision-making processes, that have often been ignored, should contribute to a more in depth understanding of the evaluation design process (Holton III & Naquin, 2005). In this way, even if better measures of programme impact are thought to be developed, and greater understanding is considered to have with regard utilisation and the quest for greater evidence, it is also considered to be important to investigate the attitudes and underlying values programme providers have to the purpose of the evaluation process and the way that they will perform the task. Scott recognises though, that decisions concerning “input characteristics and output environments” are generally theoretical rather than methodological issues (2003: 365). These, Scott recognises, are dependent upon whether data on throughput or “bridging” input and output processes are required. The question might therefore be perceived to be a value based one (2003: 372).

In the next chapter I will explore these issues further by outlining the theory underpinning evaluation models in more detail and further consider links to policy making approaches. These deliberations inform the context within which decisions about evaluation are made.

### 3. Evaluation purposes, processes and practices

As was outlined in the introduction, this study is focused on the evaluation of postgraduate programmes for school leadership development. In investigating the evaluation models designed, adopted, and applied to be implemented by programme providers it is felt necessary to draw together two main fields of study, evaluation research and the organizational decision making field. This requires appreciation of how these fields have approached the highlighted research problem, namely what influences the design of evaluations. As outlined in the introduction, the intention is to reconsider the process of evaluation by reflecting on the demands to evaluate, the designs developed to meet these demands and how and by whom these decisions are made. This chapter therefore presents a literature review concerning theories surrounding the purpose of evaluation; definitions of evaluation utilization and the factors assumed to influence it and how different levels of use are explained. The context of evaluation is also considered important especially with influences on the policy making process of ideas embodied in New Public Management. It is through describing and analysing the developments in this field that decision-making research is considered to offer a valuable contribution to the debate about evaluation. Chapter 5 will then outline different frameworks of decision-making, investigating decisions in action rather than prescriptive theory development. A combination of these outlines subsequently forms the theoretical and analytical framework for the study.

#### 3.1 The purpose of evaluation

Consideration of the purpose of evaluation is necessary to attempt to structure and inform a framework for empirical study of programme provider attitudes to how models are adopted. When attempting to comment on the purpose of evaluation, one must recognise the variety of intonations generally applied to it. Scriven states “it’s tempting to define evaluation as “whatever evaluators do” (2003: 16), but his synthesis of research literature characterises it as “the process of determining the merit, worth, or significance of things (near-synonyms are quality/value/ importance)”(Ibid.: 15)<sup>50</sup>, where the final act, or “evaluative claim” is referred to as the “So-what” question, following “evaluation logic”. The latter focuses on showing how to move from factual and definitional premises to evaluative conclusions (Scriven, 1991: 216). This question should enable the evaluator to apply the findings of what actually happened in the process under study, which Suchman referred to as the usefulness of findings for improvement of services (Alkin & Christie, 2004). But as Dahler-Larsen recognises, evaluation is a systematic data collection driven exercise, “[w]ithout data, no evaluation”<sup>51</sup> (Dahler-Larsen, 2006b: 75). There is a sense in which

---

<sup>50</sup> Scriven also interestingly distinguishes between normative and descriptive evaluation, which will affect the resulting ‘evidence’.

<sup>51</sup> My translation



evaluation is decision focused. Owen (2004) extends Scriven's focus on value and worth to consider evaluation more broadly as "a knowledge production activity". This approach focuses on knowledge utilisation for decision making rather than "logic", attending to issues concerning the negotiation between evaluators and stakeholders within identified audiences (2004: 361). Owen considers that it is these initial processes that shape the evaluation in hand. This connects his work to those theorists focusing on participation in evaluation as a key factor to improvement and greater utilisation of findings. It also acknowledges the importance of decision making in the initial stages of an evaluation beyond the framework or mandate expressed by commissioners.

It is important to consider how concepts of evaluation have developed over time. This will be important when attempting to frame the views of evaluation of those involved in the provision of postgraduate programmes in school leadership; considered also against their responses to wider demands placed upon them. Vedung (2006b) considers that evaluation is essentially based on a simple idea. He notes traditionally it was only important for public bodies to outline the principles and goals set for an initiative, and that procedures were followed and backed up financially (2006b: 109). However the public sector particularly has shifted towards focus upon results achieved through public policy initiatives, for example, through adopting management by objectives. I will return to these developments in section 3.5.

The 'trunk' of evaluation theory is described as being based on "a dual foundation of accountability and systematic social inquiry" (Alkin & Christie, 2004: 12). The 'need' for accountability<sup>52</sup> created a subsequent need for evaluation, which is also linked to the development of the advanced capitalist state (House, 1993). The basic rationale is found in the roots of liberal ideology and conceptions of liberal democracy, building on assumptions of freedom of choice, individualism, and empiricist orientation (House, 1978). While accountability "provides the rationale, it is primarily from *social inquiry*<sup>53</sup> that evaluation models have been drawn", defined as "the systematic study of the behaviour of groups of individuals" across social settings (Alkin & Christie, , op cit.). In addition, Weiss has asserted that the justification of evaluation is built upon "the contribution it might make to the rationalization of decision making" (1979: 17), where evaluation is essentially a "rational enterprise that takes place in a political context" (1987: 47)<sup>54</sup>. It is interesting therefore that literature from organisational theory and decision making has been considered to be 'ignored' when studying evaluation (Holton III & Naquin, 2005; Rogers & Hough, 1995). Such reflections are important to this study, where school leadership programmes are increasingly becoming part of a public policy trend. It is

---

<sup>52</sup> The authors define accountability 'broadly', concerned most with improvement of the activity under investigation.

<sup>53</sup> My italics.

<sup>54</sup> See also (Chelimsky, 1987; Nachmias, 1980; Nachmias & Henry, 1980)

thought necessary to illuminate how programme providers theoretically and ideologically consider the task of evaluation as well as practically. This is further dealt with in the next section. Whatever the evaluation focus, it is generally held today that findings will be *used*, even if there are different perceptions of types of use. Although evaluation utilization is not the main focus of this study, but consideration of it is still thought to influence the initial stages of evaluative activity. I will return to this subject in the final section of this chapter.

### **3.2 Theoretical basis underlying definition of evaluation**

It is also important to briefly consider different theoretical positioning of evaluation to help outline a framework for understanding how evaluators relate the processes of evaluation. While it is recognised that evaluation definitions will be based upon more than theories evaluators have engaged with, it is also considered that investigating the espoused theory that individuals favour or identify with might help illuminate the choices that they make concerning the models of evaluation implemented. It might also frame their response to demands as well as their approach to the decision making processes within their organisation. Such views might be thought to be included within their particular evaluation paradigm.

Christie (2003) attempted to account for the influence of evaluation theory on practice. She noted that there has been little comprehensive study of this subject despite its seeming necessity. Christie recognised from the work of Smith that developing instruments to ascertain theoretical point of reference is a complex task, as “evaluation practitioners usually are not proficient in theory” (2003: 9). Concurrently, the author attempted, in collaboration with eight theorists, to delineate major evaluation theories to offer a framework against which practitioner responses could be analysed. Interestingly for this study the focus of Christie’s survey was placed upon the theoretical perspectives concerning methods, values and uses of evaluation approaches. Even amongst the most experienced external evaluators surveyed, Christie found that few evaluators indicated that a particular theory guided their work, and even fewer that there was a particular text that framed their practice (2003: 13). This suggests that when attempting to ascertain the approach of an evaluator one should focus more on trying to unravel their description of the process undertaken rather than questioning of their relationship to particular theory. Christie used two dimensions to map her findings: scope of stakeholder involvement and method proclivity. In particular the second dimension “proclivity to a particular methodology” is useful for this study, even though it might be open to some challenge with regard to how quantitative and qualitative methods are presented on a continuum. Christie distinguished between external evaluators and internal evaluators, the former who tended to be educated to a higher level and more experienced in the evaluation field. Christie’s findings suggested interestingly that internal evaluators in an attempt to counteract suggestions of bias will

justify their work by choosing quantitative methods. Findings also suggested that the less experienced evaluators tended to follow the guidelines for the evaluation they should implement “strictly” as a “formula”, but overall they were most concerned with bias reduction (2003: 33). Drawing on the wider work of Alkin, Christie suggests though that there is a general propensity for both inexperienced and experienced evaluators to follow the legitimated and defensible norms of the field in an attempt to overcome challenges to subjectivity. As noted earlier, theory is rarely referred to and if done so not in its entirety (2003: 33).

The implications of Christie’s findings are considered important when interviewing evaluators about their decision making concerning the models that will be implemented. One might therefore expect there to be little reference to definite theories and texts regarding evaluation, but a propensity towards quantitative methods when focus is upon gaining greater legitimacy. In addition, less experienced evaluators will be more likely to adapt official guidelines and implement accordingly. In this situation of investigating postgraduate programmes in school leadership, it must be recognised that the context will play an ever increasingly more important part, noting that the programmes under investigation are delivered within HEIs. This means that evaluations will usually take place within or against the backdrop of QA systems. This is of course counteracted by the fact that most systems decentralise decision making concerning the methods chosen for evaluations to be implemented<sup>55</sup>. Focus on quality assurance development within HEIs is dealt with in greater detail in Chapter 4.

Hansen’s (2005b: 34ff) meta-evaluation of Danish assessment practice partly supports Christie’s propositions, revealing little documented discussion over methods and models used in particular published evaluation reports, notwithstanding a lack of discussion over possible alternatives. Hansen noted even more strongly in her survey that “only one out of five reports are theoretically informed, which [was] defined as reference to existing knowledge, other surveys/evaluations and /or methods books” (Ibid: 35). Hansen also refers though to the importance of the quality of design of an evaluation, with particular regard for how questions are formulated, models used, and basic relevance to the lens through which the evaluand is observed” (Ibid: 38). In pursuing the importance of these reflections I now turn to consideration of evaluation designs and models.

### **3.3 Evaluation designs and models**

The varying definitions and perceptions of evaluation have spawned a broad perspective of different designs and models of evaluative investigation. In their meta-review on this theme, Madaus and Kellaghan emphasise how the “conduct

---

<sup>55</sup> (Links to question 1 and section 1 more generally in the interview guide).

and nature of any evaluation is affected by how one defines the process of evaluation” (2000: 19). They go on to recognise that the plurality noticeable within the evaluation field is underpinned by “deep epistemological differences” and diverse opinion about process. As such, evaluation models tend to characterise a particular author’s view of process and subsequent suggestions for practical implementation rather than describe any widely accepted theoretical position. Numerous evaluation models have developed over time, often to match prevalent theories of organisation and management.

Without any wide acceptance of what models and designs should entail, the variation of perspectives has increased over time. As evaluation models developed, within the education field Stufflebeam (Stufflebeam & Webster, 1983) recognised potentially loosely coupled effects and uses of results in varied types of models that had been applied. He noted that many studies failed to match up to the purpose of designing and conducting an evaluation to assist judging and improving the worth of an educational object (1983: 24). Of these studies, some were *politically-oriented* “pseudo-evaluations” focused on presenting positive or negative images of a programme, “irrespective of its actual worth”; *questions-oriented* “quasi-evaluation” studies, which then apply a methodology thought appropriate for the particular questions to be addressed<sup>56</sup>, regardless of whether these are relevant for “developing and supporting value claims”; and *values-oriented* evaluations designed “primarily” to meet the basic evaluative purpose outlined above<sup>57</sup>. Stufflebeam considered that such loose coupling between purposes, designs and utilisation appeared to be exhibited by the varied perceptions of clients, practitioners and audiences involved. In his findings, clients tended to be driven towards the political models, evaluators towards the questions models, while audiences are keen to know the value of the object under investigation. Stufflebeam’s conclusion was that evaluators should be “sensitive” to their own agendas as well as that of clients and audiences, including possible conflicts. Stufflebeam suggested that evaluators should assess the relative approach of each model they intend to implement, collaborating with the client and users. As will be observed, the evaluation field has generally focused somewhere between improving methods and participation in the process. I will return to these points in the summary at the end of the chapter and suggest that this attention should also be supported by greater understanding of the internal decision processes that would appear to underlie these processes.

Researchers have attempted to understand the underlying approaches of evaluation models. House (1978) noted a division in the field between models based on subjectivist ethics, observable in both utilitarian and pluralist ideology, as well as those based on a more liberal objectivist epistemology, in which management focused models frame accountability, efficiency and quality control. According to House these ‘elite’ models generally emphasise

---

<sup>56</sup> For example accountability, testing and management information gathering.

<sup>57</sup> Exemplified by accreditation, policy and decision-oriented studies.

empiricism over theory, drawing more heavily from principles of scientific management and systems analysis, assuming a consensus of goals can be reached, which will define the focus of evaluation<sup>58</sup>. Much public sector evaluation still appears to follow this pattern or to demand information that responds to such a view. Utilitarian based evaluations build on a subjectivist ethic with an objectivist epistemology, determining what should be maximised. Pluralistic evaluation has both a subjective ethic and methodology, and as such is not generalizable, focusing rather on experience and socialization, where precedents become judgements. House's reflections appear to inform the basic competing arguments surrounding the evaluation of programmes for school leadership; while demands for a more managerial model are perceived to come from the mandator, the education field appears to consistently apply derivatives of the pluralist models (Guskey, 2000). In these models it is "particular experience" that is in focus rather than judgement of quality per se (House, 1978).

The reflections outlined above move discussion about evaluation beyond that of a purely rational exercise, for example ascertaining the input, process and effects of a specific programme as the basic operative model. Research needs to be further focused upon how processes take place within a specific context, against particular traditions and in relation to expectations and experiences of evaluative activity. Mark et al note the different traditions these models are drawn from have "influenced some evaluators' decisions about evaluation designs, each providing a way of defining success" (2000: 11). In this way understanding the views held by evaluators of the basic premise of evaluation and the purposes to which particular models are thought useful, is considered to be another important factor when attempting to investigate and enlighten their role in the decision making process.

As a result focus has now been further placed upon the design process of evaluation, especially in terms of how models are formed. Hansen (2001, 2005a) has briefly outlined distinctions of models that attempt to account for diversity within the design process of evaluations: negotiation models (what we can argue for); appropriateness models (what fits to the problem)<sup>59</sup>; routine models (what we usually do or have done before); competence models (what we can do). Hansen's categories appear to be useful heuristic aggregates but require further study. She suggests that there will, to a greater or lesser extent, be overlap from situation to situation and context to context. It does seem difficult to equate everything into a 'design' in the essence of enacting and implementing an evaluation. The point that both routine and competence may overlap heavily should require us to take a further step back and decipher how decisions are made in such organisations. Particular points of interest are how size will matter,

---

<sup>58</sup> Systems analysis will also assume an agreement about cause and effect relationships (House, 1978: 4-6).

<sup>59</sup> adaptation

whether the organization is public or private and whether those funding are internal or external commissioners.

A common reflection in the wider evaluation research field is that no model is better than another but depends rather on the actual evaluation question at hand (Krogstrup, 2006). Krogstrup considers the decision over choice of evaluation model to be more commonly normative or political rather than technical or rational (2006: 167). This may appear to be an oversimplification but it does open for possibilities not fully explored by a field that has more traditionally focused more upon describing and improving the technicalities of the process. Krogstrup adopts an evaluation definition drawn from the work of Evert Vedung, noting that “[e]valuation is a systematic and retrospective (and a prospective) assessment of processes, outputs and effects of public policy”<sup>60</sup> (Vedung, 1998, in Krogstrup, 2006: 17). As Krogstrup reflects, this begs a different question, considering what criteria will form the basis for an assessment. In an attempt to build up capacity to undertake evaluation, evaluation becomes built in and integrated in public organisations and evaluation tasks become institutionalised (Krogstrup, 2006). At the same time Krogstrup believes that there is a decreasing amount of evaluation and an increasing amount of performance measurement (monitoring) (Krogstrup, 2006: 21, 181ff). Greater focus upon performance measurement and monitoring favours a particular evaluation approach and type of information gathered. Such demands are considered as “external control” and accountability measures and are considered to focus to narrowly on particular types of information gathering at the expense of others. This factor will require a new and particular evaluation capacity building in organisations in order to release the knowledge left untapped by such processes that will reveal the social side of the organisation (Ibid.: 195ff). The significance of this statement is seen against the understanding that evaluation is too often poorly performed and utilised. If, in addition, it is poorly designed and limited in focus then it would seem important to build capacity which will progressively become institutionalised. Krogstrup agrees with Stockdill et al. (2002) that such capacity must become part of routine but at the same time remain flexible within a collective, incremental development.

To analyse an approach to evaluation would therefore seem to be helped by attempting to denote evaluation perspectives, that is, epistemological / ontological reflections on evaluation; penchant for particular evaluation models, that is, how to assess outcomes, outputs, processes, as well as the view of the evaluand under investigation. It is also considered important to understand the basic purpose or desired knowledge; guiding values; and intention for and attitude to utilisation<sup>61</sup>. Focus should additionally be placed upon applied

---

<sup>60</sup> My translation

<sup>61</sup> This is also drawn from Dahler-Larsen, 2002 and Shadish et al 1991, (in Krogstrup 2006: 69)

evaluation designs /concepts, that is, the choice of methods, those involved and the organisation of the evaluation itself. I will return to these reflections in Chapter 5. I turn first however to a consideration of how different fields as well as different countries can be perceived to have particular evaluative traditions that might be thought to impact choice of design and model of evaluation.

### **3.4 The important context of evaluation traditions**

Studies suggest that it is also important to recognise that attitudes to evaluation will also vary across professional and cultural boundaries (Bhola, 2003a; b). Bhola (2003a) views the educational evaluation field as a “normative-professional culture” around the world, but this is thought to be moderated by national traditions. The “normative tone” that Bhola describes is reflected in codified language and standards. Subsequent interpretations and translations are mediated into particular national systems, amongst other factors, by “the level of the professional cultures of educational evaluation, the level of the educational systems, and the general culture” (Bhola, 2003b: 401). This also affects the eventual utilization of evaluation information, which Bhola, outlining a form of bounded rationality<sup>62</sup>, compares to episodes of *satisficing*<sup>63</sup> rather than detailed forms of analysis. Such translating and refining of international standards into local contexts is considered a common development within public policy diffusion (Brunsson & Olsen, 1998).

Comparative research into evaluation traditions across national and regional borders has revealed some interesting differences thought pertinent to the context of this study. Karlsson (2003c) notes that while Western Europe has generally been considered to have adopted more positivistic models of evaluation stemming from research in the USA, which are progressively attached to developing forms of New Public Management (NPM), Scandinavia has appeared to approach the task of evaluation differently, with particular focus on developing a democratic ideal which too is framed within policies associated driven by NPM. While England and Norway may be considered to fit into these broad frameworks there is of course a danger of oversimplification.

Within a system that has traditionally focused on accountability and assessment, evaluation in the United Kingdom has been characterised by even greater visibility in recent times (Gray & Jenkins, 2002). There has been a general shift in public policy focus from the evaluation of management of policy and resources to management of outcomes. Taylor (2005: 604-5) sees this development since the 1990s as a “central component” of a political “self-regulating strategy of governance” within “a new regime of scrutiny”. In a NPM focused regime, political responsibility is replaced by the “virtuous circle of evaluation, evidence, [and] performance”. Rather than improvement focus, this

---

<sup>62</sup> Bounded rationality is outlined in Chapter 5

<sup>63</sup> Outlined further below.

virtuous circle establishes a “direct line of accountability” between users and providers within participatory democracy. This is concurrently observed in the increased focus upon “leading for results”; considered to be at odds with organisational “leading for learning” (Oldroyd, 2003). Evaluation has been a natural part of NPM, where decentralisation, in particular, has led to a greater degree of monitoring (Rist, Sandahl, & Furubo, 2002). This evidence based / informed focus favours evaluative disclosure of *what works* (Gray & Jenkins, 2002).

Norway could be described as a ‘latecomer’ to the concept of systematic evaluation, (Baklien, 1993; Sverdrup, 2002) especially of public expenditure (Ovrelid & Bastoe, 2002). This may reflect the small, ‘egalitarian’ nature of its society, which frames policy by consensus, participation, pragmatism and incrementalism; much of which may result from the country’s relative financial stability (Ibid.) The pressure to evaluate that has come, is not considered to have initiated in the research field (Sverdrup, 2002: 163). The lack of pressure to evaluate is also noted through the ideological underpinning that characterises much of the public administration. This has especially been noted with regard to educational evaluation. When demands for evaluation did increase, as a result of both national and international pressure, these demands tempered by the “progressive and left-radical dominance in Norwegian pedagogy”<sup>64</sup> (Imsen, 2003: 151). Imsen notes that even after pressures for more extensive assessment increased in the 1980s, there was continued scepticism to evaluation, with other terms like “research, follow-up and mapping” being used<sup>65</sup> (2003: 153). It is felt that in this context evaluations should be dialogic and democratic (Ovrelid & Bastoe, 2002), and as Sørensen (1994) notes, the evaluation model for research and higher education institutions is normally based on individual pedagogical theory, which he considers to have questionable relevance to organisational evaluation. Faced with the prospect of evaluating the impact of activity, Sørensen notes that there has traditionally been an underlying proviso that evaluation activity should lead neither to ‘ranging’ of organisations nor consequentially to differential resource allocation. Far from improving programmes, this activity might rather act as a legitimating channel. But there is greater demand to meet the need for ‘results and experiences’ that can inform political decision-making processes, especially those concerned with resource application, increased efficiency and cost-benefit (Sverdrup, 2002).

There appears to be a degree of convergence in demands for greater evaluation in both England and Norway, but the thrust is greater in the former. Much of this reflects two different evaluation cultures, where Norway has been a much later developer. There is a growing perception that Norwegian public activity needs

---

<sup>64</sup> While Imsen’s focus is at the compulsory schooling level, her comments also apply to the wider education sector.

<sup>65</sup> The scepticism continued even after the introduction of initiatives like goal-directed management, which is dealt with in the next two sections.



to be evaluated, however there continues to be conceptual disagreement about what that might involve and great variation in how it is practised, both in focus and quality. This is also the result of the variety of activities to be evaluated (Foss & Mønnesland, 2005). Evaluation research in Norway, within the Scandinavian tradition, stems much more from what Sørensen (1994) calls a quality-development model, rather than a model concerned with quantifying worth to any extent as is noted in the English tradition. This basic difference in origins of the field may well help understanding of variation in attitudes to the rationale for evaluation, its process and intention for utilization of any reported findings. Therefore, such framing, although broad, is used to inform the responses of providers of postgraduate school leadership programmes, with particular regard to their interpretation of environmental pressures and understanding of the purpose for evaluation. While it is important to recognise the evaluation traditions within the countries under study, it is accepted that this is will be both vary within and between them. In addition, the context is in flux and dynamic with more recent public policy approaches affecting this context. I turn now to investigate these developments.

### **3.5 Evaluation in an age of accountability**

In recent years changes in the administration of public policy have influenced evaluation processes more widely. On the one hand public spending cuts since the 1980s have led to decreased overall funding, and subsequently time, for evaluation, (Weiss, 1998a: 13) while on the other hand increased demands for accountability has noted a more intense drive for evidence of successful policy initiatives and investment (Dahler-Larsen, 2006b). These changes have taken place in the recent era of New Public Management (NPM) driven public policy. The UK is considered to be the birthplace of NPM, which then spread to become “one of the dominant paradigms for public management across the world”, advocated by both the OECD and World Bank, and influencing modernisation trends in England (K. McLaughlin & Osborne, 2002: 1) and Scandinavia (Eliassen & Sitter, 2007).

Hood (1991) was one of the first to expound the characteristics of NPM developing during the late 1970s and early 80s. Hood considered NPM to be a loose term for administrative reform doctrines that ‘married’ “new institutional economics” and “business type managerialism in the public sector”, despite likely contradiction (1991: 5). NPM is also described as “loose and multifaceted shopping basket” of a concept combining economic organizational theory and management theory, public choice and managerialist thought, that requires decentralization, devolution and delegation, but also greater central control (T. Christensen & Lægreid, 1999: 7) and considered “chameleon-like”(Homburg, Pollitt, & van Thiel, 2007). NPM is argued to be “neither a unitary program nor a clearly defined policy” and purports to encompass efficiency, control, accountability, decentralisation, privatisation and performance indicators (Power, 2005: 328). Greve, in considering the pioneering work of Hood (1991),

recognises two main ingredients of NPM: the use of market mechanisms and the application of private sector leadership methods, which led to fragmentation of public services and delivery, contracting and outsourcing and choice at point of delivery (C. Greve, 2006: 20). Greve subsequently notes that an implication of the implementation of NPM reforms has been that “public organisations shall deliver results. There shall be freedom for leaders to lead, and they shall have the tools to do so” (2006: 21). There is, additionally, a “movement towards ‘managerialism’ and away from ‘professionalism’” and a reduction of autonomy (Broadbent, 2007: 7).

Christensen and Læg Reid see NPM related policies as emphasizing “economic norms and values” over the traditionally legitimate ones (1999: 7), referred to as “Old Public Administration” also described as “classical” Weberian bureaucracy (Olsen, 2003), which has been considered rooted in justice and welfare for all (Møller, 2004: 190-1). But this raised suggestions of wastefulness and unaccountable behaviour. NPM focused therefore on greater efficiency (T. Christensen & Læg Reid, 2007). Olsen notes that NPM presents both “a global diagnosis and prescription: a centrally organized and rule-bound public administration is outdated and NPM represents an “inevitable shift” toward a more advanced administration” (2003: 5). Here “management by command is replaced by management by result, contract, decentralization, deregulation, commercialization and competition... and the special nature of the public sector is denied” (*Ibid.*). These processes leading to implementation of NPM based policy raise a dichotomy between ‘proactivity’ with increased freedom, which leads to greater accountability, in turn leading to increased control (T. Christensen & Læg Reid, 1999: 23) developing as a “hybrid” of “decentralization (let the managers manage) and centralization (make the managers manage)” (T. Christensen & Læg Reid, 2007: 8).

Power (2005) recognises that this “wave of reactions against the elements and assumptions of traditional public administration” draws on management theories and practices developed in the private sector has been controversial (*Ibid.*). Part of the “audit explosion”, NPM has also been assisted by the legitimacy given to “professional advisory groups” who helped its acceptance and implementation at a time when the role of the traditional professional “service providers” was being curtailed (*Ibid.*: 329). The very nature of auditing will vary according to state tradition, of which quality assurance (QA) has been a key line of development (*Ibid.*: 330). Greater control has been attempted through the definition of preferred outputs and outcomes and the forming of measures to monitor development (Broadbent, 2007). The developing “evaluation wave” has had great impact on educational policy and organisation, with greater demands for an evidence based knowledge base to mirror that considered to be found in the field of medicine (Dahler-Larsen, 2006b: 55-6). While this has been most visible at the compulsory school level, it has also affected higher education. From the cases chosen in this study it will become clear that QA has become an, if not ‘the’, important frame for understanding evaluation today, especially

within education systems. Further discussion of the basis and main trends of this development, and the context of their implementation, is important at this stage.

Reflecting over 20 years of these developments, Osborne (2006) considers that NPM might be better considered as a transition phase from traditional public administration to more complex notions of governance<sup>66</sup>, which recognise a more plural approach and the growth of networks. Despite wide consideration that NPM as a concept has been overtaken by the concept of “governance”, Greve considers that the impact of it is being first felt now, but rather than discussion on the idea and focus on decisions, interest is better placed on implementation and results (2006: 23). Christensen and Lægreid (2007) also agree that NPM continues, despite being considered dead and buried, which confirms their reflections that it is neither a “neat package” nor has it had a specific “starting point” or “destination”. Despite the flexibility with which NPM has been applied it has had a profound impact on countries implementing and adopting it.

#### *Reflections over the implementation of NPM in Norway and England*

Christensen and Lægreid (2007: 4-6) outline how scholars are divided as to what influences the implementation of NPM. One aspect relates to environmental determinism<sup>67</sup>, due to ideologically hegemonic external norms diffused through isomorphism or as optimal, technically efficient solutions. Another approach focuses on the influence of “national historical-institutional context”, based on an idea of path dependency relating the success of implementation to a proposed reform to the tradition and underlying values of the system (2007: 5). A further approach focuses more upon how the “constitutional features and political-administrative structures” of a society shape capacity to implement reform (2007: 6). The authors consider that this instrumental view relies upon an assumption of control over decision processes and a degree of rationality, which can be based on structural hierarchy or power and negotiation. Adopting a “transformative approach”, Christensen and Lægreid consider that a combination of these perspectives applies, but also recognising an adaptive, “translation” perspective (2007: 7). These approaches are also reflected and considered further in the organisational decision theories which are outlined in Chapter 5.

Christensen and Lægreid consider that NPM reforms are more compatible within the English system than Norwegian (2007: 5), where the reform process in the latter is thought to be “less harsh and combative” than the “adversarial” former (Pollitt & Bouckaert, 2000: 30). They recognise that “Anglo-Saxon” countries, with a “Westminster-style parliamentary system”, easily adopted NPM reforms due to external economic and institutional pressures combined

---

<sup>66</sup> Which Osborne (2006) refers to as New Public Governance.

<sup>67</sup> Developed from the work of Olsen (1992).

with a lack of structural hindrance and a conducive culture (Christensen and Læg Reid, 2007: 9). Norway, however, was considered more sceptical, with weaker external pressures and an equality based value system and minority controlled party system little used to radical reform (ibid.), making the country a “moderate and reluctant reformer” of NPM (Christensen and Læg Reid, 1999: 9).

Hood outlined how in UK the NPM ‘revolution’ was “led from above” and greater focus was placed on “business-like managerialism” across the public policy sector than was visible in the majority of other countries implementing such reforms (1991: 6). Hood considered that the emergence of NPM was brought on by a combination of factors, the most likely of which concerning the growing shift from a blue to white collar workforce and a mistrust of the all-pervasive state (1991: 7-8). In the UK there was a particular emphasis on “cutting costs” through improved management and structural change, combined “contracting-out, compartmentalizing and top-slicing” (1991: 15-16). These reforms could be implemented to a deeper and wider degree due to the “majoritarian” state structure and executive government (Pollitt & Bouckaert, 2000: 48). While Andresani and Ferlie (2006) agree with Pollitt and Bouckaert about the extent of implementation at the “general level”, they recognize that the HEI sector, not being under direct ministerial control, “has been more difficult to influence” compared with, for example, the Health Service. Nevertheless the authors do see the evidence of managerial speak, control by funding and increased notions of competition (Andresani & Ferlie, 2006: 416-7). As will be seen in the next section, and then dealt with more extensively in the next chapter, evaluation frameworks in HEIs have undergone change as a result of these reforms. The extent to which this has influenced academic staff and subunits will be explored further through the empirical part of this study.

Marquand notes that the post war social democratic consensus in most Western Democracies which placed full employment and mass education at the forefront of their Keynesian Welfare states<sup>68</sup> has now disintegrated, making such a political platform unobtainable, if not unelectable (Marquand, 1997: 1). NPM has therefore been one political reaction, which gained its first footing in UK under Thatcherism, even though Keynesianism had already been abandoned 2 years before Thatcher’s rise to power. The seeds of New Public Management in England were ideological will and demands for governmental reform based on public-choice economics, privatization and generic management (Pollitt & Bouckaert, 2000). Despite being considered to be the programme of the New Right, NPM under the guise of modernisation was continued and reinforced under the New Left policies of the Blair and Brown Governments, where it is noted for example that the first Blair Government introduced performance targeting, resource linked indicators and impact and outcome oriented evaluation

---

<sup>68</sup> Where governments wishing for full employment actively intervene in the economy by taxation, expenditure and monetary policy to stimulate aggregate demand, and where full employment produces ‘inflation’ to reduce aggregate demand (Marshall: 1998: 338).

(Davies, Newcomer, & Soydan, 2006: 169 - 172). Despite efforts under modernisation to “join-up” government and reduce emphasis on competition, focus is still placed upon ascertaining “what works” (Boyne, Kirkpatrick, & Kitchener, 2001; Flynn, 2007). These processes are discussed further in the next section.

While NPM reform has been “far from immune” to spreading as a result of isomorphic processes<sup>69</sup> (Hood & Peters, 2004), according to Christensen and Læg Reid (2007), there have been several waves of NPM reform which have been applied differently across implementing states. As was stated above, as a reluctant reformer, the implementation of NPM was much softer, adapting to a more socially democratic tradition (T. Christensen, Lie, & Læg Reid, 2007: 37) within a wider Nordic tradition of egalitarianism, stakeholder involvement and incremental decision making (Johnsen, Nørreklit, & Vakkuri, 2006: 207). The content of NPM based reform has been different to England. For example, while there is evidence of MBOR<sup>70</sup> and structural devolution, an introduction of performance steering was considered more problematic and there have been fewer of the more radical, contractual reform options experienced elsewhere (T. Christensen et al., 2007: 37).

Christensen et al note that implementation of NPM reforms in Norway has additionally been more fragmentary than in other countries, focused more upon different sectors developing themselves rather than on common changes (2007: 36). This offers an interesting contextual issue. The authors recognise that governance in Norway has traditionally and culturally been based upon “mutual trust”, whereas the mechanisms of NPM, particularly performance management, appear to be more greatly based upon “distrust”. The authors therefore note a complexity of development of these ideas within the Norwegian system. They note that in the 1980s Norway had only adopted the softer parts of NPM, the “least radical reform elements, like MBOR” (2007: 37). The process of NPM adoption in Norway is characterised as one of adaptation to the “historical path that Norway had been following”, which “took a long period of adjustment, translation and modification for the MBOR system to be widely implemented” (2007: 37). Læg Reid, Roness and Rubecksen (2006: 268) could not find a dominant Norwegian model of performance management, with general leeway to which MBOR model to implement, with widespread reporting about process and output rather than any attempt to adjudge outcome. Recent years have witnessed a conservative-centre coalition attempting to implement NPM reforms at a faster rate, followed by a Red-Green, centre-left alliance mixed between cautious approval and scepticism, signalling a return to some form of Weberian, “old public administration” but combined with focus upon “more market, management and efficiency” (2007: 40-1). As a result, Norway currently finds itself implementing “more market, management and efficiency”, while yearning

---

<sup>69</sup> Isomorphism is dealt with in Chapter 5.

<sup>70</sup> Management by objectives and results

for “a return to some of the main features of ‘old public administration’ and a rediscovery of Weberian bureaucracy (Pollitt and Bouckaert, 2004)”, (T. Christensen et al., 2007: 41). These points highlight the underlying values and norms in Norway. Despite the rhetoric of greater efficiency and effectiveness, there is a continual suspicion towards NPM and governance, even though the latter has been more associated with less hierarchy. However, those implemented reforms that have incorporated tendencies towards NPM thinking, like the Higher Education Reform, have included demands that require substantial administrative change and refocus. As will be seen in the next section a key part of this has been evaluative reform. A question though is also raised as to the efficacy and will behind these demands.

### **3.5.1 Evaluation under NPM**

Understanding developments in public policy making is considered important for understanding developments to the frameworks within which evaluation takes place. Wollmann (2003) considers there to be “a ‘Siamese twin’ like connection” between the two. Under NPM this relationship was considered to enter a “third wave”, where evaluation was integrated and institutionalised within organisations, based on measurement-focused, externally defined performance indicators for self-assessments to be reported upwards, that is, monitoring and feedback (2003: 2-3). But as Wollmann relates, impact is still felt from the first and second waves of change, the former “planning period” in the 1960s and 70s, focused upon evaluation to provide output evidence, and period of “retrenchment” in the 1970s and 80s, where evaluation was more focused upon identifying cost and input efficiency and possibilities for cuts and reduction.

According to Hood the concept of “public management” has been understood as “the study and practice of design and operation of arrangements for the provision of public services and executive government” where “management itself is conventionally defined as direction of resources or human effort towards the achievement of desired goals” (2005: 8). Hood, however, recognises that it is the deeper interpretation and implementation of this approach that has caused great debate from both “ideological” disciples and detractors. He also notes that NPM, argued to be rooted variously in Benthamism and Taylorism, has followed the notions of discovering and developing best practice, production engineering, leadership by trained managers and control of activity (2005: 13-18). Evaluation may simply be considered as a functional, money saving device; or an ideological, neo-liberal doctrine (Dahler-Larsen, 2005b: 362 -3). Dahler-Larsen (2007) considers that the ideology of NPM is focused on performance measurement that should contribute to decision making and resource allocation, enhancing rationality and accountability. But as Dahler-Larsen also points out, this ideological development is perhaps more symptomatic of the weakened belief that a welfare state can function the same in the modern era, rather than a sinister development designed to undermine it. The development of evaluation is neither a one-dimensional field, where many of the models exhibit “anti-

establishment attitudes” (Dahler-Larsen, 2005b: 363). Dahler-Larsen (1998, 2005b) sees evaluation through the lenses of reflexive modernisation, coupling evaluation and NPM together, observing that evaluation generally becomes a narrowed as a concept. Reflexive modernisation is drawn from the work of Beck (1997), which exhibits a greater focus upon political and economic management of risk incurred by technology as part of a risk society (G. Marshall, 1998)<sup>71</sup>. The overlap between NPM and evaluation becomes a “result-oriented measurement and monitoring” (2005b: 373). That elements of NPM are well suited to the basis and rationale behind evaluation is not in essence under discussion. However, reiterating the points made earlier, the way evaluation is recognized within a framework of NPM does appear to influence how evaluations will be designed and implemented and for some, evaluation has, itself, become synonymous with NPM (Dahler-Larsen, 2005b). The way decision makers perceive evaluation as an external demand, which as Dahler-Larsen notes is relatively new, appears to shape the form it takes and the purpose of informing. The way organizations respond to these demands is likewise shaped by their decision-making processes. In reacting to the result-orientation of NPM (Dahler-Larsen, 2005b), it is possible that any wider purposes of evaluation are joined to the perception and are set aside. Determining results is, of course, one aspect of evaluation, but not the only one.

Hood also notes a “preoccupation with control” as a major focus of NPM, analysed mainly through institutional economics, striving to develop “output- or outcome-based controls to supplement or replace input- or throughput-based controls” (Hood, 2005: 21). Osborne agrees that one of the key elements of NPM is “an emphasis on inputs and output control and evaluation and upon performance management and audit” (2006: 379). As with the wider auditing movement Power goes on to reflect that “design and operation” of the system is a model of “organizational self-observation... (with) external oversight” (2005: 333). Rossi *et al.* (2004: 13) also recognise that during this time, the control of evaluation has shifted, from the oversight mainly by social researchers, to that of those ‘consuming’ its findings, i.e. policymakers, planners and administrators. Within the wider domain, the public is now also seen as a key stakeholder requiring information about how tax funds are used as well as the overall quality and efficiency of public services in general. As such, evaluation now sits more comfortably alongside policy analysis and public administration (Rossi *et al.*, 2004: 15). These developments, as noted above, have led to polarised support and derision.

Rossi *et al.* (2004) also note that resources available for evaluation are now increasingly more often ring-fenced for certain types and methodologies that produce ‘evidence’ that will support such claims. This influence can be seen

---

<sup>71</sup> Marshall (1998: 558) recognises a sharp division between those considering reflexive modernization a strong critique of post modernization and those who is too highly abstracted and empirically “untestable”.

especially through the focus upon management by objectives and results, performance indicators, focus upon accountability and effectiveness etc., marketization, choice, critical of bureaucratic solutions to public sector organisation (Lægreid, Roness, Røvik, & Christensen, 2004; Osborne & McLaughlin, 2002). Osborne and McLaughlin also draw on the work of Hood, who had noted that the rise of NPM was linked to public spending cuts, privatisation, increased automation and internationalisation of the public service agenda (Hood, 1991). However, the authors take this further noting there to be a 'plural state', which has developed from the 1980s to the current day, where public service provision is 'negotiated' amongst major societal actors, but managed by government (Osborne & McLaughlin, 2002: 10). While recognizing that the promotion of accountability is one of the "general purposes of evaluation" (Davies et al., 2006: 165) within NPM policy it appears rather to be the preeminent purpose. Subsequently accountability is more equated with evidence; in the quest to develop evidence informed policy process which is quality assured and performance is managed. However, NPM reforms have more generally been criticised for being "ideological", "selective" and often "evidence free" (Hood & Peters, 2004) challenging the link to and use of evaluation data in policy making.

More than just being a result of adoption of NPM, evaluation is also claimed to have contributed to NPM survival as well as being conceptually changed itself (Greene, 2002). Evaluation is furthermore seen as central to NPM philosophy, but redefined as a kind of pervasive performance assessment (Van der Meer, 2007). Adopting MacDonald's (1976) conception of bureaucratic evaluation, Norris and Kushner note that increasingly under NPM evaluative activity that supports mandated policies has been "woven into the fabric of public policy" and has become routinized and institutionalised as "internal evaluation and external auditing, inspection and monitoring arrangements and performance management systems" where the results are published (Norris & Kushner, 2007: 6). The authors argue that these developments are tied to the decline of professional autonomy, decentralisation of responsibility and demand for control of efficiency and effectiveness (2007: 7). Under such conditions, while the responsibility for design and implementation of evaluations are transferred to units and subgroups, there is often a pre-specification of frameworks, formats and methodology in addition to type of data required. As was seen in Chapter 2 in England evaluators have had to respond to demands for assessment of programme impact, which has been tied to funding. However, such developments can have the opposite effect. The authors argue that demands for transparency and tying rewards to evaluation can reduce openness and honesty. Dahler-Larsen considers there to be a mismatch between NPM terminology, favouring "transparency, visibility, documentation and measurement", and the experience that evaluation processes should be "long, complex and non-linear" (2007: 18). As will be explored further in Chapter 5, demands for greater transparency do not necessarily lead to more open processes of evaluation and may work against their intention as pressures for accountability lead



organisations into a defensive role (Dahler-Larsen, 2007, 2008). A question is thus raised as to whether the aforementioned goals of NPM are achieved or whether there is another impact of such behaviour. I return to this in the section below regarding evaluation utilisation.

Norris and Kushner argue that under NPM it is predominantly at the executive level where decisions about definitions of quality and performance standards are taken (Norris & Kushner, 2007: 12). The authors agree with Dahler-Larsen (2007) that these decisions are constitutive, emphasising certain standards at the expense of others and challenging professional autonomy (ibid). The authors highlight an asymmetric power relationship, which can work to the detriment of the principle's intention. Norris and Kushner take an example of a HEI responding to these bureaucratic demands, which can "create social solidarity" as professionals unite against, but equally produces greater competition between organisations rather than cooperation (Norris & Kushner, 2007: 12). Fitz-Gibbon reflects that there has been a development towards a "totally planned managerial society in which procedures are dictated and prescribed by hierarchical systems", an "authoritarian managerialism" (2002: 141). In such systems targets are set and blame apportioned. She sees this approach based on control of the future and lack of trust of those involved in processes under investigation, contrasting with Popper's perception of "participative, democratic, organic, interacting systems... acknowledging that the future is unpredictable" (ibid.). Thus focus is placed on the quality of indicators in the system, how they are made and what processes they intend to investigate. Fitz-Gibbon therefore calls for organisation and rationale in evaluation process rather than a rejection of this method of working. This of course raises important points. There may be fundamental disagreement over the way that data can be collected, analysed and used, and the premise on which particular processes are based. This of course is not a revolutionary argument, but it means that one must look beyond the idea that programme groups do not desire evaluation per se. Groups may reject a particular demand to evaluate based on the current model to be implemented as well as their previous experience of evaluation processes. As we will see in Chapter 5, it is argued that the very act of designing and implementing an evaluation can affect the object that is under study and the nature of a programme, before the data is even collected. Dahler- Larsen (2004b, 2006b, 2007) refers to this process that is beyond the main ideas of 'use' as the constitutive effects of evaluation. I return to this in more detail in Chapter 5, relating to institutional models of decision making in organisations.

Within the developing nexus of public service provision a plethora of groups deliver the policy reform schedules. Amongst many impacts from the changes noted above, the interplay between public and private forms of organisation becomes blurred. As reforms are implemented little is known about the impact and effectiveness of the various processes that are enacted (Lægheid et al., 2004). This leads to, what Lægheid *et al* refer to as, an 'evaluation paradox', where the extent of knowledge about how goals are achieved is detrimentally

affected by the level of resources ascribed to assessment activities. In other words, promises made about effective programmes can rarely be demonstrated as being fulfilled, because they are rarely checked. The authors claim that evaluation thus becomes a mantra. Lægneid *et al.*<sup>72</sup> (2004: 151 ff) raise questions to the organisational ability for rational calculation, within their focus upon the public administration system. Aside from the ability to look ahead and assume “different courses of action and organisational forms, through planning and *consequence analysis*”<sup>73</sup>, the authors note that organisations must also be able to focus ‘backwards’, noting how skilled they are and what they have achieved via their various attempts at altering form and activity. This work ultimately connects the field of evaluation to organisational theory upon decision-making.

As seen in the previous section, NPM has emphasized streamlining across the public sector. The focal point of evaluation has shifted to the level of delivery but with greater commissioner power, creating a “new role as a steering instrument...and as a tool for consumer quality and control” (Karlsson, 2003a: 135). When decentralizing there appears to be a logic of consequence that programmes will be evaluated, which to some will appear to resemble result based surveillance (Furubo & Sandahl, 2002 in, Dahler-Larsen, 2005b). The implications of this section are interesting in the wider picture of evaluation activity. Organisations are under greater pressure to evaluate their processes and outcomes, a demand that requires greater competence within the organisation to perform such a task. As evaluation becomes more mainstreamed and is brought to the forefront of organisational thinking and agency there is an increasing likelihood that evaluation will be absorbed and adapted into organisational relevance structures, incrementally transforming the understanding of evaluation. This becomes even more reticent when organisations interact, especially when one organisation evaluates another. The implication appears to be that without any concrete external reference points for evaluation, the influence of the organisational background, processes and routines will play a greater role in the perceptions and practices of any evaluative activity that becomes standardised. At the same time organisations may influence one another as their evaluative activity intermingles. This may account for the increasing likelihood for organisations to focus on one another’s structures and systems rather than a more direct outcome measure, should that be feasible. Evaluative systems do not seem, though, to have provided the decision making information envisaged and anticipated by commissioners; neither under NPM nor Governance approaches. In addition to the complexities raised above there is also an issue concerning the utilisation of evaluation (Stame, 2006). Focus is therefore required on perceptions of how evaluation utilisation is understood; this is addressed in the next section.

---

<sup>72</sup> My translation from Norwegian.

<sup>73</sup> My italics

### **3.6 Evaluation utilization**

This section will describe how debates over the purpose of evaluation and subsequent models chosen have been driven by the underlying intention to utilise. Seen against greater demands for accountability and evidence of effects, training and development programmes in the public arena must face the problem of what information will be gathered, presented, defended and acted upon in terms of their particular strategy. This demands a discussion of how evaluation utilisation is observed and how organisations act in order to be accountable. While the eventual use of evaluation data is not investigated, the perception of why it is requested and the prospect of its use are thought to illuminate certain decisions made when an evaluation is implemented. While there are many functions of evaluation, utilisation has by many conceivably been redefined as the basic premise of it. Utilisation does of course operate on different levels in an organisation as well as the variation between internal and external uses. The implicit and explicit demand for evaluation outlined above is often understood in terms of how evaluation in turn is understood and used. An assumption in this study is that the intention and design of evaluation are affected by external demands and future purpose of use of findings. How members of organisations perceive, deliberate and ultimately operationalize these demands through their decision making process are considered worthy of study and will form part of the focus of the empirical part of this work. At the same time it has been recognised that utilisation of evaluation data is a slow process, and where not resisted, it occurs more cognitively than instrumentally (Stame, 2006: 7).

#### **3.6.1 The ideology of evaluation utilization**

It is suggested that whatever use takes place will be defined by ideological underpinnings concerning the purpose of evaluation. At the same time a 'frequent failure' to use the "conclusions of evaluation research in setting future directions for action programs" has been noted, which appears to challenge its basic intentions and rationale (Weiss, 1972: 318). Although by-functions of evaluation are found, such as testing theory or building up knowledge bases, unless it becomes part of the decision-making process, it is said to lose its purpose. Weiss later noted that evaluation is often a hierarchal decision, originating from the top and trickling down throughout a system. This had major implications for the type of decision-making under question.

The apparent lack of application of decision-making theory within the evaluation field beyond that of relatively prescriptive approaches is therefore interesting. Rogers and Hough (1995) recognise that studies of evaluation have rarely been linked to "an articulated theory of how organisations work" and often assume that organisations act rationally. Basic definitions of evaluation are observed to promote a rational choice perspective (Holton III & Naquin, 2005; Russ-Eft & Preskill, 2001). In particular, evaluation theory often assumes a linear, hierarchical structure within a 'hypotheticodeductive paradigm' rather than a complex organisation with "multiple sources of power and influence",

where the institutional context and factors shape programme choices (M. W. McLaughlin, 1985: 116), often exhibiting “counter pressures” to evaluation findings (Weiss, 1972). This contextual boundedness of evaluation will impact particularly on internal evaluators (Sonnichsen, 2000: 61). Focus should therefore be placed on how the evaluation field attempts to describe best or actual practice, and whether models aim to recommend or prescribe (Scriven, 1991). Understanding of this will help describe the purpose of a particular evaluation model and type of use that is intended for it.

### 3.6.2 Types of evaluation use

Weiss (1998b) records that the concept of evaluation use has deepened over time, changing from a pure reflection of utilization of results in subsequent programme decision making, to the wider impact of evaluations upon multiple and diverse users. The initial focus of evaluation research was on *instrumental* use, where findings are used to change parts of a programme not appearing to function and *conceptual* use, where a more holistic change of thinking takes place about broader programme aims (Alkin & Taut, 2003: 5). These forms of use are considered to be rational approaches. However, it soon became necessary to account for other uses of evaluation’ for example, Alkin and Taut attempted to go further again, seeing *use* as split into two major categories, *process* use and *findings* use, where process use is not another category “but rather another domain of use”, which may in turn “occur instrumentally or conceptually” (2003: 6), as a behavioural and cognitive change resulting from participation in an evaluation (Johnson, 1998).

Drawing on the work of Greene (1988), Owen and Rogers (1999) and Russ-Eft *et al.* (2002), Alkin and Taut also forged a distinction between *legitimative* and *symbolic* use of evaluation, where the former applies to the persuasive use of results and the latter, a legitimisation by action but without any regard for the results obtained, which may be used for political self-interest (Johnson, 1998).

Research has also sought to distinguish between use and influence (Kirkhart, 2000). Kirkhart’s attempt to integrate influence as an expanded notion of the too limiting concept of *use*, focused on three variables: source, influence and time (Caracelli, 2000). These variables present a much wider application of evaluation findings. Alkin and Taut (2003) regard ‘influence’ as an impulse that takes place within a ‘process – results dimension’, but ultimately lays beyond the control of the evaluator. Henry and Mark (2003) also agree that there should be a movement towards influence, refocusing upon evaluation as a ‘continuous process’ rather than ‘episodic impact’ (Cummings, 2002). This comment is significant moving the debate about evaluation use even further beyond a pure focus on the rational process and opening for research into ‘unintended influence’. Here the user is influenced in a more unconscious way and therefore greater respect needs to be made of this process when evaluating programmes. It would appear to add weight to part of Cousins and Leithwood’s (1986) meta-analysis that evaluation use appears strongest when all users are involved and

committed to the findings and the process is deemed to be appropriate and credible (see also Hofstetter & Alkin, 2003).

Weiss (1998b; 2005) outlined 'six' kinds of use that have been recognised in evaluation research. These are summarised in table 21, which is found in the appendix.

Within the evaluation process there are many phases, which can be initiated both externally and internally and switch back and forth between the different levels (Dahler-Larsen, 2004b). Dahler-Larsen concurs that use is not merely a function of an evaluation's aim or organisation, but rather a particular complexity. Concepts of use are problematic when one investigates use beyond what is promised or planned (Dahler-Larsen, 2007). This occurs because the concept of use is not well defined, which partly stems from a "tacit normative framework", where the question remains can one define what is good and bad use and what is misuse (Dahler-Larsen, 2007: 20-21)? The focus is therefore placed on the development of the evaluation model rather than its utilisation. "Use" appears to have been interpreted within "restricted assumptions" about knowledge, rather than, for example, a reflexive approach (Dahler-Larsen, 2007: 24). But use is an important focus, as attitudes to evaluation are noted to be formed as a consequence of the experience of how findings are actually utilised (Dahler-Larsen, 2006b: 85). This is in addition to, but develops from, understanding members' perspective upon evaluation. Therefore programme group members will be asked to reflect over utilisation in their group and wider organisation to illuminate this point further.

Researchers are therefore encouraged to map the phases of an evaluation to uncover the various purposes, roles and responses of those enacting, implementing, reporting and using evaluations. The picture is necessarily complex. Internal evaluation will not necessarily mean freedom from external influence in the same way that external control will not guarantee instrumental or conceptual use of evaluation findings (Dahler-Larsen, 2004b: 12 ff). Dahler-Larsen sees a point of consensus between control and learning focused utilization, that both see use as tied to original purpose or aim, "if one is an adherent of control all one sees is control. If one is an adherent of learning all one sees is learning"<sup>74</sup> (2004b: 15). This necessity to state the purpose of the evaluation from point one is a noticeably rational approach, believing that purpose will drive use. Dahler-Larsen refers therefore on Weiss's form of enlightenment use<sup>75</sup>, forming from knowledge creep whereby over time and space attitudes to evaluation change and are reformed as experiences are interrelated, leading to more varied interpretation and use of results. Dahler-Larsen outlines 7 overlapping types of use, and also opens for likelihood that there will different uses made of the same evaluation data (2004a: 39-40).

---

<sup>74</sup> My translation from Danish.

<sup>75</sup> The citation is not specified by Dahler-Larsen.

Of particular interest is Dahler-Larsen's reflection over constitutive type of use, which treats evaluation as a marker for interpretation, where the criteria, boundaries and values for quality of the evaluand's performance are set out (2004b: 16). Such a framework recognizes that evaluations operate beyond the formal measurements and considerations, which rather become "recipes for interpretation", especially as current behaviour becomes influenced by future evaluations. This implies that the content of organizational activity as well as the time frame for action is affected by the forthcoming evaluation. As long term effects are harder to document, organizations (in this case used by Dahler-Larsen, schools) might well adjust activity to accommodate the more measurable short term effects or that which matches the evaluation calendar.

This last point raises important questions for this study. All of the above types of use are interesting, but what are the causes whereby an organization focuses their evaluation or response to the demand to evaluate in the particular way that they do? This involves greater focus on the decision-making behaviour and procedures of organisations. Such a view appears to be echoed in Dahler-Larsen's reflections that "[e]valuations can start chains of interpretations and actions, which turn evaluation into much more than a planned activity. One takes an important first step when one recognises this, seeing evaluation not just as a descriptive activity but also as a creative one" (2004b: 17). The point here is that under NPM evaluation demands have changed significantly but one can question whether models and processes have changed. In particular discussion concerning these approaches is taken up further in Chapter 5.

### **3.6.3 Factors thought to influence or affect evaluation use**

Such diversity of use departing from more rational intentions led the evaluation research field to consider whether factors could be discovered that would explain or influence such behaviour. Review of evaluation use in American business and industry revealed that "about half of ... training programmes are evaluated for objective performance outcomes" and showed little correlation between provider skills and experience and the extent of evaluation utilization (Holton III & Naquin, 2005: 258). These findings create some areas of immediate concern for this study, particularly vis-à-vis the process of evaluation. Alkin (2004; Alkin & Christie, 2005) notes a significant split in the evaluation field, where theorists have tried to ensure greater evaluation utilization by attempting to improve the quality of information by adopting better *methods* (e.g.: Weiss, 1972; 1982, 1998b)<sup>76</sup> or increasing the *involvement* of stakeholders and users (e.g.: Cousins & Earl, 1995; Patton, 2003)<sup>77</sup>. Cousins' early work would recognise that *participatory* models of evaluation appear to improve use of results rather than "conventional stakeholder-based evaluation" (in Alkin, 2004: 325). Later, however, he recognised that responsiveness to the "context in the creation of knowledge and meaning" was important when

---

<sup>76</sup> Often noted at the macro decision orientation level (Sverdrup, 2002)

<sup>77</sup> Often believed to influence at the micro user orientation level (Sverdrup, 2002)

advocating participatory approaches. Context and culture are perceived as important within this attempt to improve evaluation use, and reflect research into organisational change and leadership function, and a supposed “organisational readiness for evaluation” (after Seiden, 2000 in Alkin, 2004: 328). That being so, Cousins’ and others’ approaches may still underplay the role of stored knowledge and standards existing within the individual and their organisation before any evaluation process is enacted. I return to this point in Chapter 5. In addition, Dahler-Larsen outlines how the developing evaluation culture focuses less on the inclusion of stakeholders, and more upon the unit or group under question becoming self-sustaining through evaluation (Dahler-Larsen, 2006b: 91). The author considers that understanding how competence is built up in the organisation will be important, given that processes are now often internalised. Focus will therefore also be required on how organisations attempt to improve the competence of those performing evaluation.

Hofstetter and Alkin (2003) note the difficulty in isolating these factors, especially from their political and organisational contexts. However, factors were identified that related to the “purview of the evaluator, the evaluator’s approach to the evaluation, and selection of users” (*Ibid.*: 213). The authors draw on the meta-analysis of seven research studies and their findings are summarised in table 22, found in the appendix. Although many individual factors have been isolated, the most interesting research built higher order categories against which utilization could be analysed. Cousins and Leithwood’s (1986) widely cited meta-analysis<sup>78</sup> recognised use to include “support for discrete decisions... and the education of decision-makers” but also more basically as “the mere psychological processing of evaluation results” (1986: 332). Such an approach is interesting for this study, in agreement with Weiss (1998a) that *intention* for use in decision models will be one of the primary motives for which models are chosen and how they are applied. Weiss noted that evaluations often seemed to compare ‘what is’ with ‘what should be’ (*Ibid.*).

Cousins and Leithwood (1986) outlined 12 factors that influence use, divided equally between two higher order categories of *evaluation implementation*<sup>79</sup> and the *decision / policy setting*<sup>80</sup>. The authors recognised the importance of developing evaluation procedures that would generate information helpful for decision-making, and that potential users should be involved in a manageable way in the planning and implementation of the process, where the majority of the 12 factors that they isolated would be present (1986: 360). However, at a

---

<sup>78</sup> This was based on empirical research into the use of evaluation results in the period from 1971 –1986. When isolating the factors influencing use, Cousins and Leithwood assume the evaluator’s role is to enact the evaluation and report the findings to a decision maker. The majority of these studies, however, were ‘retrospective’ focusing upon ‘anecdotal accounts’ of key actors (Cousins & Leithwood, 1986: 333).

<sup>79</sup> Quality, credibility; relevance; communication quality; findings; and timeliness

<sup>80</sup> Information needs; decision characteristics; political climate; competing information; personal characteristics; and commitment/ receptiveness to evaluation.

more overarching level evaluators often ‘generated unrealistic assumptions’ about the value of their findings. Drawing on Lindblom and Cohen’s (1979) view, *ordinary knowledge* appeared to be the basis for decision making and action in the majority of organisations. As a result, Cousins and Leithwood recognised that isolating the factors that influence the use of data gained by evaluation to be of key importance for enlightening more of the decision processes at hand. Such an action faces serious methodological challenges in line with the current quality of evaluation practice (Weiss, 1998a). This requires a more detailed study than is often present in more informal programme evaluations. Weiss reflects that “[e]valuators expect people in authority to use evaluation results to take wise action” (1998a: 5). Weiss also notes that the best-informed people are those running the program, but these groups “tend towards optimism and... have a stake in reporting success” (1998a: 6). These factors strongly influence the content of the findings of the programme under investigation and create challenges to their validity. Thus, the link between evaluation use and decision-making still remains unclear. Cousins and Leithwood recognised the need for a wider framework of understanding, and it is to this area that this study aims to make a contribution.

Generally, therefore, the assumption has been that the greater the rigour of the evaluation process and greater proximity of stakeholders to the evaluation process, the more likely the findings are to be used. Unfortunately empirical findings have not born out this rationally grounded supposition. When attempting to discover the causes of such different levels of use, Rich considers investigation of “routine bureaucratic and organizational roles” to be important, where utilization of evaluation information is a function of organisational decision-making “independent of the manner in which an evaluator produced and delivered the information to the organization” (Shapiro, 1984: 634). The suggestion is that these processes are understudied in relation to the application and enactment of evaluation models. While this is a key area of investigation, this study claims that the intention for utilization is also a valid area for research. Evaluation theory has often focused upon improving the operationalisation of the process, but decision-making research points further to complexity within the organisational environment<sup>81</sup>. It is intended to investigate decision makers’ attitudes and actions in formulating evaluations of their programmes.

### **3.7 Summary of perspectives on evaluation design, implementation and utilization**

When drawing together research from the evaluation field, including questions of purpose, context, quality of process and type of utilization, we are left with a

---

<sup>81</sup> The latter view, according to House (2006) has been reflected strongly in Scandinavian approaches by the work, amongst others, of Karlsson (2003b, 2003c), Vedung (1994, 1997, 1998, 2000) and Monsen (Haug & Monsen, 2002; Monsen & Haug, 2004), and these different emphases also appear to reflect over whether the focus is macro (policy) or micro level of programme evaluation.

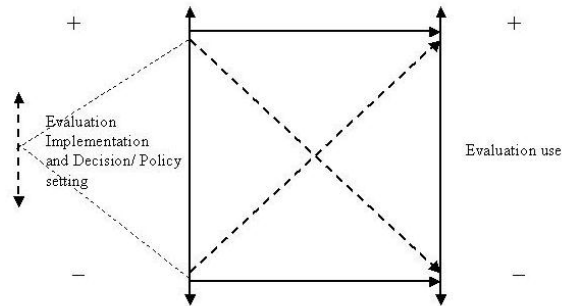


dichotomy. It is often assumed that quality evaluations, i.e. strong in methods and relevant to their context or setting, will more likely be utilized in further decision-making. However, research has shown that this is not necessarily the case, and in fact the opposite may be the case. Figure 4 below symbolises this dichotomy, recognising that future utilization of results might be dependent upon variation in the initial decision making processes focused on the adoption of evaluation models. In the section on evaluation models it was recognised that Stufflebeam (1983) had recognised the importance of assessing attitudes amongst evaluators to the models they adopt and implement. Stufflebeam (2001) also elaborated these ideas in his “meta-metaevaluation” (Henry, 2001) of evaluation approaches and models, recognising the development of around 22 different approaches of which 9 were thought to be pervasive<sup>82</sup>. Despite any disagreement over the efficacy of such models, Henry recognises what appear to be two key points for this study in Stufflebeam’s analysis. Firstly, and following on from the point made above, Stufflebeam “presumes” that models should be systematically evaluated with regard to their ability to assess a programme’s *merit and worth*. Henry is recognising a decision based process underlying Stufflebeam’s work. The second point follows on somewhat tautologically from this argument. Stufflebeam is arguing that the models are based on an assessment of value and worth for the task at hand. His choice off 22 approaches, as Henry (2001: 3) recognises, already omits certain models that do not fit easily under the categories presented. These important points appear to strengthen the argument that investigating decisions about models is an important task. These ideas, as well as further analysis of Stufflebeam’s approach to decision making within evaluation are taken up in Chapter 5.

Current concepts of evaluation use and purpose have also been challenged. One of the most interesting reflections ties evaluation to institutional theory which is an open systems perspective on organisation theory, focusing on the wider environmental context that “constrains, shapes and penetrates the organization”, particularly from a social and cultural perspective (Scott, 1995: xiv). I return to this in Chapter 5 along with a more in depth study of different models thought appropriate to investigate decision making of evaluation. Therefore, focus on the decision-making processes of programme providers, symbolised by the broken arrow to the left of figure 4, is considered to be importance. While this study does not attend to evaluation use as such, it is still considered important as an outcome that is a purpose of the evaluative process.

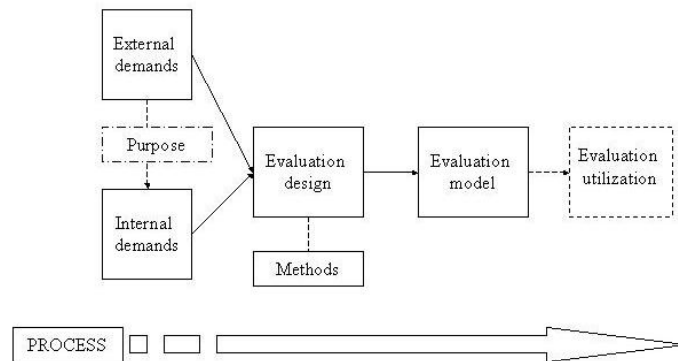
---

<sup>82</sup> Client-centred, Utilization-focused, Decision/accountability, Consumer-oriented, Constructivist, Case Study, Outcome/Value Added Assessment, Accreditation and Deliberate Democratic (Stufflebeam, 2001: 7).



**Figure 4: The problematic area of utilisation**

But a question is raised as to what happens when organisational choices about evaluation are limited or external demands to produce certain types of evidence cannot be met or fulfilled? The process of decision-making about how to evaluate programmes under investigation thus comes under question. This focus is outlined in figure 5 below.



**Figure 5: The process under study**

It is, of course, recognised that there are many more variables that influence than those outlined above, which is framed within an historical context influenced, amongst other things, by previous evaluations, organisational traditions and individual preferences. Rather than being a causal model, the figure is assumed to illuminate how the process of decision choice is important. This is thought to vary from organisation to organisation. It is that variation that might ultimately help further develop the understanding of what else influences utilisation. Although the process of utilisation is not under question here, illuminating its influence on the decision-making process is thought to be important. Therefore the box around utilisation is marked by dots. In this study it is the perception of future utilisation that is of interest and how that might affect decisions about evaluation. Linked to a greater focus on accountability for outcomes is the increased expectation that evaluation results will be utilized (Patton, 1997;

Weiss, 1998a) or at least ‘influence’ new activity (Kirkhart, 2000). Increasingly, the level to which mandated programmes are evaluated and the extent to which results are valid comes under question, even though this is not generally a new idea (Easterby-Smith, 1994; Guskey, 2000; Hamblin, 1974; Kirkpatrick, 1971, 1998). Reflecting increased demands in society for accountability, evaluation as a discipline has also adopted a ‘scientific authority’ (House, 1993: viii). House also recognises that the evaluation process has developed into a formal ‘cultural authority’ with strongly recognised political effects. It might also be perceived that specific utilisation is different from general utilisation intention. For example, organisations may well intend to utilise but be hampered by the organisational logic that guides them. Therefore, while I do not now negate the importance of evaluation utilisation, I consider that decision processes will help us gain a stronger understanding of how and why certain models will be employed.

As was noted above in section 3.6, understanding the national context is important, even when researching at the micro level. Schwandt’s research (in Dahler-Larsen, 2005b: 365) observed a Scandinavian approach to evaluation as steeped in a positive attitude towards the welfare state combined with focus upon equality and solidarity within a collectivist approach to problem-solving and policy, which contrasts with the more “logical-empiricist” Anglo-American tradition. This contributes to understanding that evaluation is less likely merely perceived as a “technical-methodological activity” but rather within a particular ideological and philosophical tradition (Dahler-Larsen, 2005b: 366). Dahler-Larsen (2005b) also notes that across Scandinavia and Europe more widely, there has been a varied understanding of the term evaluation. At the same time, the author notes that the introduction of NPM from the late 1980s has seen evaluation more markedly conceptualized as based on ascertaining success in relation to goals, results and effectiveness.

This study is therefore concerned with attempting to ascertain the demands, designs and decision makers from which an evaluation is implemented from. But understanding these better will require an understanding of the decisions made considering these areas. Therefore, the reflections from and questions raised by this chapter are thought best illuminated by an analysis of organisational decision-making. This will lead to a framework for understanding organisational decision-making, which is outlined in Chapter 5. Before that I turn in the next chapter to consider more closely the context and systems surrounding the sub-units under investigation in terms of understanding the quality assurance frameworks that HEIs are to respond to.

#### **4. Quality assurance and programme evaluation in higher education: Norway and England**

The context of the empirical investigations in the thesis, namely postgraduate programmes for school leadership development, falls within the domain of higher education. A major impact upon evaluative activity, which has advanced strongly in recent years, regards quality assurance. As the emphasis in this study is upon the decisions made concerning the evaluation model that will be implemented it is considered important to link the concept of quality assurance to research on evaluation. Discussion, though, will be delimited to consideration of the basic premises and purposes of the systems and their considered impact at the micro level of decision making within HEIs. Additionally this involves studying the demands placed on decision makers, discovering parties involved in the process and designs that are ultimately chosen. The background concerning the introduction of quality assurance in Norway and England is therefore considered important.

##### **4.1 Introducing quality assurance to higher education (HE)**

The idea of formal quality assurance systems has in recent times been introduced from external systems into the HE system. The concept of quality assurance (QA) developed strongly in the post Second World War period. Scriven (1991: 295) describes quality assurance as basically “evaluative monitoring”, noting it to be on the whole of internal and formative nature. The process of quality assurance<sup>83</sup> is considered to consist of identification of characteristics or “qualities”, defining standards or “design qualities” and monitoring the performance or “actual quality” (Blackmur, 2007: 16). Patton (2002: 147) notes that although programme evaluation and quality assurance developed from different roots and as “separate functions”, they have in recent years overlapped to the degree that “both functions can now be built on a single, comprehensive program information system”. Patton recognises that the quality movement, developing from the work of Deming and Juran from the 1940s onwards, has mainly built on the concept that quality is “meeting or exceeding customer expectations” (2002: 146ff). Vedung also agrees that quality assurance is part of one of the many forms evaluation has developed into, particularly since the 1990s, and sees it as part of the introduction of wider management “doctrines” (2006b: 105). Although Dahler-Larsen (2004b, 2006a) recognises linkages to evaluation, the author conversely appears to see quality assurance as something qualitatively separate from it, noting a shift in emphasis and interpretation of quality. Scriven, however, doubts the efficacy of such processes, unless tied to external field evaluation and therefore part of a wider system of utilisation (1991: 296). This appears to be a main feature of quality assurance systems within higher education today.

---

<sup>83</sup> Although arguing that many interpretations exist, Blackmur (2007) prefers to use the concept in a plural form referring to the idea of “qualities”.

The reform of European universities has focused upon changing the internal system of governance and operational structure from academic-community based discourse to a more generic organisational form of management focused strategic planning, resource control and administration where the predominant indicators of success are economic (Olsen & Maassen, 2007: 7). Within this wide reform of higher education the “rise and spread of quality assurance” is considered in particular to exhibit a “strong European element” (Gornitzka, Maassen, Olsen, & Stensaker, 2007: 203). Significantly, the authors note that QA reform has seen a more marked shift from institutional tradition in the formalization of standards and their movement from “the academic arena into an administrative or political-administrative sphere” (ibid.). This is most visible in the role of the assessment and accreditation agencies, who define, produce and apply the standards. Challenges, or “domain contestations”, have followed such that these processes have remained at the national level rather than being lifted to the supranational European level (ibid.). The authors argue that the study of quality assurance processes offers a good backdrop for understanding linkage between policy and practice, with little known about how standards are applied across the levels. Findings from this current research may also go some way to helping explore an area of importance raised as what “institutional conditions work as filters for or insulation against the penetration of European standards in local practices” (Gornitzka et al., 2007: 204).

A general model of QA in HE is considered to consist of a national coordinating body, institutional self-evaluation, external evaluation by academic peers and published reports (van Vught and Westerheijden (1993) in Brennan & Shah, 2000: 11). While the general model of QA appears robust, Brennan and Shah agree with van Vught and Westerheijden (1993) that there is greater complexity in terms of employment, especially with regard to emphasis, practice, level of investigation, frequency and methods used (Brennan & Shah, 2000: 11ff). An example of these weaknesses is the student questionnaires that have increasingly been seen as the basis of most quality initiatives, despite little understanding of their impact upon quality improvement and the implications of their use in support of decision making (Westerheijden, Hulpiu, & Waeytens, 2007: 305). It is also unclear how academic staff generally consider quality initiatives, whether accepting pragmatically, rejecting them as pointless ritual or with general mistrust (Nasser & Fresco, 2002; Newton, 2000; Westerheijden, Hulpiu, et al., 2007); Nasser and Fresco particularly noting a wide discrepancy in attitudes. It would also appear to be unclear how this affects future design, Westerheijden et al. referring to research into factors like mistrust of instruments and data and difficulties with subsequent interpretation due to capacity weaknesses or disagreements over perceptions of activities (2007: 306-7). Nevertheless, the authors reflect that as one moves from system level to “chalk-face” the focus of QA tends to shift from accountability to improvement (2007: 308). The fundamental effectiveness of QA is however challenged. Westerheijden et al (2007) agree with Nasser and Fresco (2002) and Newton

(2000) that the impact of QA on processes at the level of teaching needs further investigation.

Musselin (2002) recognises a convergence in policy across Europe with regard to higher education towards objectives of increasing the influence of the market and the wider society. Musselin notes that this has led to a changing role for academics in terms of management and decision making, alongside increasing demands for accountability and quality assurance. Musselin also notes that while the intention of the various policy changes has been the same, different areas for change have been focused upon depending upon the context. Further, in both England and Norway these developments defined new legislation that impacted HEIs and the formation of evaluation agencies and systems for quality assurance (2002: 1). Some of the wider implications that have been furthered from HE research into decision processes are that the impact of external reforms with regard to marketisation and increased accountability have been limited, resulting in “tensions and contradictions”. The latter, suggests Musselin, appears due to derive from an already existing difficulty, where the new expectations have “exacerbated already existing organisational inconsistency”, increasing their visibility (2002: 5-6). The intermingling of roles and tools for making decisions creates tension for academics.

*“Autonomy with accountability”*

The Bologna declaration of 1999 was followed by what came to be known as the Bologna Process, uniting 46 countries to form the European Higher Education Area (EHEA), part of this process including cooperation in Quality Assurance initiatives and developments (Gvaramadze, 2008). Westerheijden argues that the Bologna Declaration was initially undergirded by national intentions to change domestic HE policy, but later developed to produce convergence in policy across borders (2007: 77), also considered to have initially developed as a form of “sector defence” (Gornitzka & Olsen, 2006). Gvaramadze considers that there is development from quality assurance to enhancement, affording greater autonomy to HEIs, involvement of students and focus upon continual development. The author recognises the importance of the Salamanca Declaration of 2001 by the European Universities Association (EUA). The fifth of six “action areas” in the concluding statement, that were originally outlined in the Bologna declaration, dealt with quality assurance and accreditation, especially at the transnational level (European Universities Association, 2001). The declaration outlined common European values of quality, in which quality was considered as “a range of academic values... to meet stakeholders’ expectations” (Gvaramadze, 2008: 444). The Berlin Communiqué of 2003, from the meeting of signatory Ministers responsible for higher education, later highlighted the responsibility of individual institutions for QA within their national framework, followed by the Bergen Communiqué two years later which cemented the necessity for HEIs to develop “systematic internal mechanisms” for a culture of quality enhancement. Reichert (2007: 6-7) notes that the

adoption of the European Quality Assurance Standards and Guidelines (ESG)<sup>84</sup> in 2005 created a strong framework for European QA initiatives. The standards confirmed that QA responsibility is at the institutional level and demonstration of “robustness” should lead to limited external control. However, rather than highlighting formal QA procedures, HEIs should develop a “quality culture” of iterative processes. Additionally, the standards confirmed a shift towards the interests of students and stakeholders, both in terms of quality of product received and involvement in improvement of processes. Standards and guidelines for QA were drawn up by the European Association for Quality Assurance in Higher Education (ENQA) (2005) as part of the development of the European Higher Education Area (EHEA), to which all Norwegian and English HEIs are de facto members as their Ministry of Education is a signatory. The ENQA’s report recognises that institutions themselves are responsible to develop policies and procedures of QA that will, amongst other things, reflect the generic principles of the EHEA rather than specific requirements. Focus is placed on what should be done rather than how it is to be achieved (2005: 10). These principles should, however, lead to QA systems that meet the interests of stakeholder and society, which strongly includes the views of students. The EHEA recognises the importance of institutional autonomy but declares the “heavy responsibility” that comes with it. These policies and procedures will be linked to wider cooperation where “realisation of the EHEA depends crucially on a commitment at all levels of an institution to ensuring that its programmes have clear and explicit intended outcomes” (2005: 16). A result of this is that QA systems should focus on fitness for purpose and be limited in burden, providing they satisfy the wider principles. With regard to academic programmes offered, the onus is further placed upon institutions to ensure approval, monitoring and periodic review by clearly developing and declaring “explicit intended learning outcomes” and “design and content” with quality assurance in mind (2005: 16). Evaluation of programmes, reporting to the institution as mandator and relationship to national government need to be seen in relation to these developments.

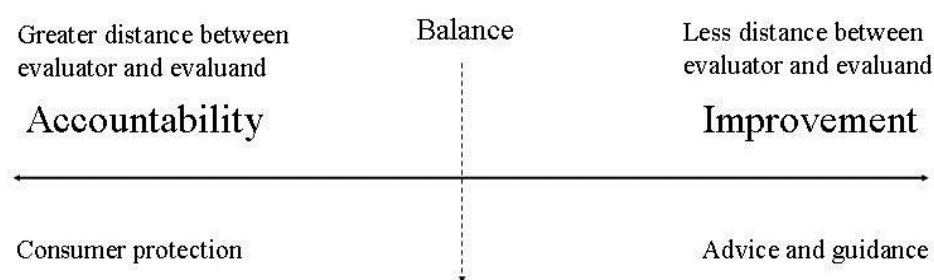
The ENQA report (2005: 11) notes that as quality assurance of programme provision<sup>85</sup> may traditionally have been understood differently from land to land so can consideration of what relationship should exist between HEIs and their external evaluator. A distinction is drawn on a continuum between an accountability focus and improvement focus. This is expressed in the figure below summarising the focus of the ENQA report. There are however tensions in the relationship. Focus on consumer protection leads to the establishment of a “clear distance” between quality assurance agency and HEI, whereas focus on improvement requires by nature a close relationship. The report also recognises the possibility of diverging opinion between the interests of the institution and that of the student body, where the former seeks greater autonomy and less

---

<sup>84</sup> Which had been developed by ENQA, EUA, ESIB, EURASHE

<sup>85</sup> Research and management are not included in these standards and guidelines.

regulation and the latter greater transparency and accountability. From these points it there is likely tension for the academics designing and presenting programmes.



**Figure 6: Balancing evaluation on the continuum of accountability and improvement**

The redevelopment of interest in higher education quality is considered to have been focused on two main issues: the relevance of learning and outputs for a “changing economy” and the efficacy in spending tax-payers money (Westerheijden, Stensaker, & Rosa, 2007: 4). The issue of the tension between accountability and improvement has mainly focused on the former and approached the issue from the perspective of the external accreditation bodies (Danø & Stensaker, 2007; Vroeijenstijn, 1995; Westerheijden, Stensaker, et al., 2007) but also some limited research has focused at the micro level on academics (Newton, 2000) considered to be more improvement centred (Westerheijden, Stensaker, et al., 2007). Vroeijenstijn recognised early on that attempting to balance improvement and accountability, particularly with regard to external quality assurance, was tantamount to sailing between the “Scylla” and “Charybdis”<sup>86</sup>, focusing primarily on how the former would drive stakeholders to develop their own independent monitoring systems, whilst focus on the latter would make any form of improvement unfeasible (1995: 33). From the beginning of discussions regarding quality assurance in European universities, Vroeijenstijn suggested that reconciling these factors would create a challenge due to the differing perception of external demands in HEIs and the expectations of Governments as stakeholders, where the former have generally resisted forms for external control despite a sense of increased autonomy.

<sup>86</sup> Choosing between the “Scylla” and the “Charybdis” might be better understood by the expression “between a rock and a hard place”. The phrase refers to two sea monsters situated on either side of the Straits of Messina such that avoiding one would cause sailors to fall foul of the other. In Homer’s *Odyssey*, Odysseus was forced to choose between losing some of his crew to the many headed Scylla or all of his crew in the whirlpool mouth of the Charybdis [http://en.wikipedia.org/wiki/Scylla\\_and\\_Carybdis](http://en.wikipedia.org/wiki/Scylla_and_Carybdis)



An particular area of importance raised by Vroeijsstijn concerns the definition of quality that will underlie the evaluative activity and who will define it (1995: 12ff), a subject that has been a constant source of interest and debate across all public service provision (Dahler-Larsen, 2008). Vroeijsstijn goes on to argue that definition of quality in higher education will naturally differ related to audience, be it provider, participant or funding stakeholder, and therefore any definition must recognise a plurality of views, as well as different aspects (1995: 13-14). Rather than just a demand that has formed externally, Vroeijsstijn argues that HEIs are also interested in quality, claiming that “to deliver quality is innate in the academic attitude” (1995: xiii), and quality assurance could easily previously have been seen as internally focused evaluation for improvement. However, in claiming that “quality can only be assured by those who are responsible for the quality: the staff and students of the higher education institutions”, Vroeijsstijn called for a positive, proactive attitude to the process rather than a reaction of retrenchment (1995: xvi). Thus the author recognises the strong potential for variety in decision making responses concerning the evaluative action.

Harvey (2004-8) summarises the debate quite succinctly, recognising that accountability appears to be about “value for money and fitness for purpose, while continuous improvement in teaching and learning is about enhancement of the student experience, and empowering students as life-long learners”<sup>87</sup>. Harvey suggests further that improvement is considered secondary to accountability, and it is often believed that focus on the latter will improve the former. Harvey challenges this perception, suggesting that accountability demands will likely merely only be complied with by academics, rather than replacing or transforming current behaviour. In addition he considers that momentum toward improvement, after initial impetus when introducing such systems, will diminish and that adding an extra formal requirement to those already engaged in evaluative activity will demotivate professionals and decrease feelings of trust (ibid.). Newton’s research suggested that while academic staff agreed that “external and internal accountability requirements had been met, there was a marked ‘implementation gap’ requiring explanation... [and] accountability and improvement had not been reconciled” (2000: 155). Harvey and Newton therefore call the tension between accountability and improvement “illusory” (2007: 230). If there is a “tension” it is between perceptions of quality at management and operational levels (Newton, 2000: 155). Harvey and Newton rather conclude that “compliance has nothing to do with improvement”, the former focused on value for money, the latter on student experience; they are “distinct” but without “intrinsic tension” (2007: 232). Commenting upon this research, Westerheijden, Hulpiau, & Waeytens noted how responses to demands for quality assurance and evaluation vary from pragmatic acceptance to rejection as “meaningless ritual. Others distrust it or feel it as a discouragement” (2007: 306). These factors will be important to

---

<sup>87</sup> <http://www.qualityresearchinternational.com/glossary/accountability.htm>

consider when discussing with respondents their perceptions of the evaluation frameworks.

#### **4.1.1 Quality assurance and evaluation in Norway**

Stensaker (2004) noted that over a period of 15 years there was a drift in Norwegian Higher Education from State to local steering. At the same time funding regulations have become more based upon “output based factors”<sup>88</sup> as highlighted in the Government white paper number 27, 2000-2001 (Ibid: 349). The White paper (KUFD, 2001: 8), to become known as the “Quality Reform” focused on quality as the paramount characteristic and demand for the education system, including the announcement that at all levels “respectable tools” to measure quality would be put in place. Bleiklie (2009) considers higher education policy to have altered drastically during the Bologna period, describing Norway as both “front runner” and “eager beaver” in implementation, even though later the process appeared to slow down. Increased involvement and decentralised responsibility to develop the evaluation tools were key to the system’s development. At the same time the system was developing from one of “authorisation and recognition” into one based upon “accreditation” (Stensaker, 2004: 349). Stensaker refers to the former process as “administrative procedure”, which was of limited scope, until the change in focus towards accreditation and the formation of ‘NOKUT’.

The Norwegian Agency for Quality Assurance in Education (NOKUT) is an independent public body established in 2002, which is also responsible for oversight of the national accreditation system of Higher Education Institutions. According to § 1-6 of the Act on accreditation, evaluation and approval of Universities and University Colleges (KD, 2005), and further deepened in the Regulations FOR 2005-09-08 nr 1040, each accredited higher education institution shall have a quality assurance system in place, from January 1st 2004, with evidence of satisfactory documentation of all processes that influence the quality of academic studies, as well as the ability to reveal weaknesses in the system. In addition, student evaluations will also be included in this process. The quality assurance system shall be re-assessed on a six-yearly cycle.

This quality assurance process at institutional level involves the institution under study preparing a self-evaluation report, building on the categories of their own choice but under criteria determined by NOKUT, which is then assessed by an external committee formed by NOKUT. NOKUT may also choose to evaluate a particular area, such as all teacher training institutions. According to §1-3 of the Regulations (KD, 2005), while the Ministry places general demands on the system and may request or decree a particular evaluation in relation to the assessment of the quality of Higher Education, it shall not instruct on the technical elements of the evaluation. This is the responsibility and mandate of

---

<sup>88</sup> Examples are number of credits and graduates.

NOKUT, but should be in line with European standards, especially in terms of accreditation and quality assurance (Ibid.).

These standards and guidelines reflect those drawn up by ENQA, outlined earlier. Langfeldt and Hovdhaugen (2006) recognise the importance of the Bologna process and increased internationalisation on the development of a quality assurance system in Norwegian HEIs. However, the authors at the same time challenge the limitations and reductive nature of the “so-called ‘student satisfaction surveys’” (2006: 25). Thus the authors question the validity and reliability of the measurements in the system. Although the authors focus mainly on the accreditation system for study programmes from NOKUT, there is obvious application to how programme and subject leaders design their self-evaluations. NOKUT’s methods, criteria and approach and relationship to government and institutions help explain the wider context in which these decisions are made.

Lycke (2004) noted that it is quality development rather than quality assurance that has been the traditional aim in HE policy and practice in Norway. A ground level “quality of studies” approach saw little coordination of processes and follow up, neither at institutional nor central levels (2004: 220). However, increasing political interest from the late 1990s, establishing the Network Council and a new law strengthening institutional autonomy, saw a shift “from ‘grass root’ engagement to leadership responsibility” where the increasing emphasis was on quality assurance, accountability and systematisation (ibid.). Lycke also recognised from a review of expert reports within NOKUT that the criteria for evaluation were all “grounded in management theory and experience”, departing from Norwegian tradition in education (2004: 225). Lycke, being surprised by reports of enthusiastic acceptance and cooperation from academics to QA initiatives in Norway, considered that her research appeared to diverge from that of Newton (2000) who found that academics in UK were generally in opposition to the basis of QA demands (2004: 226). Her findings are however taken from expert reports of evaluations rather than investigations directly involving staff. Lycke did, however, find a “dilemma” in attempting to balance accountability and improvement and steering and democratic processes.

In Norway, NOKUT (2003) make it quite clear that the responsibility for ensuring satisfactory quality of educational studies on offer rests with the provider institution itself. Self-evaluation based on a comprehensive quality assurance system is a clear requirement for all accredited institutions. Quality is defined generally as that which satisfies students, meets accepted academic goals, and is relevant to societal demands based on prevailing standards and criteria for accreditation of institutions and study programmes (Ibid.). But, at the same time, NOKUT is a control organ as their focus is placed on the

‘quality’ of quality assurance systems rather than quality of education<sup>89</sup>. This confirms the responsibility of the institution itself to evaluate academic activity. Stensaker (2006a) considers this to be development of a quality culture which may be an embedded part of Dahler-Larsen’s concept of evaluation culture (2006b). Unlike the evaluation culture it might appear, though, that this process might limit the onslaught of the pervasive evaluation culture within strong groupings.

How respondents at programme level perceive the importance and impact of these external demands will be interesting. In this situation those responsible for developing evaluations at programme level are thought to be responding to both external and internal demands. Discussion concerning evaluation designs and models chosen as a response to these perceived demands should enlighten the decision making process. For example, do programme providers attempt to respond concordantly or discordantly to their mandators? Do chosen designs seek to be instrumental, symbolic or conceptual?

Stensaker reported from an evaluation of the implementation of quality assurance systems in Norway that in order to be considered “meaningful<sup>90</sup>” and “appropriate” academic staff needed to be involved in the design and operationalisation process and leadership should have integrated their focus on quality within their traditions as well as building towards future needs (2006a: 10). This is also to be tempered by questions of pressure in the system. While one might perceive the setting up of quality assurance systems as exerting greater central control over the HEIs, a deeper investigation is required into the motives, goals and structures built to achieve improved quality and greater control. Stensaker (2006b) suggests that despite the perception of increased critical demands on institutions there is, in fact, greater leeway than might otherwise be believed. In particular Stensaker questions whether the establishment of NOKUT led to the proposed differentiation between the political and technical approaches to institutions. While the Ministry has developed and outlined concrete, detailed demands and “standards” for Institutions to follow (2006b: 15), NOKUT’s formal framework is much more vague and more open to local interpretation (2006b: 17). Stensaker thus questions whether NOKUT’s control function might be interpreted as symbolic when considering that minimum standards are difficult to establish, that they open for interpretation and that there is little to distinguish between accreditation and other forms of evaluation allowing for disagreement over qualitative interpretations of findings (2006b: 16-17)<sup>91</sup>. As a result of this, NOKUT appears

---

<sup>89</sup> Recent developments have seen moves to amending and reframing the role of NOKUT to additionally operate as an advisory body.

<sup>90</sup> Translated from the Norwegian text.

<sup>91</sup> There are some specific demands that Stensaker also recognises ... percentage of positions of special academic competence etc.

to have (2006b: 18) either adopted or developed the role of mediator between Government and Institutions<sup>92</sup>.

Stensaker (2006b: 24-25) reporting further on the OECD report into tertiary education in Norway (T. Clark et al., 2006) notes how the country differs from other OECD countries in that QA is considered less rigorous. Particularly and interestingly for this study, Stensaker notes that focus is placed at the institutional level rather than the study programme level. This is evident in the policies of the largest higher education institutions in Norway, whereby the results are aggregated up through the system and compared primarily at faculty and institution level. Stensaker also focuses upon the breadth of representation within the Norwegian committees and decision making groups as differing from international movements. This difference is important, but also reflects the underlying desire that such processes should be as democratic and representative as possible in Norway. Such an appearance is an important factor. This raises an important issue concerning the perception of demands within the evaluation system. In discussing the likelihood of the current system leading to increased quality, Stensaker (2006b: 30) suggests that the focus upon quality assurance through increasing institutional capacity in concert with external peer review in Norway is much more indirect than across the rest of Europe. This approach reflects the view that quality assurance in Norway is thought to involve more tasks than just an assessment of quality (Langfeldt & Hovdhaugen, 2006; Stensaker, 2006b). Stensaker (2006b: 28) also agrees with Langfeldt and Hovdhaugen (2006) that there are by and large less precise tools that judge the academic quality, thus producing a much more general evaluation approach.

#### **4.1.2 Quality assurance and evaluation in English higher education**

Bauer and Kogan recognise that universities in the England have traditionally been “almost wholly free from state control” (2006: 27). Since the 1960s the UK as a whole has been understood to have undergone 5 major periods of higher education reform, moving from expansion of the welfare state with increased demands for university access and massification, through economic stress and cutbacks, to requirements for greater financial control and quality assurance. The latter developments began in the 1980s along with the implementation of NPM by the governing Conservative Party. Despite focus under the New Labour Government shifting during the late 1990s towards social inclusion, there has been a continued emphasis on market mechanisms in higher education. A shift in control and greater freedom for HEIs during the 1990s after the dissolution of the binary system also brought about greater competition despite stronger financial control (2006: 33). In 1992, universities were authorised to award their own degrees.

---

<sup>92</sup> Stensaker (2000, 2003, 2006a, 2006b) accounts for these developments both from an historic and contemporary standpoint. The significance of these points should not be undervalued and are relevant in relation to the perceptions of institutions when designing their evaluation initiatives.

In England universities had formerly been “autonomous and self-regulating”, but a policy shift saw research and education evaluations become under the jurisdiction of funding bodies (2006: 35). Of greatest interest to this study, the evaluation of teaching and institutional audit was ultimately awarded to the Quality Assurance Agency (QAA) (ibid.). The responsibility to develop systems, assure quality and maintain standards remains part of the role of the individual HEI. This requires them to assess and account for the assessment of their students, as well as develop procedures for “design, approval, and the monitoring and review of programmes” (The Quality Assurance Agency for Higher Education, 2003: 2). HEIs should monitor against achievement of “stated aims and the success of students in attaining the intended learning outcomes”, usually by the programme team. “Periodic review” of programmes was designed to take place on a five-yearly cycle, based on external peer-assessment and validation of programme aims. In addition external examiners assess student achievements and standard of output (ibid.).

The QAA was established in 1997 to provide “an integrated quality assurance service for UK higher education” (QAA, 2003: 3). The Agency describes itself as an independent body funded by HEI subscriptions and contracts with the main funding bodies. The QAA describes its responsibility as “safeguarding” wider public interest, assuring “sound standards” in qualification and “encouraging continuous improvement” in managing quality systems (ibid.: 3). Bauer and Kogan noted a shift in emphasis from the QAA in 2003 towards a “lighter touch”, after criticism from universities over the “increasingly prescriptive evaluation frameworks” (2006: 35). Focus was shifted towards auditing the QA systems within HEIs. At the same time, there has been a notable shift towards a “professionalisation of teaching” to “enhance quality” (ibid.) and uphold academic excellence as the “leading and most prized criterion” (2006: 38) supported by the advent of “clear and explicit standards”, “subject benchmarks” and a “code of practice” for managing academic standards and quality” (QAA, 2003). Bauer and Kogan point out that the QAA’s activity is not linked directly to funding. In 1985 the Research Assessment Exercise (RAE) was established, which has eventually been taken over in England by the Higher Education Funding Council for England (HEFCE), which had developed from the Universities Funding Council (UFG)<sup>93</sup>. This significant adjustment saw the government beginning to devise goals for universities, as well as applying the “legislative and financial means” to meet them, with decreased input from academics (Bleiklie, 2006: 43). The RAE was thus an incentive based system, the difference being that it was a policy tool rather than an internal academically devised one (ibid.: 44)<sup>94</sup>. The combination of these NPM based demands towards “a normative framework for public accountability, managerialism and

---

<sup>93</sup> Which was initially the University Grants Committee.

<sup>94</sup> Bleiklie (2006) offers an excellent overview of changing policy dynamics and regime shifts towards Higher Education.

market values” pressurised HEIs in England to develop “new institutional structures, modes of management and even conceptions of autonomy to ensure their survival” (Askling & Henkel, 2006: 87).

The role of the Higher Education Funding Council for England (HEFCE) has been identified as both mediator between the Department and HEIs and steering mechanism (Broadbent, 2007). Broadbent recognises that resource allocations are increasingly used as a steering mechanism, “used to ensure that organisational systems achieve that which is required of them” (2007: 4). There is a “movement towards ‘managerialism’ and away from ‘professionalism’” and a reduction of autonomy (Broadbent, 2007: 7). In 2001 in the UK programme assessments were substituted by institutional audits, without necessarily implying greater trust within the HE system (Westerheijden, Stensaker, et al., 2007: 7). Since 2004 the UK has seen public policy defined standards in terms of what information should be available with regard to the quality of HE and these are controlled through national accreditation (Blackmur, 2007: 15-16).

Henkel considers the British evaluative system of higher education as the “most comprehensive and intrusive yet devised”, auditing QA systems, provision of education and research output (2004: 91-2). The new focus on transferable skills as a “conception of knowledge”, challenged educational and disciplinary traditions and ethos (Henkel, 2004: 98). Henkel (2000) notes that quality assessment in England was a governmental initiative, coming to the forefront in the 1980s with the drive for greater accountability across the public sector which should include HEIs. Henkel recognises that these moves simultaneously challenged the concept of “institutional autonomy” and “collective public accountability” (2000: 70). Case study research showed that universities experienced quality initiatives to be focused more on accountability and “obtaining demonstrable value for money” (2000: 84). The increased demands for improved processes and reporting had also led to institutions appearing more structured, operating with a more “managed order” and greater consistency under “more generic concepts of quality” (2000: 94). At the micro level academics considered the developments of these systems as removed and distinct from the educational exercise of their roles (2000: 99). Additionally, the influence of quality assurance was considered by academics to be “pervasive”, considering themselves under “continual scrutiny” (2000: 96-7) in what felt like a “zero-sum game” using time allocated for improving their work to “meeting the demands of an administrative concept of quality” (2000: 99). Henkel goes on to note that this often collided with academic notions of quality being discipline centred (2000: 106). However, one interesting finding from Henkel’s work<sup>95</sup> was that departments were often motivated to greater collective work processes as they responded to demands to develop their quality assurance policies, even though this did not change their “basic educational values” (2000: 111). While the focus of this current study is not on academic identity per se, investigation is

---

<sup>95</sup> which was mainly focused upon academic identities

undertaken into perception of roles and decision processes within subunits and related to decision processes in the wider institution.

#### **4.2 Summarising the policy shift in evaluation and assessment in higher education institutions**

The massification of higher education and the size of HE systems appear to partly explain the need for greater formality of management (Westerheijden, Stensaker, et al., 2007: 2-3). Linked to the prevailing public policy climate of reform and regulation outlined in the previous chapter, quality assurance is considered by the authors as “here to stay”. Defining the purpose and form of QA are, however, not easily rectified not least because the purposes of HE are so diffuse and even though the central focus should be on student learning, the potential outputs are so difficult to identify, (Westerheijden, Stensaker, et al., 2007: 4).

Kogan notes that “there has always been evaluation in higher education”, based on certification and validation of knowledge and its producers (2004: 3). However, the advent of massification, increased competition and “political suspicion” of profession power has increased the importance of, as well as changing the nature of, evaluation (2004: 4). Evaluation developed as an “instrument of public policy”, highlighted further under NPM (Henkel, 2004: 86). Quoting Neave’s (1998) idea of the evaluative state, Kogan goes on to recognise the continuing shift towards evaluation for “policy adhesion”, as well as a posteriori focus on product control rather than process investigation, with a purpose of steering HE more closely towards “national priorities”. This was also problematic as it was combined with ex ante / posteriori financing (Neave, 2004). Kogan sees this as a collision with the traditional technology of HE, where academics set the agenda and quality criteria (2004: 6). The underlying problem is one of intention and decision for evaluation structures, that is, whether they should be “purgative or developmental” (2004: 8).

Vedung refers to the evolution of a “special evaluation tradition” within higher education<sup>96</sup>, whereby “professionals themselves carry out the evaluations against their own professional, mostly unwritten and tacit, quality norms in self-evaluations and against quality norms of their peers” (2003: 42). The underlying emphasis of such models is “dialogue, discussion and deliberation”, rather than goal-attainment or effects per se; an “exercise in professionalism” rather than “scientific exercise” (2003: 64). At the same time he recognises that a democratisation of the evaluation process has increased the role of the “ordinary” stakeholder. There is a tension between the democratic focus of evaluation theorists favouring greater participation and the increasingly consumer oriented approaches more closely associated with New Public Management reform, as outlined in the previous chapter. The former is inclusion

---

<sup>96</sup> While Vedung applies this work to Sweden many of the observations are drawn more generally from and apply to the wider evaluation field.



and improvement focused where the latter focuses more on accountability and outcomes. Vedung recognises that in complex fields such as education there has been a principal of the public sector being “profession-driven” (2003: 64-5). This is tempered slightly by adopting peer-review processes to ensure some degree of parity with other public sector arrangements. Public policy changes and greater demands for accountability as well as impact assessments create a tension for these processes, as will be seen in subsequent sections of this chapter.

Reichert (2007) outlines six “pre-conditions” for effective QA, split between individual and institutional responsibility. She argues that individuals must trust that evaluation will offer some benefit, which will involve exposing weaknesses and using “time and effort” to rectify them. At the same time, institutions must build their autonomy, develop strong leadership capable of addressing change, and provide the necessary resources for change and development. This is supported by her recognition that a key limitation in quality enhancement raised by HEIs has not always been the “nature” of the QA initiative but the resources available for follow up (Reichert, 2007: 6). Problems it seems are already evident to HEIs, but either ignored or shelved. Westerheijden, Hulpiau and Waeytens (2007) recognised systematic variations in the general QA model noting there to be distinct phases in the design and development of QA schemes. These phases are thought to be linked to the social and policy context which affect the hierarchical development of these processes (2007: 298). Aside from external issues like HE policy, political climate, economy and demography, the authors also note the importance of “internal dynamics”, which result from “the learning effects that result from the actors playing their part in subsequent rounds of quality assurance” (ibid.). This can be *positive learning* when academic staff who have developed capacity for self-evaluation become more engaged in contributing to the improvement of the institutional quality culture. The authors also recognised *negative learning*, which took place when staff “learn to play the tricks” of QA without it affecting the “internal life” of the quality of teaching and research, also known as “window dressing”. There are further perceptions of underlying problems with QA at the micro level. Harvey and Newton consider the “contention” to be how quality can be improved by “asking an amorphous group of academics to identify their strengths and weaknesses” (2007: 226). The authors suggest that these activities, along with the arrival of external “raiding parties” passing “summary judgement”, might lead to policy compliance, regulation or control without affecting quality per se. The authors consider this a bureaucratic process removed from the basic activities of education and research. Henkel (2002) also reflects that changing the structures within Higher Education in England as a result of legislation have not resulted in a significant decrease of ambiguity within organisations. It is to this topic that I turn to next.

### *The structure of HEIs and accountability*

Henkel agrees that these demands are clearly part of the wider public sector progression towards new public management, whereby HEIs have been “required to increase their efficiency and to subscribe to various forms of quality assurance” (2002: 29). The author reflects, however, that this creates a challenge for HEIs, as they must balance commerciality with academic standards and an increased tension arises between “mediating” central policy and maintaining and strengthening their institutional autonomy (2002: 30). Henkel noted that in England this was reinforced by a greater degree of direct intervention from the state as well as the requirement to improve management structures.

As part of these wider reforms, Henkel notes how organisational structures within HEIs have come under increasing attention, especially with regard to the long standing notion that decision making is based upon collegiality and community. The author recognises that this perception may have been overemphasised, and notes that there has often been difficulty in resolving conflicts that are endemic within generally loosely coupled systems (2002: 30). It is interesting though that she further states, drawing on the work of Bargh et al., that accountability has not in of itself been merely about improving and tightening structures, but rather that accountability appears to be viewed with regard to societal interests including those of the state. It might appear, though, that the latter have become synonymous with central demands.

Within these accountability focused systems, QA plays an increasingly more important role. Henkel’s research notes that the policies linked to QA are linked to the allocation of resources, which in turn is linked to organisational reputation (2002: 33 - 34). Henkel suggests that as a result “academics are under constant scrutiny by senior managers” within a system of “growing insecurity”, whereby academics feel they must meet the needs of administrators rather than the other way around (ibid: 34). This is of course a nuanced position, and Henkel recognises that despite all the changes within the system, leaders continue to insist that it is the academic at the base level that drives “institutional success”.

Becher and Kogan (1992: 169) offer three, “potentially conflicting”, modes of accountability: public contractual/managerial, professional and consumerist<sup>97</sup>. They note that all these modes influenced higher education from the 1980s. Brennan and Shah recognise that these are related to internal decision making in HEIs, coming to the heart of value structures across groups as well as those held by individuals (2000: 33). The public contractual mode is about performance related to collective policy models, whereas professional accountability focuses more on the intrinsic quality of a particular subject related to the values within

---

<sup>97</sup> These modes also appear similar to Bleiklie’s (1998) conception of universities as government agencies, cultural institutions or corporate enterprises which later developed into 4 expectations and are outlined further down (2004).

that academic field. The consumerist mode is more greatly focused on responding to market demands. These three modes offer a useful framework against which interview responses can be considered. They are, however, anticipated to exhibit a degree of overlap.

The development of HEIs has led to four major expectations or templates of how they should be organised via: academic quality, collegial coordination, social responsibility and business enterprise (Bleiklie, 2004). The fourth template is a departure from more traditional ideas of Higher Education, where the HEI is considered a “producer” of “quality” services. Bleiklie remarks however that whilst “quality and ‘quality assurance’ are emphasised as fundamental goals, the most important expectation... is the *efficiency* with which it produces useful services... to the benefit of the *users* of its services” (Bleiklie, 2004: 14). This became increasingly more noticeable in both Norway and England from the 1990s, where despite different emphases of policy there was a central “concern” about costs of HE and greater interest in the “product” (2004: 18).

As was noted above, in recent times the shift in England has been from higher education run by state bureaucracy to greater autonomy for HEIs in an attempt to produce a more flexible, deregulated governance driven system “responsive to contextual (societal) demands” (Westerheijden, 2007: 75). At the same time, Westerheijden notes that both sides of the “North Sea” were increasingly interested in value for money, rate of return to society and economy, opening for market mechanisms and consumer choice. Change should be improvement focused and instrumental, rather than incremental, and HEIs held accountable (2007: 76). Images of quality higher education were thus redefined.

Questions are raised, then, as to how much change is visible as a result of legislation to change the management style and decision structures across HEIs. Despite the development of more hierarchical structures Henkel claims that there is still a “high degree of organisational complexity and ambiguity”, and this appears to be moderated by age and culture, or how “traditional” the HEI under investigation is (2002: 35). Such categories are of course difficult to define, but Henkel (2000, 2002) suggests that academics are less likely to express themselves as managers within more traditional universities. In addition, ambivalence was a common feeling amongst academics concerning being a manager. Henkel’s findings could, perhaps, be considered also to illuminate attitudes to decision structures and processes, in this case concerning what the focus of an evaluation should be, how should it be decided and who should be involved in that process. There are of course many other factors that will influence such attitudes and perceptions, including prior experience of those involved, both inside and outside of the organisation of which they are members. The most important point is though that the introduction of such principles into the academic arena will be tempered by institutional values and traditions, which will often become a source of greater ambiguity, despite attempting to bring

greater clarification (2002: 37). The development of quality assurance systems, as well as external organisations, designed to control and oversee implementation and progress, is described by Dahler-Larsen as part of the institutionalisation of the quality wave (2008: 67). The author notes a shift of focus towards the quality of way organisations act rather than what they offer, with greater interest in environmental acceptance than resolving instrumental tasks (*ibid.*). These perspectives will be outlined in more depth in the next chapter.

### **4.3 Developments in the quality culture**

Discussion has surrounded the development of “quality culture” within HEIs, which Gvaramadze considers to require a common definition of the concept of quality (2008: 445). The author recognises however that the concept is “a contextual phenomenon” that needs appreciation of the specific HEI as well as the national context. This was the conclusion of the EUA and noted in their policy position (EUA, 2007 in Gvaramadze, 2008: 445). The EUA position highlights the “inextricable link between institutional autonomy and accountability”, recognising a reciprocal robustness between the two (European Universities Association, 2007). Affirming the European Quality standards, focus in the policy position is placed on developing internal improvement focused processes that emphasise shared values that develop professionalism and creativity rather than managerial processes; are fit for purpose but also linked up; where leadership frames the processes and follows up; while developing “non-bureaucratic” quality units; ensuring that data is used to measure institutional performance.

Gvaramadze notes that these “bottom up” processes are characterised by transformation at the programme level and enhancement at the institutional level (2008: 445). Enhancement requires that leadership develops a common vision and enables mechanisms to be linked to institutional objectives, further highlighting institutional autonomy, less bureaucracy and continuous improvement. It still requires greater transparency, but the purpose will be for external evaluation linked to quality mechanisms rather than control. Stakeholders should be involved and the goal is to maximise effectiveness (2008: 446). Interestingly the subject of individual development is also taken up, under the guise of “quality as transformation”. Under this process skills of both teacher and student are enhanced and empowerment takes place to bring “value added” to the latter and making evaluation participative and learning centred (*ibid.*). Gvaramadze considers it difficult to align these views, noting difference of opinion within HEIs, particularly at sub-unit level and across academic fields. The requirement for universal participation amongst multiple interpretations is difficult to build upon, when “each programme as a university unit has its own identity and culture”, and therefore requires unique indicators to be built into each programme (2008: 452). The author notes that developing a “quality culture” involving increasing the responsibilities for greater stakeholder

participation is complex for students, staff members and institutions as a whole. How they respond to the challenge is vital to understand. Evidence suggests that so far this has been far from easy in practice (Harvey & Stensaker, 2008). What Neave (2004) noted, however, as problematic was the acceptance of Bologna during the time of rise of the “evaluative state”. While each of the signatories was being called to adopt generic structures they were at the same time adapting their own accountability systems, particularly with regard to evaluation. Crucially the two processes have impacted and strengthened one another, credence being given to the control functions of the evaluative state as well as the Bologna process being allowed to develop further its own agenda (Neave, 2004: 32).

These reflections raise important questions for the subunits under investigation in this study. Musselin questions the complexity of the new role for academics within the changed organisational framework that often develops as a result. In addition new tools to enable decision making are developed, but these are “not neutral” and often provide more information than has previously been available (2002: 6). These reflections appear to fit with the reasoning behind the focus of investigation for this study. The importance of the role of the academic is explored, in the light of the demands placed upon and within their programme group, wider in their institution alongside those forming within the external environment. In addition, focus is placed upon the designs for evaluation and the tools that are developed to produce the data required at these different levels. The increasing amounts of data that are produced are also discussed, although an increase in information might not only be interpreted as positive and presuppose effective utilisation (Feldman & March, 1981).

Bleiklie (1998) recognises how the development of the “evaluative state” combined with the acceptance of “corporate management ideals” has signalled a shift in attitude from public policy makers to HEIs. However, rather than a wholesale shift in values, ideals and expectations, the author recognised multi-phase development of “different layers of expectations that gradually have been piled upon one another in keeping with the historical transformations the university has undergone” which led to alteration of institutions rather than new creations (1998: 310). There appears to be evidence in this study of the way that these ideals generally rooted in NPM have influenced the academic and administrative actors in different ways and to different extents. But the general pressure towards more complex arrangements of assessment and evaluation and the linkage to both funding and assessment of programme purpose and content create difficult dilemmas and choices. To do one thing well creates pressure on individuals and groups to do other things in a more limited way than they would like. And this is even situations where actors are in general agreement with the premise to evaluate more effectively. Bleiklie reminds us that the developing tautology that is HEI policy creates more tension in the system; moves toward firmer and more effective management noted in greater standardization and performance indicators are combined with the contrasting moves towards less

centralised control to improve “efficiency and flexibility” where each HEI makes its own decisions of allocation (1998: 310). This is emphasised in the varied approaches to evaluation, and specifically to quality assurance, alongside the wider structural changes.

### *Higher education governance in flux*

Following on from the changes in public policy and their impact on evaluation, this section considers further their impact on the context of this study, the field of Higher Education<sup>98</sup>. Ferlie et al (2009) outline different conceptions of higher education governance, roles of the state and narratives of public sector reform. They note that there has been a shift in governance of HEIs, from the Mertonian idea of autonomous institutions where education develops separate to public policy and reform and academics are seen as professionals with freedom, where power lays with the faculty (2009: 3). The authors offer two developments since the disintegration of the Mertonian model. The first observes the state as an interventionist mediator between society and HEI, where the state assumes greater control over the public sector and attempts to streamline processes to benefit the knowledge economy (2009: 4). The second “conception” swings more towards market governance of HE, where teaching and research are “commodities rather than public goods”.

Ferlie et al (2009) also see “redefinitions” in the role of the state since the 1980s. The first redefinition is a “more restricted and managed [public] sector”, which developed from New Right policies and placed pressure on the Mertonian concept of autonomy. This was particularly evident in England but also began to influence countries like Norway that had more strongly held onto concepts of institutional autonomy and academic freedom. Aside from funding changes, the governance changes brought an increase in “intermediary bodies” along with the shift from “ex-ante control in favour of ex-post evaluation” even though HEIs continued to be administered by rules (2009: 8). The second redefinition exhibited a “hollowed out” or weakened state, where a blurring of boundaries between other actors and nations leads to a governance model of public management and greater supranational influence (2009: 8). As will be outlined further below, the influence of initiatives like the Bologna process and declaration have had strong impact on national ministries and further down through the various levels. Ferlie et al also note how this increase in number of stakeholders and diffusion of power has complicated the notion of accountability. The impact has been felt down to the micro-level. The third redefinition concerns the “*democratic revitalisation* of pathological and over bureaucratised traditional forms of public administration” (2009: 10). This has led to greater stakeholder involvement in public policy making, more generic propositions and the idea of evidence based and informed decision making. This

---

<sup>98</sup> Issues concerning the impact of the subject area of School Leadership development were outlined in Chapter 2.

has also infused the governance structures of HEIs, with the introduction of non-academics onto the boards of Norwegian Institutions and into English research boards and councils (ibid.).

While each of these redefinitions appears at odds with the others, the result has been that they intermingle to develop a more confusing framework for HEIs. However, within this study it will be important to discover how these changing relationships have been perceived to impact on the decision making processes at the micro level, in this case with regard to evaluative activities. Ferlie et al. (2008) also recognised in the work of Kogan et al (2000) that HEIs were seen to insulate themselves from external policy demands. There appeared to be limited diffusion from macro to micro levels, which as will be seen in the next chapter, characterises Institutional models of organisation.

Ferlie et al. considered HEI reforms in relation to wider public sector reforms, considering three linked explanatory narratives<sup>99</sup>, New Public Management, Network Governance and Neo-Weberian (Ferlie et al., 2008) but later reducing the focus to the two former (Ferlie et al., 2009). As these models were dealt with in greater depth in the previous chapter, I attempt here to highlight the authors' comments in relation to HEIs and evaluative activity. Under NPM, the UK has been recognised to be a "main adopter", or "outlier" (Paradeise, Reale, Goastellec, & Bleiklie, 2009), where a smaller, results oriented and efficient public sector has been evident, whilst Norway as was noted earlier, has been a more reluctant adopter implemented a less pervasive programme, for example as exhibited in the 2002 "Quality reform". However, as was also noted earlier in Bleiklie's (2009) research, when it came to Bologna, Norway was considered to adopt at least initially like an "eager beaver".

One of the foci of NPM infused policy affecting HEIs is the "explicit measurement and monitoring of performance in both research and teaching; development of audit and checking systems" (Ferlie et al., 2009: 13). The Network Governance (NG) narrative exhibited an adjustment and more democratic version of NPM, where the state, whilst being hollowed out, influences more than directs, by devolution of power and relinquishing some control to supranational bodies (ibid.). Audit control is "dampened", with lighter system and more self – regulation (2009: 15). The authors note some evidence of gentle shifts in this direction (Ferlie et al., 2008). Ferlie et al. (2009) consider that the "Third Way" responses of the New Labour Blair governments in the UK reflected a reaction against the overload of control, not least the costs of, initial NPM based policies. They also consider, however, that the origins of network governance to many extents predate NPM. A third narrative, the Neo-Weberian sees the state as main facilitator with representative democracy involved, in this context, in HEI scrutiny, maintaining the view of a specialist public service

---

<sup>99</sup> The authors adopt the expression narratives to highlight their mixture of technical, political and normative elements (Ferlie et al., 2009: 11).

(Ferlie et al., 2008). It can, in many senses, be conceived as an alternative to NG (Paradeise et al., 2009). QA systems with academic ownership but referring to consumer needs, shift from ex-ante to ex-post controls, and greater results orientation. Interestingly Ferlie et al consider the importance of evaluation as tool within this framework of narratives, considering how different elements of QA frameworks have been attached to them. For example the authors see steering as part of the QAA, and the RAE as an NPM initiative. Budget allocation and human resource focus are seen as a fusion of NPM and NG. While decentralisation and alliances are more part of NG, the rise of intermediate bodies can be linked to all three narratives. This has brought about a clash of perspectives, where NPM has been considered more efficient, whereas governance is thought better than exclusive relationships, and evidence better than good faith against the idea that generally NPM destroys collegiality and shifts focus away from genuine quality (Ferlie et al., 2008). Drawing on the conclusions of Paradeise et al. (2009) it might be said that regulation by a combination of these alternatives appears evident in both England and Norway, even though the weighting towards the market orientation of NPM was much stronger in the former and more as a result of “linear implementation” (2009: 225), than the “institutionalization of collective action” associated with NG in the latter (2009: 246). It will be interesting to see how the weighting of these different narratives might shape the decision processes concerning evaluation at subunit level in the organisations under study.

Changes in the policy environment have brought a shift in the organisation of higher education institutions. Universities have been seen as moving from “administrative bodies to strategic actors” (Paradeise et al., 2009). The complex reform package implemented to enable such a shift has been coupled to wider public management reform, including increasing organisational density, diversifying funding<sup>100</sup> and strengthening micro-management, resulting in governments attempting to “steer but not row” (2009: 218).

#### *Shifts in perception of quality*

Although not the main focus of this study, it is important to recognise that evaluation as part of quality assurance systems will be framed by different attitudes to and perceptions of quality. While I do not fully intend to explore this concept, it is important to recognise that there is an inherent complexity in defining and agreeing what quality is (Øvretveit, 2005) which affects the way an evaluation is designed, implemented and reported and for the original purpose for which it was intended to be used. Dahler-Larsen (2008) argues that there has been a general shift in the “Quality Paradigm”. Consumerism has replaced production focus and measurement of quality has slowly shifted from objective

---

<sup>100</sup> Funding is increasingly linked to performance evaluation, including research output and student throughput. For example, in Norway about 40% of public funding is linked to teaching and research performance (Paradeise et al., 2009).



to subjective qualities, moving from an inspection of technical quality to quality assurance systems and from control to regulation and development (Dahler-Larsen, 2008: 27-8). The author argues that at the same time the concept of quality has been internationalised, due both to the definitions formed by international organisations as well as cross border agreements and treaties. The focus of quality shifts from attempting to find an intrinsic characteristic of an item to focusing on extrinsic measurement against a standard, as quality becomes “organisationalised”, a characteristic definition against an organisational system (Dahler-Larsen, 2008: 29). This takes place alongside the institutionalisation of evaluation and quality assurance into organisational activities that will also be considered in the next chapter.

Quality assurance has been described as a tool to ensure accountability as well as compliance to national policy (Harvey & Newton, 2007: 225). The new quality movement, and its part in NPM, has been described by detractors as a “modern Taylorism”, which is noted particularly from the separation of task between management focused upon design and worker on delivery (Øvretveit, 2005). Recent focus on ascertaining and improving quality within the public sector has greatly influenced the field of education, and in doing so upon approaches to the training and developing school leaders. Within the field of school leadership this has often been challenged as alien to the more widely accepted idea of focus on building learning organisations, particularly in Norway in reaction to changing focus of educational policy (Afsar et al., 2006; Møller, 2004, 2006b). Demands upon those providing postgraduate programmes in school leadership development appear to reflect some of these wider developments observed in the Higher Education field, as Henkel puts it, that “knowledge and learning [are] defined as key economic and social drivers” (2002: 29). As will be explored further in subsequent chapters, this was observed as respondents affirmed, especially in England, the increasing demands that their programmes should demonstrate impact, especially to account for how school leaders might work to improve results within their own organisations.

#### *Greater focus on standards*

While professional groups may claim that they best know what the needs of programme participants are, there are wider definitions of who the customer and end users of a particular public service are (Furusten, 2000; Øvretveit, 2005). This raises the question of what standards will be applied to judge quality. Standards are described as the “operational definitions of the intended level of service” and where measurement is “the assessment of the level of service achieved” which can be externally specified, assigned from customer requirements or interpreted from a combination of the two (Øvretveit, 2005: 544). The process, particularly of measurement, is though far from straightforward and requires a level of definition of standard which rarely goes unchallenged (Kushner, 2001; Øvretveit, 2005). This can be noted in the field of

education and especially with regard to the accreditation process of programmes. Although this part of QA is observed to be popular with those wishing to “reduce harm to the service user from incompetent professionals and dangerous services” it is also recognised to be “unpopular with many practitioners... takes time and bureaucracy to implement [where] resources might be better used for other actions to improve quality” (Øvretveit, 2005: 548). But, how parts of this process like self-assessment and peer review are monitored, controlled and utilised raise important issues with regard to how decision makers approach the task decentralised to them. That QA is often also tightly tied to particular standards also raises concerns. There can, also be a dichotomy. Standards may be presented as a concrete set to be applied and followed, while concurrently having a stronger underlying rhetoric. Standards may thus be seen to mean much more than they actually say. Standards for quality are often based upon assuring inputs, processes and outcomes, and although the purpose is often thought to outline a process easily understood which can be implemented within the current structure, this structure might be too weak for the task or unwilling to perform it (Øvretveit, 2005: 548). Standards are now envisaged to be on the supply side, developed outside of the organisation from which they are intended and downloaded into organisations (Furusten, 2000). As such standards have moved from being ‘procedurally’ interpretable “guides to position and progress... with... a tolerance for an essential lack of precision” to meaning “a measured target for purposes of justification” (Kushner, 2001: 121). Kushner refers further to the reductive element of such behaviour, whereby complexity is diminished and context becomes less important.

#### *From deliverer to user*

There is also great debate over the extent to which one holds a “customer perspective” and who is included in such a concept. Øvretveit (2005: 554) notes that although the “primary beneficiar[ies] of education” are considered to be the actual students enrolled on a particular programme, there are other customers in the wider society. In the case of school leadership training and development programmes one might perceive Government, school owners, teaching staff, parents and pupils as indirect customers or recipients of the benefits of these enterprises. This has implications for understanding the purpose and focus of a particular programme, but also importantly for the intentions with evaluation and the evidence of effects sought, which returns us to the questions raised by Guskey (2000) and Leithwood and Levin (2005) with regard to level to which programme impacts are evaluated.

Such an attitude is considered to influence the decision making regarding designs of evaluations, particularly when decisions are decentralised, to varying degrees within organisations. A complex web of different purposes, ideologies, competencies and SOPs will influence such decisions. As Øvretveit reflects in particular, the different viewpoints over quality maintains “professional

boundaries and autonomy”, where “language may be the professions’ last defense” (2005: 557). The purpose of this study is not to isolate these different phenomena from one another, but rather to gain an overview of how they are interpreted by decision makers and influence evaluation design. Questioning the interpretation of demands, as perceived both from external and internal mandators is assumed to be one important framing factor.

Although the “main driving forces” of quality assurance are external to HEIs and the “ultimate responsibility lies in the hands of the (nation-) state” (Schwarz & Westerheijden, 2004: ix), more needs to be known how such developments impact decision making at the micro level, particularly as most accreditation and assurance systems devolve the forming activities to this level, while maintaining the control mechanism higher up. While at a macro or meso level this might encourage focus on the “blurred boundaries between accreditation and evaluation” (Schwarz & Westerheijden, 2004: ix), the necessity to quantify quality for the mandator creates an added demand at the local level to produce evidence that will in essence be “used” for both publicity and programme improvement in the student marketplace which now encompasses the wider world. Universities appeared to struggle with the dichotomous value shift towards increasing managerialism combined with market orientation and decentralised responsibility linked to NPM (Kogan, 2004: 3). One of the results of decentralisation of power to institutions has been an increase of centralisation of decision-making at the institutional level, reducing collegial governance at lower levels in order to increase cohesion across the institution (Amaral, Jones, & Karseth, 2002: 289). This, however, is a process rather than fait accompli.

#### *Implementation effects at the macro-level*

Paradeise et al. recognise the complexity of factors underlying the introduction of QA frameworks at macro level, and while reform processes introducing greater focus on managerialism and control might be similar across different boundaries, they mostly “remain path dependent and... incremental” (Maassen & Olsen, 2007: 197-8). This only adds to the complexity within each country’s HE system, and further into each HEI. Despite the recognition again that this study focuses at the micro level of institutions, there does appear to be one important difference that comparative research of reform policy within Higher Education between Norway and England revealed. Evidence of managerial reform was noted across both countries, but one major difference was the lack of “soft budgetary restraints” in Norway compared with England (UK). In UK there was emphasis in policy upon financial control, efficiency and value for money, as well as greater focus upon stimulating competition between HEIs. These were not evident in the research in Norway.

The shift from ex-ante to ex-post evaluation and monitoring since the 1980s has been evident in England through the introduction of RAE, quality audits and self-assessment and the creation of national evaluation and accreditation

agencies. The later developments were reinforced more widely in the 2003/2005 Bologna conferences (Paradeise et al., 2009: 220). However, despite apparent convergence of HE policy across in particular West European borders, Paradeise et al. maintain that change has been more incremental rather than planned. They do however concede that NPM has been the “cornerstone” of public service reform in England (UK), particularly the “equating of strong management with managerialism and greater control (2009: 218). The authors consider implementation of the reforms in UK to resemble more linearity than in other countries (2009: 233). In addition, the picture of Norway as “reluctant reformer” stems from an observation that “old patterns” tend to reassert themselves and slow down the process of policy change. They point to the resistance to the 2002 Quality reform as an example of hindrance by localism and incrementalism as the academic field reacted to the promotion of teaching over research.

#### *Impact of QA on decision making within HEIs*

The centralisation of power with regard to NPM shifted power from academics to the government and other stakeholders, while increasing responsibility to them for their “survival and prosperity”, related especially to performance (Henkel, 2004: 86). Internally, HEIs also saw greater centralisation of decision making and managerial control over issues like research, curriculum, teaching and learning methods and quality assurance (Henkel, 2004: 94). Henkel saw this as challenging traditional conceptions of authority, hierarchy and security. It also led to greater involvement of administrative staff and non-specialists within the core professional practice. At the same time it appeared to increase the collegial effort within departments as self-evaluation required greater collaborative effort and mutual support for the survival of both work place and subject.

This raises an important question with regard how these reforms have affected the decision processes within HEIs. Paradeise et al. found evidence that at subunit level the “inner life” was “out of reach” of the centralised authority, continuing the “ideal of collegial autonomy” (2009: 230). They did find however that the assessment control and “full cost accounting” evident in the UK had increased the pressure on academics in a way that was witnessed in few other places<sup>101</sup>. These findings concur with Bleiklie and Kogan who compared “drastic change” in English higher education policy with Norwegian emphasis on “continuity and gradual change” (2006: 15). These ideas will be explored further in the next chapter.

Brennan and Shah’s (2000) comparative research focused on the frameworks for quality assurance and their impact on institutional management and decision-making, with the latter focus on decisions and impact resulting from findings. The authors consider that quality ‘assessment’ is regarded to concern “the

---

<sup>101</sup> The only other example the authors give is the Netherlands.

traditionally private 'inner worlds' of higher education institutions and the political and social contexts which are increasingly important on shaping these worlds", linking the "private micro" and "public macro" worlds (Brennan & Shah, 2000: 1). They also note that the influencing factors at national level were thought to be similar across European boundaries (2000: 10).

Interestingly despite arguments favouring the idea of "inner-life", Westerheijden recognises that the development of QA systems and greater external demands have also on occasions appeared to lead to more attention at the micro level, which increased academic interdependence and cooperation but also increased stress (Westerheijden, 2007: 83). This seems to strengthen the need for further investigation at micro-level.

#### **4.5 An exemplifying case: evaluating impact of the Quality Reform in Norway**

The Quality Reform obviously placed focus on HEIs improving their evaluation systems and preparing for greater accountability that followed their increased freedom. Michelsen and Aamodt (2007) note the expectation that quality systems would permeate HEIs and take focus from the level of individual engagement to a common direction at system level. Evaluation of the introduction of the Quality Reform revealed that around 85% of those in HEI leadership declared that developing a new quality assurance system has been their "most central task after the introduction of the Quality Reform" even though there is individual variation (Michelsen & Aamodt, 2007: 48). The authors do recognise that systems have generally improved and coordinated focus upon quality, mainly at the programme and subject level, as well as on results and governance. In addition the new systems, rather than replacing the old, have often run concurrently with them. However, they also recognise that it is the administration that has generally been responsible for developing the quality system, and academic staff have had varied involvement (Ibid : 49), which has led to both a bureaucratising and professionalizing effect on the evaluation processes. The former has led to a shift in focus from academic to administrative processes, even though the latter has built up competence within the system. It would therefore seem to strengthen the question of how evaluations within programme are formed and take place. In addition, the authors recognise that it is still too early to suggest how these processes will be assessed and followed up in practice.

Another paradoxical consequence, according to Michelsen and Aamodt (2007), has been the suggestion of less involvement for academic staff and students in the quality assurance process than anticipated. They claim rather that these parties have been "decoupled" by the bureaucratising / professionalizing process, with resistance from individual members and the handing over of tasks to those thought competent (Ibid : 50). They summarise that there is an "increasing tendency towards formalisation and centralisation of the quality

assurance process at places of learning, a high degree of administrative adaptation to external standards where it is currently only possible at some places of learning to uncover attempts to adjust the quality systems to internal needs and the ambition to develop” (Michelsen & Aamodt, 2007: 51). Such behaviour has been observed more widely when standards are introduced without any real evidence of the standard being practised (Brunsson & Jacobsson, 2000).

Stensaker (2006a) recognises that a key part of this reform has been the strengthening of the administrative involvement in the quality assurance process. Sanctions for not having an approved system are serious for a HEI, whereby they will lose the right to accredit new study programmes and have to rely to central approval, a process that inevitably runs more slowly than competitors within the HE sector (2006a: 8). However, despite the aim of improved quality based upon the individual institution’s traditions and specialised focus, Stensaker also recognises that many merely adopt the central criteria set out by NOKUT, while others imitate, translate or copy the structures of those institutions that have already achieved accreditation. Stensaker recognises that quality assurance is a relatively new phenomenon in Norway (2006a: 9), but that as with general trends there has been a higher degree of centralisation and formality in the system. At the same time, as was suggested earlier, the approach from NOKUT appears to have been somewhat more informal, vague and little defined (2006a: 10) based more on bargaining than compared to European standards (2006b).

#### **4.6 Summary**

This section has focused on the development of quality assurance in higher education in England and Norway, the contexts chosen within which empirical investigations will take place. Quality assurance was linked to and considered in relation to evaluation models and imperatives outlined in the previous chapter. This was followed by consideration of recent policy and practice developments within both countries. HEIs in both Norway and England had been previously subject to evaluative assessment, which in the former is considered to have “contributed to the readiness” for new initiatives (Askling & Henkel, 2006). Interestingly both countries have appeared to exhibit a level of convergence, even though England has increased central control whilst Norway has devolved greater responsibility to its HEIs. There was however significantly greater pressure for financial control and subject focus in England compared with Norway. While there have been moves at supranational level to provide models appropriate for QA in higher education institutions, there was a notable difference in detail between the countries. Brennan and Shah saw from the outset that there had been variation between countries, as well as variation between and within HEIs within the different countries (Brennan & Shah, 2000). Brennan and Shah see particularly a contrast between the traditional autonomy of units within HEIs and the shift in focus for organisations towards

“collectivity, transparency and accountability”, promoting “evidence” over professionalism (2000: 16). The authors recognised that quality assessment could have impact on decision making, where the underlying assumption is that its transparent nature ought to “support greater rationality” especially increasing the use of evidence, but they noted that this was affected by “power and influence of existing interest groups in institutions” (ibid.). Mission-based evaluation focuses upon “fitness for self-defined evaluation”, as is mainly found in systems in the USA, whilst standards-based evaluation concentrates on establishing “fitness of purpose” through considering output factors and defining information necessary (Westerheijden, 2007: 81). As Westerheijden notes, when governments demand information on graduation rates, HEIs “have an incentive to increase graduation rates, *ceteris paribus*<sup>102</sup>” (ibid).

A question arises as to what possibilities there are for providers to ascertain programme effects and whether indeed one can isolate any. This leads to another difficult question for programme providers, how does one justify the content of programmes? The literature studied recognises a waning of academic justification for programmes combined with perceived challenges upon academic freedom. There are suggestions of a combined challenge to institutional autonomy and academic freedom. Having said that, within these frameworks there is also a focus upon local responsibility for developing and implementing evaluation.

In the frameworks and documentation for the subunit cases outlined in chapters 7 to 9 there is a notable local responsibility for developing evaluation (QA) methods within the common framework. Results from the individual evaluations are collected, summarized and reported up through the system where findings become more generalised into an institution wide image. The focus of quality assurance, as recognised earlier, is match to student satisfaction, recognised academic goals and societal relevance. Such an approach matches general format of quality assurance systems, where quality is assumed to encompass user wants and needs against a mandated standard at the lowest cost possible – or client quality, professional quality and management quality (Øvretveit, 2005). With these aims being most reported on, one might question whether or not programme design and effect is relevant.

As was briefly discussed in chapter 2 debate over the evaluation of programmes for school leaders reflects these limitations.

The next chapter focuses on decision making about evaluation, offering a framework for the empirical study.

---

<sup>102</sup> All things being equal

## **5. Unravelling evaluation processes: focusing on decisions *about* rather than decisions *from***

Discussion in the preceding chapters focused upon a review of literature concerning evaluation, particularly the purposes of evaluation, and how models to be implemented are chosen, with particular regard for eventual utilisation. In chapter 3 it was recognised that the wider evaluation field appears to some extent to have become polarised in debate in relation to whether focus, when developing a model should be placed upon improving methodology or increasing the extent of participation. While these two foci are not mutually exclusive, they were seen to be applied by groups operating under different paradigms leading to polarisation with debates about evaluation. In this study I have also suggested that these continue to be areas of importance for consideration. However, when considering the evaluation literature presented in this review one of the areas that appears to have had less focus than others concerns the process of decision making, and particularly with regard to the decisions about the design of the model to be implemented. Focusing in on this area of the evaluation process is thought to be of importance in helping inform evaluators, commissioners and users in considering how demands are interpreted, what agendas are involved, which may hopefully lead to discussion why an evaluation might be used differently to original intentions, if it is used at all. A question might therefore be raised in regard to whether the design response matches the demands rendered. It is, of course, recognised that many factors will influence the way an evaluation will be designed. It has also been noted that it is difficult to “[establish] precisely when the decision process that results in the implementation of an evaluation process starts [as well as problems with following] the course of the decision” (Hansen & Borum, 1995: 322). While considering the limitations of researching decisions about evaluation, the focus of this study is placed on the way they are perceived to be taken, and the framework within which they are taken. This is also interesting in a time, as was outlined in the previous chapter with regard to quality assurance, when evaluation is becoming reframed.

There are many varied definitions of decision making. Simon offers a broad definition, including the recognition of and attention to problems, ascertaining alternatives, and evaluating, choosing and implementing solutions (1993: 394-5). Simon noted that decision making could be rational, non-rational or irrational. I will return to these concepts when dealing with decision process models in section 5.5. With regard to the models, some greater understanding is needed of how they can be employed to investigate these evaluation processes. In this current study focus has been rather placed upon how evaluators interpret the demands of mandators, and what decision responses are forthcoming. To achieve these ends this chapter considers research and theories concerning decision making processes in organisations. To inform this study I will first outline briefly the discussion that has taken place concerning this subject within



the evaluation field. This is extended by discussion of the decision making theories that have been applied, particularly in regard to the questions raised in their own organisations concerning the using multiple models to understand decision situations. Discussion is also made of how applying multiple models can be seen as offering a template of perspectives and concepts, and describe how this is thought to enable investigation in this study. As will be seen in the next chapter, a template is merely an initial starting point to frame an investigation that subsequently develops as the study proceeds based on the interpretation of data.

### **5.1 Making decisions about evaluation**

Whilst the evaluation field has not focused strongly on decision processes there has been interest from some commentators. A framework of particular interest to this study is drawn from Stufflebeam and colleagues writing in the early 1970s. This was interesting research. While the context in which they were operating has developed even more since the introduction of NPM, they raised issues that begin to recognise the importance of the perception of evaluation and the judgements made about it. While also attempting to offer a new definition of evaluation, the authors considered the role of the professional evaluator in supporting the decision process. In doing so they linked evaluation specifically to decision alternatives, defining it as “the process of delineating, obtaining, and providing useful information for judging decision alternatives”, where evaluation was still seen as value judgement for improvement purposes (1971: xxv). Within this work the authors outline the so called CIPP model developed mostly by Stufflebeam, that is, context, inputs, process and product, which was thought to provide better information on which to base decisions about programme improvement, survival or termination. The model is interesting in the sense that Stufflebeam and his co-authors consider how mechanisms of human decision making will affect the evaluation process and therefore how evaluators can organise their work to increase the focus and quality of their models and ultimately improve information flow. They saw the evaluator as acting on behalf of decision makers, implementing tasks to inform subsequent decision processes, describing the evaluator as:

“an extension of the decision maker’s mental process... negotiating each step of the decision process by working with him to delineate the information which is needed, by obtaining this information, and by helping the decision maker to use the information” (1971: 93).

While this approach is mainly related to large scale evaluation, observed against of educational programmes in schools and in addition referring to that performed by external evaluators, there are pointers that are relevant for this study in terms of focus upon the importance of understanding decision making. The authors pointed particularly to the problems of lack of supportive “evidence” available from evaluations, with evaluation being “seized with a

great illness” (1971: 4), the symptoms being avoidance of task, anxiety for judgement, unresponsiveness to demands, scepticism over utilisation, misadvice and incapability to perform adequately and other structural problems (1971: 4-8). While this list of problems is not exhaustive, it does highlight many of the major issues confronting the evaluation field at all levels, and still after over 30 years the issues raised appear to be very pertinent and seemingly requiring greater investigation. In describing the “etiology” of these problems, the authors noted five major issues: problems with the underlying definition of evaluation, understanding of the decision making process, the influence of values and criteria, the problem of administrative levels and research models.

In a recent study Tourmen (2009) focused on decision making with regard to programme evaluation, but more directly on the use of programme theories and how decisions might differ related to evaluation experience. The author’s study focused on individuals’ interpretation of evaluation situations, with regard to design decisions related to given terms of reference (Tourmen, 2009: 12). There are similarities between Tourmen’s intentions, based on her recently delivered PhD, and the focus of this study. Tourmen considered the importance of evaluator actions, goals and results observed, as well as their response to context and reasoning for choices (Tourmen, 2009: 14). At the same time, Tourmen’s study was more interested in individuals’ actions with regard to the application or otherwise of theory and the role of knowledge. Comparison was also made between the actions of experienced and beginner practitioners in terms of scope and focus of the evaluation. As will be described in the ensuing chapters, my study is more directed towards the decision processes in subunits, within one academic field and professional context, exploring the decision behaviour of groups related to internal and external frameworks and views of evaluation.

#### *Definitions of evaluation*

As was outlined in chapter 3, there have been numerous competing definitions of evaluation. The aim of this section is not to produce a definitive definition but rather to recognise that the definition or perception of evaluation by evaluators can be thought to frame responses to demands placed on them. In Stufflebeam et al.’s model of problems, the authors identified three broad types of definition: *measurement*, *congruence* and *professional judgement* (1971: 9ff). These categories are thought to be appropriate for analysis within this current study, in attempting to ascertain the way programme providers generally define evaluation and how this might influence their choice of design or model and how these views will be shared when decisions about evaluation are made. The measurement definition encompasses those holding a more instrumental view of data collection relying on reliable measurement tools and “instrument development”. (1971: 10). This is considered to limit evaluation only to variables that are already considered to have a degree of manipulability and be measurable (1971: 11).

The second definition of evaluation focuses upon the “congruence between performance and objectives; especially behavioural objectives” (1971: 11). This “Tyleristic” approach forms the basis for evaluative study attempting to discover changes in behaviour as a result of the implementation of educational programmes and the learning processes involved. Attention can be both process and product oriented (1971: 12). While many of the principles mentioned were a positive improvement for data, such an approach imposed “very narrow technical constraints”, focused more on global objectives about evaluation behaviours, where everything was assessed in terms of “effects on students” and their achievements (1971: 13). The authors argue that this approach also had often been applied to initiatives indirectly linked to student activity. Today with even greater demands for efficiency and accountability and subsequently increased focus upon student outcomes, such thinking appears to be at the root of the ‘impact’ movement. The problem raised continues; adopting behaviour as a criterion directs evaluation towards a post ante focus, becoming a “terminal process”, with decreasing focus on formative assessment.

The third definition of evaluation highlights the “judgmental process”, which seeks to explore how value is placed upon data, which the authors consider the previous two models to take for granted (1971: 13). In this case, however, evaluation is based upon and consists of “professional judgment”, where evaluations are built upon the perception of the group under investigation. While such approaches offer ease of implementation based on the analysis of experienced experts, reliability and objectivity are questioned. These reflections reflect to some extent the current situation seen across educational organisations today, especially with regard to assessment of models based on self-evaluation within quality assurance systems.

The authors subsequently redefine “[e]ducational evaluation [as] the process of delineating, obtaining and providing useful information for judging decision alternatives” (1971: 40). There appear though to be many tautologies within their construction regarding the link between evaluation and decision making. This is noted within their next major problematic area.

#### *Decision making*

The second major problematic area noted by the authors is that of decision making, specifically that there is lack of knowledge about it and its nature, lack of taxonomies to interpret it through and lack of methodologies that might be employed to improve use. Stufflebeam et al. (1971: 16) challenge a rational view of the evaluation process where evaluators are thought to identify goals, implement a strategy to investigate processes and isolate outcomes. They also observe the decision making process to be complex, describing it as developing through four incremental and hierarchical stages: awareness, design, choice, and action (1971: 50ff). While the authors tentatively forward the “disjointed

incrementalism”<sup>103</sup> decision model, outlined by Braybrooke and Lindblom (1963) as an alternative to a rational perspective, they are most concerned to point out that lack of understanding of the decision process has “hindered [evaluators] considerably in determining what evaluation methodologies are most productive and what kinds of information, under what circumstances, would be most valuable” (1971: 17). In addition, they recognise that decisions are not isolated events and take place within a tangled chain of processes that will need “unravelling” (ibid: 18). As will be observed in later sections of this chapter, this approach continues to build upon the idea that decision is regarded as choice, even if the parameters of that choice are bounded.

Part of this argumentation was later taken up by DeYoung and Conner (1982). Once again this was more rooted within the activity of external evaluation. In this work, however, there is an interesting focus upon decision making during the planning and designing phases of evaluation. The authors claim that an evaluator’s “implicit or explicit choice of a decision-making model has significant impact on the conduct and fate of the research, since it includes assumptions about how the research should be planned, implemented, and used” (1982: 431). The authors further suggest that the “evaluator’s *perception*<sup>104</sup> of organisational decision making influences the way he conducts research” (1982: 435). The work, however, is limited by the fact that they compare only two decision models: rational and incremental, and that both of these are closed systems theories of organisation. However, they do helpfully propose that “the way an evaluator designs and implements an evaluation, as well as the way he reacts to events during the course of the evaluation, is affected by the role he adopts”, suggesting that this “too often” occurs implicitly without “awareness of the operative decision-making model in a ... program or of the implications of his or her decision” (1982: 438). The authors imply three major issues. Firstly, particular evaluation roles can be adopted in relation to initiating design and choice of model in order to improve the process and subsequent utility. Secondly, these roles will be affected by the developing decision making processes and programme activity rather than the initial goals of the programme, and thirdly, the lack of understanding and awareness of the implication of their own decisions about how to evaluate, should make evaluators act more explicitly in their choices, especially adopting incremental approaches.

Therefore, choosing the rational model and focusing on goals is liable to result in failure to interpret events correctly, whilst choosing an incremental model is thought to make evaluators capable to understand and grasp the “intraorganisational power struggles”, presenting them the opportunity to respond to developing demands and promote change as necessary (1982: 435). However, having only one operational alternative to rational, goal following approaches does create difficulties, especially as the authors subsequently

---

<sup>103</sup> Outlined in more detail in section 5.5.2

<sup>104</sup> My italics.

appear to assume that incremental decision making has some kind of quality as a panacea. While the authors offered a more limited approach for understanding of decision processes within organisations, the strength of the argument appears to be in raising the issue of the importance of understanding evaluators' perception of the decision making process that influence them.

The limitations in Stufflebeam et al.'s model were more problematic. The authors in conclusion recognised that they "had not been able to identify an existing theory of decision making that... permitted a heuristic application to the evaluation problem" (1971: 331). They recognised that their own construction of the decision settings and process was limited, and that their CIPP<sup>105</sup> model, although useful, was based on taxonomy that they had not grounded fully in decision research. Therefore, while their problematic areas still appear pertinent, their model for evaluation requires further reflection. One area is identified here, seeing the importance of the decisions made about evaluation in response to the context of demands placed. In attempting to explore this a little further, this study will focus more on how reflections of organisational decision making apply to mainly internally led processes or self-evaluations.

Shadish and Epstein (1987) later recognised the generally limited number of studies of evaluator role and the influence of their choices over models. In their study of evaluator attitudes, which partly investigated demands, the authors recognised a distinction between "academic" and "service" orientation in evaluation, where the former is characterised by an interest in scientific development and theoretical understanding and the latter to provide data and decision making information for mandators. Developments since this time have seen a fusion of these categories, as mandatory demands purposed to ensure accountability have ensured that, for example, academics within HEIs previously considered to have academic freedom are faced with multiple and competing demands to evaluate their activity.

#### *Values and criteria*

The third major problematic area is that of values and criteria. This problem addresses the "implicit or explicit value structure" that is placed on the assessment of evaluation data and what degree of agreement there is concerning these (Stufflebeam et al., 1971: 18). They also raise a further interesting question regarding how the values of the evaluator will relate to their interpretation of the behaviour of practitioners under study (1971: 19), which has also been developed to challenge the idea that evaluative judgements can be value free and split from presentation of fact (House & Howe, 1999). Shadish and Leviton (2001) consider evaluation to be a question of a value judgement and maintain that such judgements are rarely rational. While individuals might on occasion be able to detect their biases, these are most often revealed when

---

<sup>105</sup> Evaluating Context, Input, Process and Product to make a value judgement.

confronted by an opposing viewpoint (2001: 184). The authors argue for greater inclusion of stakeholder values through a descriptive (views of others) rather than prescriptive (evaluator driven) process, at the expense of evaluator domination (2001: 187). This again is linked to the evaluator's relationship to or perception of the ultimate use of findings. Perhaps of even more significance though is the way that the evaluator's values are considered to influence the interpretation of the demands for information, as this will underlie the whole evaluation process. This is likely of course linked to the first area in terms of attitudes to measurement. The authors recognise that increased focus on accountability for public programmes demanded different responses from evaluators in terms of the models they initiated and implemented. There is also recognition that the way that criteria for an evaluation are selected will represent an arbitrary values system (Stufflebeam et al., 1971: 26).

The topic of this study focuses predominantly upon self-evaluation processes in a context of peer-review and external assessment. Vedung notes how these encompass "value criteria [that] are professional conceptions of merit" understood and accepted within a specific field, as quality criteria are too complex to develop more generically (Vedung, 2006a: 403). Vedung recognises the tension that arises as a result of such processes, in that criteria of "merit" and "performance standards" vary considerably across groupings (2006a: 404). Despite weaknesses, Vedung considers the approach to be more suitable than any other. The biggest challenge appears to come when the degree of complexity is played down and external mandators impose what are considered to be reductive frameworks of quality and performance standards, as are often considered associated with NPM frameworks (Vedung, 2006b). Such frameworks produce a general perception that evidence of achievement must be forthcoming, and where it is not, doubt must exist over the ability to deliver (2006b: 140).

#### *Administrative levels*

The fourth major problematic area is that of administrative levels (1971: 19). Here Stufflebeam et al. challenge educational evaluators on their overemphasis upon the micro level of observation and analysis. Such attitudes to evaluation can lead to problems with developing instruments, aggregating data and confusing purposes and processes of data collection. The strength of this problem area would seem to be in relating it to the underlying values and definitions of evaluation held by the evaluator, although in this case the emphasis appears to be on the preferred methods, skills and experience in performing the task.

#### *Evaluation and research compared*

The final problem area is linked quite closely, but considered to offer "the greatest challenge" and that is "overcoming the idea that evaluation

methodology is identical to research methodology” (1971: 22). This idea is taken up in more recent studies of evaluation models (Gulbrandsen & Stensaker, 2003; Krogstrup, 2006). This questions the constraints that can be placed upon the evaluation activity, that it should replicate scientific investigation, via experimentation and the assertion of assumptions and propositions, for example as found in statistical techniques (1971: 25). This area is interesting in relation to the context for this study as many of the informants are themselves involved in research activity within Higher Education Institutions. Stufflebeam et al.’s problematic areas are summarised within the table below:

**Table 2: Categories of problem related to evaluation decision making (after Stufflebeam et al. 1971)**

<b>Stufflebeam et al.’s category of problems</b>	<b>Focus</b>
Evaluation definition	How the underlying perception of definition of evaluation defines the model chosen to be implemented.
Decision making	How can knowledge about the decision making process improve evaluations and their utilisation?
Values and Criteria	What are the criteria by which evaluation data will be interpreted, and whose values weigh heaviest?
Administrative levels	What is the point of focus and level of analysis of an evaluation?
The research model	How is evaluation methodology different to research methodology?

Stufflebeam et al.’s reflections and subsequent model are framed as a response to demands for information from a decision maker to an evaluator. This normative presentation serves the purpose to highlight general deficiencies and suggest future action. As was also noted, Stufflebeam et al. considered the decision process to be a complex, but hierarchical form of incremental decision making. This was as a result of, like DeYoung and Conner (1982), focusing mainly on comparing incremental decision models with a rational decision approach. While the relevance of these models is not disputed, the open systems political and institutional models are considered to offer greater illumination of the processes at hand. These models will be outlined before being applied within the alternate templates for this study.

However, the main contribution to this study from the work of Stufflebeam et al. is in regard to the factors they outline that might be thought to influence the

evaluation process. As has been outlined, the five “problems”, consist of the influence of definitions, the decision process, values and criteria, different administrative levels and the challenge of comparison with the research model. In this study the focus is placed upon understanding one part of this process, where the responses to demands within the cycle of evaluation are considered. Thus, perception of the interaction between evaluators and mandators / commissioners initially appears to become a greater issue than focus upon how to meet demands. Therefore, focus is more upon the decision making surrounding the adoption and implementation of evaluation models. However, as will be seen in subsequent sections, the evaluation process is considered to be a complex interweaving pattern of events, which also challenges the proposition that the decision process is hierarchical and linear. An individual’s (or aggregated to group level) underlying definition of evaluation is also thought to be more strongly linked to issues concerning values and criteria in these initial decision phases as well as at the post data interpretation stage.

In recognising that the focus of this study is decisions made in response to a demand to evaluate, two areas are considered important to highlight before looking particularly at decision making. The first concerns how to understand the process of evaluation, which is outlined in the next section and the second concerns the perception by decision makers of the evaluation context, which will be outlined subsequently.

## **5.2 The evaluation process: “elements, actors and rationales”**

In the previous section it was recognised that interest in the decision making process has mainly focused on improving the post-evaluative decision process, that is a utilisation focus. It is proposed here that improved understanding of decisions concerning design of evaluations and the interpretation of demands would greatly enlighten this process further. Debates over the intrinsic nature of the evaluation exercise are bound to take place, if one accepts House’s recognition that the modern *profession* of evaluators “rests on collegial, cognitive and moral authority” (1993: ix). It is therefore important to recognise where the focus of this study is placed. This section briefly deals with a framework to analyse components within evaluation processes, particularly with regard to “elements and actors and rationales”, beyond a linear-rational view and acknowledging an inter-related process (Dahler-Larsen, 2004a).

### *Elements of the evaluation process*

Dahler-Larsen’s framework is an “analytical construction” which recognises that parts of evaluation processes will overlap, in both substance and chronology. The author chooses therefore to refer to these parts as “elements”<sup>106</sup> rather than

---

<sup>106</sup> My translation from the Danish word “momenter”, which can also be translated literally as “moments”, but I prefer to use the term “elements”, which seems to span spatial and content distinctions.



phases, considering 8 elements within the evaluation process (Dahler-Larsen, 2004a: 41). These are outlined in summary form in the table below. Within these elements there will be multiple decisions made, that are at the same time also influenced by those made across the different elements, reinforcing the idea of overlap. Elements of the evaluation process are therefore considered to be intertwined and recursive (Dahler-Larsen, 2004a; Dornbusch & Scott, 1975).

**Table 3: Elements of the evaluation process (after Dahler-Larsen, 2004a: 41-45)**

Elements	Description
Initiation	The point at which concrete initiative is taken to implement an evaluation, which can also begin to determine ownership of the process.
Agenda	Where purpose and themes for the evaluation to answer are decided and the evaluand is determined. This includes implicit and explicit standards and values to assess the evaluand against.
Knowledge management and organisation	Decision concerning “who does what”, as well as access rights to information during the process.
Design	Choice of evaluation model and methodology, a decision concerning “effects”. Influenced by credibility of designs as perceived by involved parties.
Data Collection	Point at which data is collected. Seen as a distinguishable from design phase, as decisions made and influence over the process at this point can afford distinct change.
Analysis and Summary	Selection, interpretation and presentation of results and findings
Validation	Social definition of validity of data, building on analytical choices but also on context, relevance and attention attracted from decision makers.
Consequence	Preface to legitimacy or response to information gathered, based often on recommendations from evaluators.

In this study I have mainly focused upon the first four of the elements suggested by Dahler-Larsen: “initiation”, “agenda”, “knowledge management and organisation”, and “design”. However, while I do not specifically investigate all of these processes, as they are considered to be recursive it is proposed that they will impact current and future decisions iteratively.

The “initiation” element involves trying to understand where the demands for an evaluation come from. Dahler-Larsen recognises that this can involve the

directions of a concrete mandator or be the result of a more general legal requirement (2004a: 41). Of importance would seem to be how movement from a general desire to a specific demand for an evaluation is interpreted by those responsible for implementation, as well as their perception of ownership, i.e., who the mandator is.

The “agenda” element refers to the juncture at which the purpose and the themes of evaluation are decided as well as how the evaluand is to be distinguished (Dahler-Larsen, 2004a: 42). It would appear important to attempt to illuminate how evaluators perceive the interpretative framework that is to be employed, and in what way these, as Dahler-Larsen suggests, might ultimately be reinterpreted at later stages. This supports the need to investigate the implementation of evaluations, in this case with respect to academic groups evaluating their own programmes, through a process focused exploration of their perception of events and how they have been involved in them and any decisions that have been taken.

“Knowledge management and organisation” involves the course of action of apportioning responsibility and ascertaining the degree of insight and influence for different parties (ibid.). It will be important to explore how evaluators perceive their degree of control over this process, as well as in what sense different mandators play a part. The “design” element of the evaluation process deals with how model and methodology are distinguished. As was noted in the summary in the table above, this involves understanding what degree of effects from a particular programme or activity are expected to be discovered and this will portray a particular image of an evaluand (Dahler-Larsen, 2004a: 43). The author also proposes that the credibility of certain methods and view of evaluation within a particular field will necessarily influence the choice of model and methodology. As was noted earlier in relation to the propositions of Stufflebeam et al., understanding of the informants’ values and preferences will be important in informing this study.

#### *The importance of context to evaluation decisions about programme impact*

Alongside understanding the values and preferences of the actors involved in choosing and designing models and implementing evaluation it is also considered important to understand the influence of context on these decisions. As we will see from the following sections, the context of the programmes under study is considered to be more complex than is assumed under simpler models of decision making that focus mostly upon goals, implementation and results. Understanding how the context might influence the “success” of a programme requires a wider interpretation of variables and factors thought to be of influence. However, perception of how context might influence the impact of a programme is also thought to vary under certain epistemological modes of thinking and values systems. When investigating the designing and implementation of evaluations of programmes it is therefore felt important to

gain an understanding of how evaluators perceive programmes can impact those taking them and how such potential changes might be measured, in addition to any wider expectations for the programme from stakeholders. Questions about context might therefore be as much about individual attitudes and values as they are about mapping out the particularities of the environment. The wider factors, or context of the programme, are then thought to influence evaluation designs. Respondents will be asked to reflect over the framework that their programmes are set in and their perceptions and responses to it. The major legislative and policy demands and structures that set the agenda for evaluation of postgraduate programmes for school leadership in Norway and England were outlined in chapter 2, in order to understand the wider context in which decisions are made within the framework of this study. This overview is considered against the responses of programme providers at the micro level, who are asked to relate their perception of the demands placed upon them and the way that they follow these up. There are of course many degrees of freedom when attempting to collate such information. But, I reiterate that the exercise at hand is an attempt to explore the decision making made concerning they type of evaluation to be implemented and to further develop a conceptual understanding of this process.

The next section outlines linkage between the perception of the context and how respondents view the possibility to ascertain programme impact in relation to it. Such a discussion further informs respondents' attitudes to evaluation, their responses to demands placed upon them and their deliberation over which models to implement.

#### *Activist and determinist perception of context impact*

Stake (1990) considers the importance of research into the “situational context” of evaluations and how that might influence both evaluation design and subsequent utilisation of findings, an area of focus thought to have been previously little studied, or even ignored. The contexts of educational programmes are considered to be multifaceted, encompassing the “temporal, physical, spatial, social, political, economic etc... [where] an educational practice has its habitat, its milieu, its frame of reference, its zeitgeist—not one but many contexts” (1990: 231-2). Stake goes on to recognise that part of evaluation is to illuminate what influence this context has.

Part of Stake's propositions is thought to be of particular relevance to this study. Stake forwards an idea of evaluators as activists and determinists, although he recognises that these descriptors are normative categories and that some kind of continuum would most likely explain attitude and behaviour (1990: 241). In a simplistic summary, activists will see the potential for programmes to impact strong change over their context, whereas determinists will reject such a possibility and see context as dominating. As Stake intimates, it is unlikely to find many evaluators of a strongly deterministic persuasion as such a position would be oxymoronic to the task ahead of them. However, evaluators might be

disposed to one or the other of the viewpoints which will in turn affect their approach to evaluation design.

While Stake (1990: 241) appears to focus primarily on the perceptions of external evaluators to the “control potential” of providers over their “program destiny”, it would seem that this idea also bears much wider relevance in terms of current thinking in evaluation, and particularly to the context of this study. Writing in 1990 Stake mapped out a key issue as the evaluator’s attitude to the concept of ascertaining programme impact, extending his idea of activist and deterministic perception. Impact is an expression which appears to have been reinvented or gained new vigour under NPM in recent years. Those of an activist persuasion will “over-dignify the concept of impact and set unrealistic standards for success”, which Stake claims appears to characterise “Western thinking” with regard to programme evaluation, contributing to “desired change” (1990: 241). Those of a more deterministic persuasion will “devote too little of the design to the discovery of effects”, focusing rather on process description, “elegance of purpose... quality of arrangement... covariation of endeavors, and to the intrigue of the story... [where] [a]ccomplishment may be treated as ephemeral and without substance” (1990: 241- 2).

The significance of these descriptions is not so much in terms of how evident extremes on the continuum will be. Stake recognises that these “might degenerate into mere tendencies of strictness and leniency” (ibid) in evaluator personality, but that it is also important to attempt to illuminate the underlying the epistemological worldview of the evaluator towards the idea of ascertaining programme effects and impact. As I have already pointed out, Stake appears to focus particularly on the discussion regarding the supremacy of situational context over programme delivery with regard to external evaluation. In this study the attitudes of those evaluating their own programme delivery are under investigation. While it is not fully clear to what degree the categories that Stake proposes, i.e. activist and determinist, are suitably efficacious descriptors to use for respondents, it is still considered relevant to follow up Stake’s charge to the evaluation field and investigate programme providers’ attitudes to accounting for, in this case, some semblance of programme impact. This focus appears even more relevant with regard to the recent discussion concerning the increasing interest from mandators and commissioning bodies in being presented with evidence of programme impact, also evident within the context of this study of school leadership development. It is recognised though that this might be more indirect with regard to HEI programmes compared with nationally mandated courses and training programmes. HEIs in Norway have developed programmes for local mandators to which they must report back to, but as will be seen in later sections, these vary both in their demands and competence in analysing data.

### **5.3 Organisational decision making theory applied to evaluation**

It is thought necessary to outline a brief overview of the theories that underlie models about and approaches to decisions in organisations. To investigate decision processes within the context of this study the main models drawn from these different research fields are considered as templates which guide both the data collection and analysis (King, 1998, 2004). This process will be outlined more fully in the next chapter. The rationale behind this approach is that while the study of organisational and decision making theory has been widely applied, there has been less impact on the field of evaluation (Dahler-Larsen, 1998). It is also considered that no single model will offer a complete explanation for the investigation of such a multi-faceted and complex process (W. R. Scott, 2001, 2003), especially when the investigation is based on the perception of those involved in such a process, even though it is recognised that there is a challenge in using multiple models with regard to implications for interpretation (Pfeffer, 1981b: 29). The considered implications of these various theories for this study and a wider view of evaluation are also outlined.

As has already been stated, there are two major issues that are raised by these reflections, firstly that evaluation is itself a rational process and subsequently that decisions about an evaluation will take place within some kind of rational process. As such it is thought important to consider challenges to the concept of rationality. In considering this it is not only the concept of rational decision making that is being challenged, but partly the very concept of a decision itself and whether this can be investigated. This is recognised by Christensen, who proposes that:

“The idea of decision is a theory. It assumes a connection between activities called the decision process, pronouncements called decisions, and actions called decision implementations. The decision process brings together people, problems, and solutions and produces a decision. The process may involve problem-solving; it may involve bargaining, it may involve some system of power. Whatever the mechanism, the process generates an outcome. That decision, in turn, is converted into specific actions through some variation of a bureaucratic system” (S. Christensen, 1979: 351).

Christensen frames the decision process as consisting of more than merely a “decision”, where problems might be specially constructed or ignored and decisions might never be implemented. At the same time, the process of decision making will be separated from the outcome. Drawing on Olsen’s work, Christensen notes that decision making can often be seen as a “ritual act” (1979: 383). This can be linked to Dahler-Larsen’s (1998) idea of evaluation as ritual reflection, where process is little linked to outcome. The life of the programme seems to develop independently of certain structures and coupling is tighter in the lower segments of the organisation. This will be dealt with further in

subsequent sections. Christensen's research additionally noted the importance of ideology and culture as driving the form of decision process (1979: 352).

Brunsson (1982, 1990, 2007) also challenges the perception that decisions are merely about choice, recognising widespread elements of irrationality within the decision process. Brunsson (1990) recognised three additional roles of decision processes in addition to the making of choices: mobilising organizational action, allocation of responsibility and providing legitimacy. Brunsson notes how organisations often struggle to achieve "collective action", and will use decision processes to cement commitment to a desired or planned activity (1990: 48). The result might be a limited or "biased" set of alternatives to be considered or committed to. Within these irrational processes there is less demand for data than in the rational, predictive designs, the former being more planning oriented (1990: 49). This might conceivably also be a condition imposed upon those further down the system.

Another complexity in assuming decision making to be about making rational choices from alternatives, concerns the allocation of responsibility (Brunsson, 1990: 51). According to Brunsson, rational theories place the responsibility upon decision makers as the cause of events, carrying out an intended action. In such interpretations if decisions are assumed to be about making choices then decision makers can be identified and held responsible. However, decision makers may have less or more influence upon events than is believed or anticipated depending upon how they execute their role and the choices put in front of others (1990: 50). Responsibility can be won as well as shunned, decisions can be made more or less visible and decision makers can attempt to show that they had no choice (1990: 52). One implication of such an approach is the importance of understanding how the "values, beliefs and perceptions" of decision makers inform the choices offered; the type of decision process that unfolds and the responses that ensue (1990: 51). In this study, discussion concerning the purposes of evaluation and premise of the programmes will be under focus in the various subunits, as well as how the evaluation process is thought to be part of the wider organisation and the degree to which the groups have control over the process.

While Brunsson's study was focused upon "decision-oriented" organisations, for example councils and boards, the sub-sites under study here are academic programme provider groups, particularly those responsible for their design and evaluation. These groups are considered interesting, as the part of the basic premise of these programmes is to focus upon decision-orientation and evaluation in others, in this case the active reflection over school leadership and the school leader role and associated processes. There is a sense in which programme providers are asked to consider how their own activity reflects the values emphasised within their own programmes.

Some recent studies have attempted to link evaluation research and decision theory. Hyyryläinen and Viinamäki (2008) attempted to account for eventual evaluation utilisation the “demand side” of evaluation, by considering decision makers views in relation to “rational”, “non-rational”<sup>107</sup> and “boundedly” rational (sic) decision models. In attempting to consider the different models in relation to utilisation, the authors question the assumptions of rationality related to decision makers, as well as the decision making process. The authors noted that rational models continued to be championed by evaluation theorists and practitioners, accentuating problems and searching for solutions on the “supply side”, while non-rational models focused on problems between “decision makers and other stakeholders” (2008: 1236-7). The authors concluded that the “Boundedly Rational Model” offered a solution by “lowering the expectations of rationality of decision makers”, while also recognising that not all information will be used by decision makers who are at the “core of the analysis” and who act non-rationally from time to time (2008: 1237). This appears to be a simplification of the process and models, suggesting that information is used, not used or only partly used. With regard to the process, some greater understanding of the recursive and complex interaction within organisations needs to be understood.

Part of the problem is that evaluation methodologies are considered to be more often rooted in education theory rather than organization theory and as a result activities are thought to be rarely framed with improvement of decision-making in mind (Holton III & Naquin, 2005). While the relationship between evaluations and decisions has received attention, much of this research has focused on the subsequent utilization of information. In essence, research has focused on decision maker / stakeholder values and how these might be incorporated into an evaluation to increase the relevance of findings and likelihood of use within the subsequent decision making process. While this is not an unimportant phenomenon, it is argued here that too little focus has been placed on what shapes the decision to evaluate, especially with regard to the underlying values of those deciding how the evaluation will take shape and their intentions for utilization. This may become more evident when programmes are evaluated internally (Love, 1991, 1998), especially when the results are sought after by external mandators. This can also challenge the “romantic myth” that internal self-evaluation is merely a positive, learning focused, democratic process, compared to external evaluation which is negative, control focused and undemocratic (Dahler-Larsen, 2006b: 83). Looking at the purposes of evaluation, the chosen models and the events surrounding implementation are considered of great importance. Questions are raised as to whether there is a mismatch between internal and external demands, and in what way they affect one another and what kind of data fits what demand. Recognising the complexity in the decision process concerning evaluation is thought to offer the field a different vantage point in the search for improved practice. As has

---

<sup>107</sup> Apparently similar to Ambiguity models.

pointed out in relation to the varied works of Greene (1988; Greene & Walker, 2001) and the “Danish school”<sup>108</sup> consisting of Dahler-Larsen (1998, 2001), Hansen (2001, 2005a, 2006) and Krogstrup (2006) such complexity requires insight from the wider field of organisational and decision theory. In the next section I also consider how decision making has been considered within the higher education field.

#### **5.4 Decisions and evaluation in higher education**

As was noted in earlier chapters, this study investigates decisions about evaluation partly in relation to quality assurance systems within HEIs in England and Norway. It is considered therefore important to outline briefly some of the limited examples of linkage between evaluation and decision making within the field of Higher Education. During the early 1990s increasing focus was placed upon the assessment of higher education. This was noted especially within OECD countries and an ensuing report highlighted the role of evaluation as part of the institutional decision making process (OECD, 1994). Once again the focus becomes one of utilisation of findings but there was no registered focus upon decision making about evaluation.

One interesting comment in the findings comes in relation to the status of evaluation noted in the final section of the collection of OECD reports, where it is suggested that evaluation’s “purpose is to inform and clarify decision making, yet it is not itself a decision making process” (Cazenave, 1994: 201). In this study it suggested that this position requires greater nuance, especially with regard to how decisions are intertwined throughout the whole evaluation process and not just with regard to making a decision in relation to the data provided from it. In the next section I consider how the area of decision making has generally been applied within research in HEIs.

##### *Application of Decision theory within Higher Education research and practice*

If the application of Organisational and Decision Theory within the field of evaluation is claimed to have been relatively limited (Holton III & Naquin, 2005), then to some degree the opposite appears to be the case within the context in which this study is based. Higher Education provision has observed a great deal of the development of Organisational Theory, especially related to different models of decision making and the analysis thereof (Musselin, 2002). Musselin notes how application of decision theory within the field, in addition to the use of the Higher Education field to develop such theory further, led to the development of different decision models that were then further applied within research into HEIs. This suggests active interaction. The ensuing models from

---

<sup>108</sup> There is no particular recognised school as such, but I have noted with interest the contribution made by these scholars towards widening the evaluation debate in this direction.



such research varyingly encompassed collegial, bureaucratic, political and ambiguity perspectives of decision making. An example used in this study can be found in the work of Hardy et al (1983) outlined further below.

Musselin (2002: 1) also outlines how the different models that have been developed, particularly during the 60s and 70s, have tended to concur and attempts have been made to combine them, relative to the domain under study and the understanding of how they may interconnect. Musselin further reflects that application of these models has both aided and restricted research. It has aided research by suggesting “ideal types and thus [giving] clues to apprehend and to reduce the organisational complexities of universities” but has restricted by leaving the “analyst to [decide] on the right type in which to place the university under study instead of opening the black box”, which may also be problematic if the models do not match the evolution of the context under study (2002: 1). The author suggests that focus ought to be placed on exploration of specific aspects within the organisation of higher education, rather than the development of new models or qualification of decision patterns (2002: 2). As has been outlined earlier, the intention of this study has been to look at further factors that might influence the decision process concerning the development of models for evaluation, while attempting to apply the various decision models and as such illuminate the process further. In doing so, it is hoped to reflect further upon Musselin’s suggestion. In agreement with Musselin, I adopt Scott’s view, outlined earlier, that none of these approaches offer a full and final position from which one can interpret such processes, but as Musselin also alludes to, certain models will offer greater explanatory power depending on the domain.

Hardy et al (1983: 411-2) adopted Mintzberg’s view that HEIs correspond to a “professional” rather than “machine” bureaucracy, reflecting the complexity in rationalising mission and activity as well as the relative looseness of coupling and decentralisation of decision processes and necessity of specialisation at the base level. The authors recognise 3 interlinked levels of decision making control within HEIs: at the individual academic level, by central administration and by the collegiality. These are variously characterised by decisions based on professional judgement, administrative fiat, and collective choice, the latter of which the authors see can be categorised further within collegial, political, garbage can and rational models (1983: 412) the latter of which Hardy later noted to include bureaucratic perspectives (1990b). These processes can be observed in figure 7 below. The authors’ reflections are interesting in relation to the framework of this study. There is recognition that many decisions are taken at the level of the individual academic based on their professional judgement. Professional judgement is a category building on the identity of the individual academic. Hardy et al. recognise that individual academic staff (professors) have traditionally had “a great deal of autonomy over research and teaching because of the difficulties of supervising or formalizing this work”, thus they see decisions related to “basic missions” controlled at the individual level (Hardy et

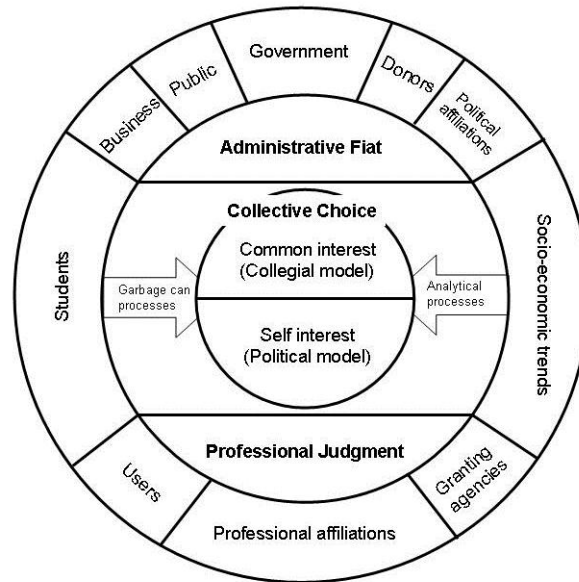
al., 1983: 412). While this autonomy in recent years has come under threat, there is still some degree of the individual professor controlling their own areas. The authors claim that this is to do with the principles of “pigeonholing”, whereby programme responsibility is decentralised and the overall portfolio is loosely coupled (Ibid: 413). The authors recognise that this does make programmes free from a degree of “external influence”. At the same time programmes are shaped by a standardisation process rooted in the occupational socialisation of the responsible academic. Interestingly the authors noted the increasing importance of student feedback and the influence this could have on programme focus. This can also be linked to Stufflebeam et al.’s reflections over professional judgement as outlined in section 5.1. There is an important overlap here with the role of central administration.

The third category is interesting as it relates to decisions made by “collective choice” (1983: 417). Hardy et al see these decisions as evolving “out of a variety of interactive processes that occur both within and between departments”, between academics and administrators and across levels. The authors outline what they see as three “phases of interactive decision making”- identification, development and selection. Similar to the elements of evaluation described by Dahler-Larsen (2004a), Hardy et al describe processes that do not take place sequentially but are complex, developing cyclically and in “interrupted” fashion. These processes are outlined in the table below based on Hardy et al.’s findings.

**Table 4: Phases of interactive decision making (Hardy et al., 1983: 417-418)**

<b>Phase</b>	<b>Description</b>	<b>Actors involved</b>
Identification	Where the decision need is recognised, and situation diagnosed. In time some decisions will arise out of routine.	Individual initiative
Development	Alternatives and solutions developed. Progress related to power relationships between academic and administrative staff.	Ad hoc groups / task forces
Selection	Screening, evaluation, choice and authorisation of development phase. This will take place at various levels and can be a “cumbersome” process.	Hierarchy of permanent groups and administrators

The authors see the collective process as combining “collegial and political processes, with garbage can influences encouraging a kind of haphazardness on one side due to cognitive and cost limitations... and analytical influences on the other side encouraging a certain logic or formal rationality...” (1983: 423). These are observable in the figure below.



**Figure 7: Three interlinked levels of decision-making in the University (Hardy et al., 1983: 414)**

Following up this research, Hardy later adopted a framework drawing further upon Blau's (1973) "coexistence" perspective, where bureaucratic and academic/professional features of discipline and innovation are juxtaposed (Hardy, 1990b: 208). The intermingling of these forces is combined with a decentralised form of power and responsibility. These propositions are found to be interesting in relation to this study as even though professional values are thought to drive the work of academics, there is still a sense of bureaucratic standardisation in the way study programmes are carried out related to the particular professional socialisation (Hardy, 1990b: 209). In addition, sub units "afforded an autonomous, loosely coupled (Weick, 1976) existence" are "hooked up to the larger organisation via hierarchy" and this takes place through institutionalised processes (ibid.). Drawing further on the work of Beyer, Hardy discusses the relationship between bureaucracy and collegiality, where collegiality is considered to be decentralisation at the departmental level, where academic faculty members have a high degree of influence, and bureaucracy as departmental centralisation, where faculty members have less influence compared to heads of department (Hardy, 1990b: 209). These dimensions are also compared to the relative autonomy from central administration influence. Hardy notes that while there has been a great deal of research into the structural characteristics of HEIs, accepting the proposition of similarity to Mintzberg's professional bureaucracy, this research has had "little to say about how decisions are made within this context" (1990b: 210). It is therefore considered important to relate this research to others models of decision making within organisations.

This is thought necessary to address the issue of decision making with the empirical study, framing the perception of respondents of how the subunit sees organisational processes developing and their place within that framework.

### **5.5 Models of decision making and decision processes in organisations**

In this section I outline different decision making theories that have supported models for use in organisational analysis. These models are thought to be useful in analysing the processes of designing and implementing an evaluation. While the theories outlined are considered generic, I believe they have explanatory power for this study. While drawing more widely on organisation theory to map the development of the approaches outlined here, I focus on they have been adopted into research models and analytical frameworks by Allison (1969, 1971; & Zelikow 1999), Peterson (1976), Pfeffer (1981b), Hardy (1990a, 1990b; Hardy et al., 1983) and Dahler-Larsen (1998). These models are thought to fit within a wider institutional framework as described within Thompson's macro-organisational view (1967, 2003) and, later by Scott (1995; 2001, 2003). As was outlined earlier, Dahler-Larsen (1998, 2001, 2006b) applied organisational theories more specifically to the evaluation field, while Hardy's work builds upon research into strategy formation within the HEI context using the models for analysis purposes (Hardy et al., 1983). A pertinent point for this study is how the authors were particularly focused upon "realized" rather than "intended" patterns of (or positional) decisions and actions, implying a search for consistency in the responses. Thus rather than taking isolated decisions and actions as examples of the process, an attempt was made to consider patterns, longitudinally, in the approaches made. As will be seen, this was thought to bring a particular challenge to the interpretation of theories of ambiguity (1983: 408).

Part of the purpose of outlining the theories the models are built upon relates to the challenge outlined above that decisions are not necessarily rational but are thought to be varyingly influenced by individual and organisational characteristics. This is coupled to similar challenges within the evaluation field in terms of process of development and implementation. It is therefore considered wise to further connect research on evaluation and its utilization with decision theory. Applying decision making models should deepen understanding of other factors influencing the evaluation process in organisations, besides the already mentioned foci of improving the quality of information (Weiss, 1972, 1982, 1998b) or involvement of stakeholders and users (Patton, 2003).

The major models are presented that build upon wider decision-making research, from rational to naturalistic approaches, each of which is thought to be important when attempting to understand evaluative processes. The implications of using these models, or frameworks, are considered further in section 5.8.

### 5.5.1 A rational approach- “action as choice”<sup>109</sup>

A rational theory of choice is generally characterised as “instrumental behavior taken in order to achieve desired ends” (W.R. Scott, 1995:138), where organizations are regarded as “instruments designed to attain specified goals by organizing a series of actions in such ways that they lead to predetermined goals with maximum efficiency” (Mathieu et al., 1992:33). This is based on the assumption that human behaviour is rational, and that decision-makers are assumed to enter decision situations with clear and known objectives, enabling them to make thorough analysis of the external environment (Allison, 1971; Das & Teng, 1999; Pfeffer, 1981b). Theoretically, rationality has been emphasised as the most essential condition in making decisions, in what is seen as a comprehensive, normative process in which individuals gather information, develop alternatives, and then objectively select the optimal alternative. Problems are evaluated in relation to stable goals and optimal courses of action chosen from a set of alternatives (Pfeffer, 1981b; Weick, 1976). Decision-makers strive to do what is best for themselves or their organization (Beach & Connolly, 2005). It is assumed that decision processes are consequential and preference-based (March & Heath, 1994); in the sense that action depends on anticipating effects of current actions and that consequences are evaluated in terms of personal preferences. The “economic man” operates within the rational organization making optimal choices in a highly specified and clearly defined environment, simultaneously ranking all sets of consequences from the most preferred to the least, and finally selecting the alternative leading to the preferred set of consequences with value-maximizing effects (March, Guetzkow, & Simon, 1993; March & Heath, 1994; Newell & Simon, 1972).

As research on organizations developed it became clear that rational models needed moderation. Research on decision-making was initially carried out to prescribe what should be done, rather than to ascertain what decision-makers *actually do*; as was also noted by Scriven (1991) concerning evaluation. In spite of their utility for predictions, pure versions of rational choice are hard to accept as credible descriptions of actual individual or organizational agency. Second generation research began to focus more upon how decisions are actually made in organisations (Beach & Connolly, 2005). Expected utility became tempered by the belief that people hold irrational preferences, and that they behave inconsistently across similar situations (March, Guetzkow, & Simon, 1958), where not all alternatives are known, nor all consequences considered, nor all preferences evoked at the same time (March & Heath, 1994). The core idea of this ‘bounded rationality’<sup>110</sup> (March et al., 1958) is that individuals are “intendedly rational... [but] constrained by limited cognitive capabilities and incomplete information” (March & Heath, 1994: 9), but also stemming from the

---

<sup>109</sup> The section titles 5.5.1, 5.5.2 and 5.5.3 are drawn from Allison (1999).

<sup>110</sup> For a fuller description of bounded rationality see Simon (1997) and Gigerenzer and Selten (2001).

external constraints and limitations of allocated time and resources in the individual's environment (Das & Teng, 1999)<sup>111</sup>. As a result actors tend to 'satisfice', or reduce the information-processing load by selecting the first alternative option that meets the minimum standards of the decision maker. These are discovered sequentially and result in "repertoires of action programs" which operate as alternatives in "recurrent situations" (March et al., 1993: 191). It is also considered though that due to the division of labour in organisations with regard to decision making, that greater rationality can be achieved by increasing participation and delegating with regard to specialism (Hall & Tolbert, 2005: 128). In the case of evaluation it might be thought that division of labour throughout the organisation would lead to the design of models suitable to produce answers to means-goals questions. It is also suggested within the discipline of HRD that bounded rationality is a little understood phenomenon, especially when considering the complexity of the decision making setting and processes (Herling, 2003). Modern naturalistic decision making theorists have developed Image theory, in which the process of screening for decision alternatives is thought similar to the concept of satisficing (Beach & Connolly, 2005: 165). This adjustment of formally rational models, and particularly Simon's (1976) work on decision processes which recognised distinct sub-processes, has been criticised as continuing a linear understanding of decision processes (Langley, Mintzberg, Pitcher, Posada, & Saint-Macary, 1995). Simon refutes these criticisms, considering that sub-processes like agenda setting, problem representation and alternative searching are identifiable but do not form any particular set order and are most often recursive (1997: 127). A question is raised about how one can aggregate further to link to this evaluation purposes. There might be many reasons why evaluations are not used in organisations. Of course the point here is to try and understand how decision making about evaluations might be affected by the way the organisation collects and uses information.

Rational theories were adapted into an organisational decision making model by Allison (1969, 1971; & Zelikow 1999). The 'rational actor' model focuses on action as deliberate choice, assuming a *logic of consequence*, focused upon a unified actor<sup>112</sup> acting rationally with one set of preferences, perceived choices<sup>113</sup> and a single estimate of resultant consequences (1999: 24). The response to an action will therefore be to question why the actor acted like they did, drawing inference from the macro context at hand. Within this model, analysis will often focus on the context, judging whether the behaviour was truly

---

<sup>111</sup> Bounded rationality connects with rather than opposes rational and psychological models but offers a description of the decision process rather than merely the outcome of it (Gigerenzer & Selten, 2001).

<sup>112</sup> Albeit in Allison and Zelikow's model a 'national' one.

<sup>113</sup> Peterson (1976) views this statement as problematic, and illustrative of a confusion over the nature of rational behaviour that led Allison (1971) to reject rational models too quickly.

value maximising or not. In its simplest form, which Allison has been criticised for focusing too heavily upon, the model links purpose to action, and behaviour is considered to be a consistent, value maximising choice within specified constraints. Pfeffer argues that in order to determine whether organisations are “best described in terms of the rational model” processes must be investigated, rather than observing decision outcomes which will be “consistent with rational choice, if the appropriate goals and preferences are assumed” (1981b: 21). Therefore this study will focus on perceptions and descriptions of decision processes by organisational members rather than describing particular outcomes and attempting to trace goals set. Allison did not consider this model to offer the best explanatory power, but he appears to have applied much of the theory on bounded rationality and satisficing in his second model, outlined in the next model explanation.

Dahler-Larsen considers the approach to evaluation within organisations adopting the rational model, recognising an instrumental use focused upon gathering information in relation to the “continuance, discontinuance or adjustment of particular activities<sup>114</sup>” (1998: 31). In the rational model the information gathered from evaluations can be followed up with certainty, or as Dahler-Larsen considers, mechanically. The author recognises, however, limitations to information gathering, dealt with more fully below. Dahler-Larsen notes that classical approaches to evaluation have often built upon a rationally based model with its emphasis on following up clearly defined preferences (1998, 2005b, 2006b) and measuring quality (2007, 2008), but interestingly there has been increasing attention in recent years to such models following the growth of the evidence movement. Hardy (1990b: 211) notes that “the bureaucratic features to be found in universities reflect administrative rationality” and the attempt to maximise effectiveness, which are more related to the ideas of administrative fiat within a closed system. To some extent Hardy combined rational and bureaucratic models; however these do not address the same issues, the latter dealing more with rule-based behaviour and incremental change. The bureaucratic model is outlined in section 5.5.2 and would seem to fit more closely to Hardy’s idea of administrative fiat.

#### *A particular challenge to rational models: information in organisations*

One limitation of rationality could be that available information in an organisation is systematically incorrect (Feldman & March, 1981). Information might not be gathered for making decisions, but rather as a flow that enables organisational members to make meaning of their surroundings. Organisations gather more information than they use, and they do not systematically use all the information gathered when making decisions. Organisations tend to credit information gathering, and therefore the status of the gatherer is increased and decisions are legitimated, even if they are not properly based upon perfect

---

<sup>114</sup> My translation from Danish.

information. Information might be gathered in an instrumentalist perspective, as a means to increase the legitimacy of an individual's role and activity (Feldman & March, 1981). In these cases, decision making becomes as much based upon intuition as it does on cognition (Beach, 1990: 25). Representations of the decision process are reduced to only the most necessary information, upon which the final decision is then based (Beach & Connolly, 2005), or the chooser's "definition of the situation" (March et al., 1958).

Greve (1995) recognises that although the gathering of information can have symbolic value, the process itself can be considered important and can encourage the involvement of people across the organisation. While this appears to apply generically to decision making, one can see links to the evaluation process. Evaluation appears to a lesser extent to reflect the rational, evidence obtaining activity that it can be purported to be and yet organisations continue to demand it and members continue to contribute to it. The mere act of doing it appears to gain some symbolic value. The fact that it should encompass both quantitative and qualitative methodology creates further interest in the process. As Feldman and March (1981) noted, the production and focus on statistical data often underlies the demand for evidence of objectivity. The authors note that there is often misunderstanding regarding the nature and purpose of information (1981: 175). Organisations cannot process the information they have; the organisation, as well as those working in it, is limited. The authors also claim that the types of information gathered are often just those that are available, and may not serve the purpose to which they are used. Another part of their argument is that the information that is gathered is done so in a "surveillance mode rather than in a decision mode" (1981: 175), the idea being that it does not appear to have any "apparent immediate decision consequences", but rather monitors the environment for the existence of any surprises, or lack of them (1981: 176). This activity is not necessarily deemed to be formal, based on explicit calculation, but rather operates more haphazardly as the gathering of gossip. There is a sense in which it would appear to reflect the way evaluation systems develop within organisations, with an eye on the demands of the watching world and an eye on the internal responses of participants.

The possibility that evaluation may be used as information to provide evidence of rationality and good management has come under increasing interest (Weiss, 1998b). Greene and McClintock (1985) report on a study into the nature and role of information used in program development of adult and community education. The study is reported as focusing upon whether "information gathering, exchange, interpretation and reporting met needs for program decision making and accountability", a framework which the authors say was drawn from literature on evaluation utilization and organizational decision making (1985: 528). The authors noted how the information needs of stakeholders might be influenced by "perceived and actual models of decision making models in the organization" (1985: 529); in this they drew upon Allison's model. They also drew upon Feldman and March to consider how information was used in the



organisation and Maynard-Moody and McClintock's (1981) framework for considering the degree of certainty towards programme goals and their attainment. There are interesting similarities to this study; the main purpose of their approach was to look at how information was used in program development and decision making focused across a sample of stakeholders, but in this current study focus is specifically placed on how the models of evaluation are developed and how the groups implementing the evaluation respond to their context, which includes internal and external demands as well as their own approach. The question though remains, how do groups make sense of these demands in a complex environment? This question will be dealt with more in relation to the subsequent models presented, but first I turn to the organisational behaviour model another closed system model.

### **5.5.2 Organizational behaviour – “action as output”**

Beach and Connolly (2005: 126) refer to this as the structural model, where constraints on information are felt to be compensated by structuring the organisation such that task and expertise are matched. Division of labour can lead to organisations being able to perform tasks that individuals cannot do alone (Allison & Zelikow, 1999). Such activity develops organisational policy, leading to “efficient, but inflexible” behaviour, which is predictable and stereotypical (Beach & Connolly, 2005: 139). Rule based action will also constrain the organisation at the same time as it constructs a “unique identity”. The organisation performs to a logic of appropriateness rather than that of consequence observed in the rational model (Allison & Zelikow, 1999: 154), where “neither preferences as normally conceived nor expectations of future consequence enter directly into the calculus” (March & Heath, 1994: 57). Rule following is a form of systematic reasoning based upon recognising the decision situation, comparing it with the identity of the individual and organisation making the decision<sup>115</sup> and questioning the normal rules of behaviour, that is standard operating procedures (Ibid.). At the same time open systems models later proposed that organisations will be constrained by their environment and must adapt to change, leading to a propensity to homogeneity via isomorphic behaviour (Allison & Zelikow, 1999; DiMaggio & Powell, 1982). An example might be that QA systems generally are much the same. This recognition needs of course to be tempered by the knowledge that there are varying and shifting demands outside of that in relation to different impact / funding demands. The great challenge is to design organisations that operate efficiently with a minimum of information, in order to avoid overload.

However, where organisations exhibit looser control change often tends to be incremental (Beach & Connolly, 2005). Incremental theories have affected the

---

<sup>115</sup> March and Heath recognise that actors have multiple identities, while organisations can exhibit tighter and looser forms, where learning cycles are often incomplete (reflected also in Scott's (2003) idea of rational, natural and open systems). This is further outlined below in relation to ambiguity models.

evaluation field more than any other decision making approach, seen primarily through the work of Lindblom and Weiss. Weiss refers to decision accretion, the “build-up of small choices... the gradual narrowing of available alternatives” (Weiss, 1976 in Alkin & Christie, 2004: 29). This is similar to Lindblom’s (1988) description of ‘muddling through’, where in complex processes there is little possibility to derive a single utility function, the impact is too hard to measure, information about consequence is limited, and options are often unknown (Beach & Connolly, 2005). Administrators must often act without clarification and subsequent knowledge of objectives (DeYoung & Conner, 1982). This contributes to the assumption that “rational evaluation based on normative theory is impossible”, and policy is built more by ‘incremental’ change of failures in the status quo, where policy is shaped more out of what is to be discarded than what is to be attained (Beach & Connolly, 2005: 148ff). Ignorance of these processes may offer a more negative tinge to Patton *et al.*’s reflections that “program development is a process of ‘muddling through’ and evaluation is part of the muddling” (in Hofstetter & Alkin, 2003: 205). The model was also developed to outline a focus on action and implementation as well as on thought (Beach & Connolly, 2005). Connolly observes this taking place within a decision cycle, a combination of the cognitive and emotive schema of a decision maker that intermingles with the perceptual environmental schema to “dictate goal-directed action” forming ‘consequences’ (Beach and Connolly, 2005). Thus decision-making is observed to take place in and be affected by the environment or context. This appears to bridge the perceived gap between incremental evaluation and implementation. Such recognition would also strengthen the need for more descriptive and naturalistic forms of research that recognise the role of the individual within their organisation and environment.

These theories were adapted by Allison (1969, 1971; & Zelikow, 1999) into a second decision making model. The model’s “explanatory power is achieved by uncovering the organizational routines and repertoires that produced the outputs that comprise the *puzzling occurrence*<sup>116</sup>” (Allison, 1969: 702). Thus decision is explained as the “by-product of organisational behaviour” where “capacities and constraints [are emphasised] both in choice and implementation” (Allison & Zelikow, 1999: 386). The basic unit of analysis observes action in relation to output resulting from organisational routine, which is visible within previously established standard operating procedures (SOPs) which provide cues and where “existing organizational capacities” denote the range of options and define and confine decisions (Ibid.: 164ff). Actors are perceived as being loosely coupled groups within a larger organisational setting, operating to an organisational mission statement that defines the area of activity and objectives to accomplish<sup>117</sup>. Allison’s model has been challenged for its explanatory power,

---

<sup>116</sup> My italics

<sup>117</sup> This is close to Halperin’s organisational ‘essence’ where the viewpoint of the dominant group guides the mission and capabilities (1974, in Ibid.)

which should be taken into account when used in analysis. For example, Peterson (1976) saw this second of the 'unitary' models more tightly knitted to the rational decision model than Allison, despite agreeing on its basic assumptions. Bendor (2003; Bendor & Hammond, 1992), however, considers the link between rules and behaviour to be rather more subtle than Allison suggests, resulting in more complex activity. Bendor also considers that Allison has misunderstood and "inverted" Simon's model, where the use of decision rules may be rather more complex than just "simple, unsophisticated, or predictable" behaviour, especially when activated by a "random shock" (1992: 309). As such behaviour may not be as constrained as Allison reflects. Organisations may be smarter than the individuals comprising them and less sluggish, by specialisation and division of labour (Ibid.: 312). Routine may thus be more positive. Organisations learning from the past may have therefore gained a greater grasp of the rules and operating procedures they employ (Pfeffer, 1981b: 22). Pfeffer agrees with Simon's (1964) consideration that "goals operate as systems of constraints... which decisions must satisfy" (ibid.) and through which they "learn and adapt" through these rules and SOPs (1981: 23). Pfeffer does though consider, as was noted above, that these are less deliberate choices and more evolutionary developments from policy and procedure based on more limited information search and precedent (1981: 23-24). Dahler-Larsen also considers this from a learning perspective exemplifying the opportunity to use evaluation for incremental improvement of specific activity and enlightenment for future action (1998: 33-35). Dahler-Larsen also draws on March and Olsen's (1976) concept of the learning cycle, which is interrupted under complex decision processes. I will return to these reactions more briefly below when discussing decision process models and ambiguity theories, including the garbage can model of choice.

Debates about the incrementalist understanding of decision processes also led Etzioni (1967, 1986, 1989) to offer "mixed scanning" as an alternative model to both rationalist and incrementalist models of decision making. Whilst limiting the detail required in making "fundamental decisions" as outlined by rationalist models, the approach allows for exploration of "longer-run alternatives" underplayed by incremental approaches (Etzioni, 1967: 385). Rather than only focusing like the incremental model upon the "smallest possible units of change", Etzioni claims that mixed scanning, or the humble decision making model, offers an overview of the decision situation that is rather more like "an entire library of encyclopedias under perpetual revision" (Etzioni, 1989). Drawing on the development of the mixed scanning approach (Etzioni, 1986) it might be fair to say that, if organisations evaluate their programmes separately and on their own merits they are acting incrementally. If however they first develop and adopt guidelines that will enable them to assess the quality of their programmes they are using mixed scanning. This search for a "nestling relationship" is how Etzioni draws distinction between fundamental and incremental decisions (Etzioni, 1986: 10). The criticisms offered by Etzioni are

linked to the theoretical developments outlined in section 5.5.3, through the political model.

As was noted above, Hardy combined the rational and bureaucratic models, focused upon substantive rationality, where “decisions reflect attempts to find solutions to maximise effectiveness”, assuming clear goals, information use, selection of optimum outcome and sufficient resources following (1990b: 212). As has been seen, though, Allison (1971) chose to hold these models separate. One might therefore ask if it is correct to split these models. A closer reading of Scott’s analysis suggests the problem might be a misinterpretation of Weber’s definition of bureaucracy, and a failure to recognise his distinction between technical and formal rationality, where the former “emphasizes means-ends rationality” and the latter “the orientation of action to formal rules and laws” (2003: 49)<sup>118</sup>. In rational models there is an emphasis on technical rationality of decisions, whereas bureaucratic models often emphasise formal rationality and matching choices to existing patterns, procedures and rules within organisations. Adler and Borys (1996) also outline a more nuanced view of bureaucracy away from the “pejorative connotations” often attributed it, pointing rather to the existence of both enabling and coercive types of formalization in organizations. The authors recognise the importance of goal congruence to develop an enabling form. Formalization has positive impact when employees are enabled to manage tasks, without merely simplifying routines. This involves building in organisational learning from experience of best practice. This avoids bureaucracy being viewed merely as “rigidity, goal displacement, and authoritarian command and control” (1996: 84). It is also noted that despite recognising his rational focus, interpreters also failed to recognise the importance Weber placed upon context with assumptions regarding open systems, which would influence institutional models (Evan, 1993: 5; W. R. Scott, 2003: 50), which are to be outlined in section 5.5.4.

#### *Decision process models and ambiguity*

Pfeffer notes that “decision process models” suggest greater randomness and less rationality than bureaucratic models, combined with a lack of known goal preference and maximisation (1981b: 25). At the same time, unlike the political models that will be discussed in the next section, behaviour cannot be attributed to power relationships and bargaining amongst different groups within the organisation. Decisions are related to processes rather than preferences. As has already been noted, the rational and bureaucratic models in their purest forms can fail to address the issues of complexity and ambiguity thought to influence decision making. March and Heath note that theorists attempt to treat ambiguity as the exception rather than the rule, hoping to eradicate rather than embrace, seeing it as disorder rather than something ordered and a part of the decision

---

<sup>118</sup> Scott builds on the work of Kalberg (1980) in relation to interpretations of Weber.

process that needs to be controlled (March & Heath, 1994: 192-3). Instead the idea is introduced that organisations can be loosely coupled<sup>119</sup> characterised by “structural looseness” (Lutz, 1982), noted particularly in relation to educational organisations, with their unclear goals, preferences and technologies and fluidity. With regard to the suggestion that organisations either act by logic of consequence or appropriateness, in situations of growing complexity belief is constructed in an organizational setting, which in turn is affected by environmental ambiguity (March et al., 1976). As Scott outlines, the normative pillar is stressed, and social context is important (2001: 67). Concurrently, the ‘complete cycle of choice’ is limited and breakable and complications arise in the process of individual belief and decision making (*Ibid*). March and Olsen identify 4 sources of ambiguity: ill-defined intentions; imperfect understanding of actions and consequences; multiple views of history and variable patterns of organisation. They conclude that they:

“remain in the tradition of viewing organizational participants as problem-solvers and decision-makers...(but) assume that individuals find themselves in a more complex, less stable, and less understood world than that described by standard theories of organizational choice; they are placed in a world over which they often have only modest control. Nevertheless, we assume organizational participants will try to understand what is going on, to activate themselves and their resources in order to solve their problems and move the world in desired directions” (March, Olsen, & Christensen, 1979: 21).

Despite hierarchical control systems and standard operating procedures, organisations still struggle with the problem of “decision coherence” (March & Heath, 1994: 192-3). In such organisations, decentralisation and delegation will often loosen links among the subunits, which can then lead to inconsistency of action over time as well as internally (*Ibid.*: 194) and decision and implementation become loosely coupled (*Ibid.*: 196). Although, as Gamoran and Dreeben point out, two issues are often ignored, firstly that “not every connection... is a loose coupling [and these vary] from one context to another”, and secondly that there is too often a lack of identification of the mechanisms that hold the particular system under investigation together (Gamoran & Dreeben, 1986: 613). As such one must not ignore the controlling nature of regulation and resource constraints as well as custom and situational wisdom. Pfeffer (1981b) included such recognitions within the political model.

Against the backdrop of ambiguity theory and challenges to organisational choice, attention and learning, Cohen, March and Olsen outlined the garbage can model of decision making. This they described to fit the context of “organized anarchies [which] are characterised by problematic preferences, unclear technology and fluid participation” (1972: 1). One of the major propositions of

---

<sup>119</sup> An idea originally perceived by Weick (1976, 1989)

the model is the “partial uncoupling of problems and solutions”, where decision making departs from a purely problem solving activity, and depends rather on a fortuitous alignment of problem, solution and decision maker (1972: 16). The important factor when ascertaining organised anarchy is “intention”, the lack of “conscious choice or planning” (Pfeffer, 1981b: 27). This is also combined against a backdrop of temporal sorting, whereby alternative claims on attention are modified by the time available to make decisions which can bring some kind of order (March & Heath, 1994: 199). The authors argue that the symbolic effects of decision-making can be as important as any decision that appears to be made, signalling values and beliefs in the organisation as well as acting as a socialising process for newer members (Ibid.: 212ff). HEIs are considered to lack cohesion in relation to goals, aims and objectives, and neither do they exhibit hierarchal tightness in structure, leading to more complex decision processes. This was considered particularly evident in the period leading up to the 1980s when HEIs were seen to be administrated but not managed, with a lack of collective definition of goals and performance control (Paradeise et al., 2009). These ideas are further developed within the Institutional perspective, outlined in section 5.5.4.

The model outlined above does, however, raise some difficult issues. Hardy agrees with Musselin’s perception that the Garbage can model relies on a situation of slack resources, which both consider to be fairly removed from the experience of most universities, especially in more recent times (Hardy, 1990a: 400ff). However, it appears hard to argue against the wide disparity of goals within such complex institutions (Patterson, 2001) and the fact that the authors recognised that under times of pressure and clear external demands there would be less flux in patterns of participation and restriction of potential solutions (Cohen et al., 1972). In addition, it was noted that problems identified as important would be dealt with first, finding a context in which they can be decided and attaching solutions to them, relative to organisational constraints like structure and time, and degree of temporal sorting (March & Heath, 1994: 201ff). Questions are further raised as to balancing between focus on the interpretation of decisions and the likelihood of implementation. March recognises, along with Brunsson (1982)<sup>120</sup>, that managing such processes raises a challenge when attempting to align talk and action. The authors suggest embracing ambiguity rather than attempting to eradicate it (March & Heath, 1994). As discussion turns towards the idea of managing organised anarchies, Hardy suggests a move towards, or integration with, political models (1990a: 401). These will be dealt with in the next section.

---

<sup>120</sup> See also (Brunsson, 2002, 2007).

### 5.5.3 Political model (bargaining) – “action as political resultant”

This model is the most complex of those outlined by Allison and appears to resemble participation models of decision-making. Although building on decision making research generally, this framework also explores and accounts for “advantages and disadvantages of member participation” in decision making (Beach & Connolly, 2005: 128). Allison’s model, however, also draws heavily upon his political theory background but at the same time is an attempt to show the role of the individual in organisational decision-making. The model is focused upon viewing “action as political resultant”, where “outcomes are formed, and deformed, by the interaction of competing preferences” via a multiplicity of ‘players’, those “actually engaged in the interaction” (Allison & Zelikow, 1999: 225). In this section then I begin with a brief outline of Allison’s (1969, 1971; & Zelikow, 1999) third model, comparing it to a revision made by Peterson (1976), which is thought to be of particular interest to this study.

Allison’s model accounts for power structures within the organisation. Individuals are more active and strategic, and affected by their epistemic community as much as their organisational background (Allison & Zelikow, 1999). This is interesting in terms of the impact of programme group cohesion, and also whether they are academic members of staff. In less formal, more established organisational decision making processes there is often “deference for seniority” and for “recognised domains of interest or expertise”, and where there are smaller groups the aim is often for consensus. Pfeffer considered that some of these processes reflected control mechanisms within bureaucratic models, the difference being that their use was more divisive, not producing coherent goals (1981b: 28). Framing of the issue is important to its outcome, as is identification of “games and players”, because knowledge of leadership preferences is not sufficient to explain agency (Allison & Zelikow, 1999: 257ff). Explanation of the process at hand is observed to be complete when the pattern of ‘bargaining’ and those responsible for the decision and enacting are acknowledged and their patterns of behaviour and outcomes of their action are identified. The results of bargaining will “seldom perfectly [reflect] the preferences of any group or subunit within the organisation” (Pfeffer, 1981b: 28). Hardy also draws on the political model developed by Baldrige associated with HEIs (Baldrige, 1971), considering it as a “counterpoint” to the idea of consensus in decision making (1990b: 211). Greater focus is placed upon self-interest, bargaining and influence. At the same time it is an open systems model, recognising the influence from the external environment (Dornbusch & Scott, 1975; Pfeffer, 1981b; W. R. Scott, 2003).

When applying this model, Peterson (1976) saw a distinction between pluralist and ideological bargaining, while recognising that these are neither mutually exclusive nor exhaustive. The pluralist model appears to be incremental in

form<sup>121</sup> and primarily concerned with preserving “immediate electoral or organisational interests” (1976: xi) by maximising support, while compromising group demands. The ideological model was motivated by “broader, more diffuse interests... [which are] deeply, ideologically, committed to...” (1976: xii). Where power is shared a decision will often result from a political process, which appears more chaotic and parochially oriented than is evident in the preceding models. This often results in unintended outcomes where power comes through the ability to persuade, but is not random<sup>122</sup> (Allison & Zelikow, 1999: 256). The ideological decision-making model appears to exhibit some characteristics of image theory’s screening process considered the most important part of decision-making (Potter & Beach, 1994). This overlaps with the discussion of bounded rationality in section 5.5.2.

Pfeffer pointed out the weaknesses in the rational and bureaucratic models that disregarded divergent interests and goals within organisations (1981b: 27) and at the same time failing to recognise decision processes that were not concerned with maximising goal attainment (1981b: 29). In addition, if preferences could be ascertained to be consistent, ambiguity models would be seen to have fewer efficacies. Pfeffer goes on to suggest that it is power relationships rather than goals that account for outcomes. Pfeffer and Salancik had noted that organisations, while containing elements of bureaucratic models, often operate as coalitions, as described in the political perspective models of Baldrige and Cyert, Simon and Trow (Pfeffer & Salancik, 1974: 137). However, predicting whether decisions will represent one or the other was considered by Pfeffer and Salancik to be an important area of study. This activity is clearly framed within a resource dependency perspective. Within the power research programme there is a clear division between functionalist, critical and post-modern approaches. Slovic and Lichtenstein (1971) consider that the models here mainly represent developments within functionalist research. Dahler-Larsen warns against considering that interests and preferences can explain everything, and avoiding the question of where they come from, relying too heavily on following the logic of consequentiality (1998: 45). In terms of evaluation, the model might assume that evaluations are always implemented strategically, with a preconceived end in mind or tactically to avoid some imposed action (Dahler-Larsen, 1998: 44). The development of the political model has, however, been criticised as both underdeveloped and overcomplicated (Witt, Andrews, & Kacmar, 2000). Witt et al. define it as a grounded in “phenomena in which organizational members attempt either directly or indirectly to influence other members by means not sanctioned by formal standard operating procedures or informal norms, in an attempt to achieve personal or group objectives” (2000: 342).

---

<sup>121</sup> Extending Lindblom’s muddling through model

<sup>122</sup> This model borrows heavily from the political theory of Neustadt concerning presidential power.



### *Collegiality*

It should once again be reiterated that these models are thought of as ideal types. They are interesting as they have been developed within a similar context to that currently under study. While Allison's models were developed from a macro-political perspective, the propositions from Hardy are based upon research within Higher Education Institutions. These findings can be compared to the writings of Dahler-Larsen (1998, 2001) who has considered the evaluation process itself from organisational-theoretical perspective. These models draw on the same theoretical frameworks. The model of particular interest within Hardy's work is based on an understanding of collegial decision making processes (Hardy, 1990a, 1990b, 1991; Hardy et al., 1983). According to Hardy it is based on the traditional idea of consensus based decision making within an academic community, which is competence related rather than positional (Hardy, 1990b: 210). Hence challenge is made to traditionally understood concepts of power.

While Childers (1981) had found the structurally focused bureaucratic and collegial models to load on similar factors, separate from the process driven political model, Hardy considered the collegial model to be more closely related to the political model. The distinctions between collegial and political processes were thought to relate to actor interest. Collegiality is perceived as more likely, under conditions of commonly accepted ideology, mission (Hardy et al., 1983) or saga (Hardy, 1990b). Hardy notes how Clark's idea of organisational saga (B. R. Clark, 1972), was developed from data which revealed how organisational members were bound to the wider goals of the organisation through their loyalty and commitment. The model "presupposes shared norms, values and premises about organisational purpose, and a commitment to institutional objectives which often seems to revolve around excellence" (Hardy 1990b: 211). The conditions under which this is thought to occur are "small prestigious units or departments with charismatic leaders, or when there is sufficient slack to accommodate disparate goals" (Hardy et al., 1983: 419). Hardy et al. contrast these conditions with those expected to cultivate political decision models, that is, "conflicting goals" and interdependence of interest groups, scarce resources and critical issues to explore (ibid.). The underlying difference between the two models is that in the collegial model participants are guided by common interest and focused upon consensus, whilst in the political model they "seek to serve their self-interest". These are, again, ideal types, and the authors speak of trying to ascertain underlying motivation rather than observing behaviour when attempting to draw distinctions between the two models. At any rate they are expected to be found in combination in organisations. However, according to Hardy a "collegial" approach is not an "inevitable, or even normal, state of affairs in a university... [it] is not automatic – it has to be created" (1991: 137). Recognising this, power should be used to "avoid overt conflict where possible or, if not, to secure agreement and elicit collaboration when differences of opinion do occur" (Ibid: 139). The concept of underlying motivation might be considered to be important when considering this behaviour. With regard to the

collegial model one might also perceive of the ideological conviction to proceed consensually when forming decisions, at least in principle. The latter reflection is important, as Hardy et al., recognise that those promoting collegiality might overstate consensus levels, and those promoting political models might overstate the degree of conflict. It might be better to consider some form of continuum against which to consider decision activity.

#### **5.5.4 An institutional approach - action as ritual decisions with constitutive effects**

There are many different but linked strands of Institutional theory (IT). Peters (2000) identifies at least seven research programmes, including sociological, political, rational choice and historical strands, recognising difficulty in presenting a common core of unifying ideas. A core idea, however, appears to be that structures “do matter”, even though at different levels of complexity, and that these structures persist and regulate members’ behaviour (Peters, 2000: 4-5). Institutional forms are seen to be symbolic, and formal procedure is considered to legitimate the organisation rather than measure performance (Meyer & Rowan, 1977). The institutional model has developed from Weber’s work on bureaucracy, developed in Merton’s work and later by his students, including Selznick (W. R. Scott, 2001: 22). Focus of these studies was upon how organisations could be transformed into institutions over time, but constrained by their environments, with a value-infused and not just mechanical character (ibid.).

Scott notices how the process of institutionalisation has been observed to be a variable, more noticeable in organisations with “more diffuse goals and weak technologies” (2001: 24). The sociological strand of what was considered to have developed into New-institutional theory has therefore promoted research into open, natural systems with their compliance to institutionalised rules rather than the focus of closed natural systems upon autonomous technical efficiency (Meyer & Rowan, 1977; Rowan & Miskel, 1999; W. R. Scott, 2003). The models generally reject perceptions of “autonomous actors... operating with unbounded rationality in order to pursue their self-interests” (Rowan & Miskel, 1999: 359). What is not agreed upon is how to define an institution, how structure will matter and what “factors... shape behavior within institutions”, particularly with regard to whether the sources of preferences are endogenous, as held with normative approaches, or exogenous, as held at the other of the continuum by rational choice institutionalist theorists (ibid.). Rowan and Miskel consider in particular that new-institutional theories investigate how embedded socially-organised environments arise and what affect these have on collective, social action. Institutions may function by formal codes, informal and socialising norms and values, and more cognitive schemata and scripts, depending on the degree of institutionalisation (Rowan & Miskel, 1999: 359-60). These factors are recognised in Scott’s “three pillars” of Institutions: regulative, normative and cultural-cognitive (2001: 51ff). Scott outlines how these “ingredients” are built upon different assumptions regarding compliance,

order, diffusion, logic, indicators and legitimacy built upon varied ideas of social reality and behaviour. The elements of the regulative pillar “include rules, sanctions and surveillance systems”, whereas the normative concerns “values and internalised expectations regarding appropriate ways of behaving”, whilst the cultural-cognitive regards “shared conceptions concerning the nature of reality and means-ends relations” (W. R. Scott, 2006: 886). The latter framework continues to recognise that these “internal interpretive processes [are] shaped by external cultural frameworks” (W. R. Scott, 2001: 57) and the hyphen emphasises “common symbolic systems and shared meanings” (W. R. Scott, 2003: 136). While “all fully fledged institutions are complex composites of these elements”, Scott recognises that the distinctions will represent the “primary source of meaning” within particular organisations and therefore utilising them differently in analysis will lead to different conclusions (W. R. Scott, 2006: 886). Scott’s summary of the pillars is presented in the table below.

**Table 5: Scott's three pillars of Institutions (Scott, 2001: 52)**

	<i>Pillar</i>		
	<i>Regulative</i>	<i>Normative</i>	<i>Cultural-cognitive</i>
Basis of compliance	Expedience	Social obligation	Taken-for-grantedness Shared understanding
Basis of order	Regulative rules	Binding expectations	Constitutive schema
Mechanisms	Coercive	Normative	Mimetic
Logic	Instrumentality	Appropriateness	Orthodoxy
Indicators	Rules Laws Sanctions	Certification Accreditation	Common beliefs Shared logics of action
Basis of legitimacy	Legally sanctioned	Morally governed	Comprehensible Recognizable Culturally supported

### **Decision processes**

Within an institutional focus, questions are raised as to how decision processes unfold, while challenging assumptions that the process is a result of some definable choice. Immergut (1998), commenting on New Institutionalism, sees a central tenet in the recognition of complexity in ascertaining human preference. Research does not attempt to aggregate personal preference, but rather emphasises the rules and regulations within the institutional context that shape decisions where “mechanisms for collective action do not measure the sum of individual preferences. Instead, they allow us to reach decisions, even when there may be no clear-cut consensus” (Immergut, 1998: 138). To Brunsson

“decision” itself is more “institution” than “choice”, recognising the activity of decision making to be rule based and taken for granted, with underlying expectations of intentionality and rationality; extensions of wider belief in individuality and individual behaviour (Brunsson, 2007: 1-2). In addition, organisations are often considered as individuals, expected to take rational decisions leading to greater effectiveness and efficiency (2007: 3-4). The author views decision processes rather as social phenomena, which although tied to choice are distinguishable from it, and often result in different processes than might otherwise have been intended. As a result Brunsson challenges the increasing pressures for greater rationality in decision processes, idealised through increased promulgation of quality assurance focused tools.

Therefore, in considering decision making, understanding the process and level of institutionalisation within an organisation will be important, albeit in this case through the perceptions of individual subunit members. Peters utilises Huntingdon’s criteria to distinguish this (in Peters, 2000: 8)<sup>123</sup>. One of Huntingdon’s criteria surrounds the concept of autonomy, representing the “capacity of institutions to make and implement their own decisions” (ibid). Peters argues that “autonomy” might be part of the manifestation that institutionalisation has “occurred” rather than an indicator of the concept. The author argues that better measures involve understanding the capacity for management and procedural standardisation within the respective institution, or the sense in which the organisation is thought to have become deinstitutionalised (Peters, 2000). The latter process is thought to be more compelling during periods of increased or changing coercive governmental regulation and performance standards, which are often combined with other exogenous factors like funding and resource supply issues (Oliver, 1992). Such processes lead to less coherence throughout the institution (Peters, 2000). It is reiterated that Institutional research is more often focused at the macro level, but there is recognition that theories have relevance at the subunit level (Rowan & Miskel, 1999; W. R. Scott, 2001). Of particular interest here is the sense in which the values in a particular subunit and decision processes that take place are thought to be congruent with those throughout the organisation, whether decision making behaviour is regulated by embedded organisational norms, myths and symbols throughout the organisation and within the wider environment (Rowan & Miskel, 1999).

Pressures from the environment are viewed differently within this understanding. Within this research paradigm came the recognition that organisations might have a technical and an institutional environment (Brunsson, 2002; Meyer & Scott, 1983, 1992). Brunsson summarises these propositions by suggesting that “the technical environment evaluates and supports an organisation in terms of its products and results”, whereas “the institutional environment judges it on its structures, processes and ideologies”

---

<sup>123</sup> Peters does not cite the particular work.

(Brunsson, 2002: 6). Beyond demonstrating results and processes, Brunsson recognises that there is a necessity for organisations to reflect wider societal norms, which have increased in number in recent times. The environment for organisations appears more fragmented and more deterministic, with demands increasingly more heterogeneous. Brunsson suggests that organisations must demonstrate both efficiency and legitimacy. This is movement away from the perception that evaluation should merely be a rational process focused the outcomes of an event compared with the goals originally set. As has already been outlined, research findings have suggested that there is much more that affects the process of evaluating than the quality of the design and thoroughness of implementation before one even begins to try to understand what happens with the results. It is therefore proposed to continue investigation of the interactions and decisions made surrounding design and implementation. In this study this is achieved through in depth reflective interview. In the concluding part of this study I will also argue for increased use of observation and longitudinal methods of research. Such methods will of course require informants to agree to much more intrusion within a subunit of limited membership.

### *Criticism*

Institutional approaches have been criticised in terms of their efficacy due to their reference to deeper structures within organisations as well as their greater focus upon macro-analysis. Another issue concerns the complexity or ease of change in an institution, which relates back to the question of espoused values and “degree of integration” (Peters, 2000: 6). The strands appear to focus on different types of organisation and processes within and between them. Interestingly Peters suggests that Institutional theory itself would benefit from adopting a multiple lens approach from the various strands. These two particular points of criticism are linked.

Firstly a consideration of the level of focus is important, responding to the idea of overemphasis at the macro-level. An issue related to this concept concerns the process of diffusion of ideas across institutions to gain legitimacy. One of the areas within this point relates to the concept of institutional isomorphism, recognising how organisational forms appeared to develop in concert with those considered legitimate within a particular institutional environment (W. R. Scott, 2001: 153). A combination of processes associated with Scott’s pillars is considered to lead to organisations and their subtypes becoming more similar or “structurally isomorphic” as time passes (W. R. Scott, 2003: 164) as they incorporate institutional rules from their particular field (W. R. Scott, 2003: 215), suggests that decision responses would be guided by acceptance of appropriated behaviour within an institutional field. DiMaggio and Powell (1983) outlined an analytic typology of processes encouraging isomorphism: coercive isomorphism, stemming from “formal and informal pressures” from organisations dependent upon and wider societal cultural pressures; mimetic

isomorphism, stemming from the development of “standard responses to uncertainty”, easing the pressures of search; and normative isomorphism, stemming from professionalization and the collective definition of work conditions and practices. Universities are seen to be particular cases of such processes. As was noted in table 5, Scott links these mechanisms to the 3 pillars. At the same time, in response to environmental demands organizations may decouple their operational core from their normative and formal structures and thereby retain some degree of autonomy over processes while being seen to be legitimate, but which leads to greater organisational ambiguity (Meyer & Rowan, 1977; W. R. Scott, 2003).

The degree of determinism to which ideas are thought to be implemented through imitative isomorphic processes has been challenged. One problem, as has been accepted by DiMaggio and Powell, is that institutionalism does not appear to account for deviation and change, “but only increasing conformity and isomorphism” (Jennings & Greenwood, 2003: 198). Strands within the new-institutional paradigm have begun to focus more upon the translation of ideas across boundaries. Although much of this work continues to be focused on macro-processes, this is also significant when considering decision making processes, giving greater attention to the agency of the organisation adopting and adapting ideas from within their environment. De la luz Fernández-Alles and Valle-Cabrera (2006) attempt to respond to criticism and reconcile institutional and organisational theories. New Institutional theory is thought to explain organisational agency as resulting from “both institutional and competitive pressures”, which shows how a combination of these will influence behaviour related to the varying “power relations and legitimacy provided by stakeholders” (de la luz Fernández-Alles & Valle-Cabrera, 2006: 511). Both competitive and institutional pressures must be managed “in order to obtain the social support of stakeholders” (2006: 512). In terms of wider applications of this theory, de la luz Fernández-Alles and Valle-Cabrera in their literature review note how “legitimacy” and “efficiency” are to be linked rather than being polarised, implying that organisations need to focus on obtaining social support necessary for survival, which in turn will lead to “greater access to resources” (2006: 512). The authors argue that some apparent emphases within this field on the “passivity” and “homogeneity” that lead from “mimetic processes” may be overemphasised. There is greater degree of agency than was originally argued. However, as Brunsson recognises, the institutional environment emits volatile norms, making it difficult to combine legitimacy and efficiency (Brunsson, 2002: 7). This is thought to lead to the formation of two organisational structures; a formal, norm adopting structure, suggesting adaptation to changing demands through ritual behaviour and an informal organisation, which is more action based internally, representing the reality of actual behaviour.

Sahlin-Andersson argues further that “imitating” organisations are not “passive adopter[s] of concepts and models defined and spread at the macro-level”, but edit or translate them, to a greater or lesser degree, creating new meaning (1996:

92). She therefore challenges the passive diffusion evident in the earlier institutional theorising. Such an approach focuses more upon organisational identity and the processes of change within the local organisation as well as continuing to investigate the adoption of broadly held ideas. Greater focus was also placed on identity forming within organisational fields, although the latter are not necessarily objective in form (1996: 72). Sahlin-Andersson also notes however that in the search for successful models of organisation, we are confronted by rationalisations of successful practices that become ideas sought out by organisational members or carried by “editors” who circulate them. Røvik (1998, 2002), recognises that certain organisational ideas become widespread and implemented by others as “recipes”, for other reasons than are noted through a more “rationalistic-instrumental approach”, where these recipes are “tools in the hands of rational actors (managers) attempting to design effective and efficient organisations” (Røvik, 2002: 114). From a New Institutional perspective<sup>124</sup> focus is rather placed upon the processes that legitimise, particularly those values that are central to “the modern world such as rationality, efficiency, renewal, development, democracy, individuality, and justice” (2002: 115); the latter drawing heavily on the work of Meyer (1996). Røvik (2007) considers these ideas and “recipes” to spread through decontextualisation, ready to travel across fluid organisational boundaries and reform by contextualisation, whereby organisations drawing them in from the environment adopt and adapt them. This is seen as a “translation theory” perspective. Røvik in essence distinguishes between “virus theory” and “translation theory” (Røvik, 2007: 56ff). The former is focused upon what ideas can do with organisations, considering how resistant or prone certain organisations are to the influence of ideas and the different ways these ideas affect them. The latter, as was seen above, focuses rather on what organisations can do with these pervasive ideas.

#### *Constitutive effects*

An issue arising for this study concerns the concepts of legitimacy and authority, which will impact the response and approach within an organisation to evaluation demands (Dornbusch & Scott, 1975; W. R. Scott, 2006). The concept of legitimacy within Institutional theory differs from the approach, for example, within resource-dependence theories which consider it a resource (W. R. Scott, 2001: 59). Scott notes that within IT, legitimacy will be a “symbolic value” presented and visible to the environment, whatever the pillar that it is most closely tied to. Legitimacy develops from ritual behaviour and shared meaning, but again is interpreted differently by the different pillars, where the cultural-cognitive will be at the “deepest level”, with its focus on “preconscious, taken-for-granted understandings” (W. R. Scott, 2001: 61). Regulative approaches will consider legal requirements and formal demands as regulative rules, whereas

---

<sup>124</sup> As well as being held by symbolists and social constructivists (Røvik, 2002: 114).

“constitutive rules” are associated with the cultural-cognitive approach, rules that are considered to create, influence and have meaning for an event or process within a particular context, reflecting an influence of external meaning, rather than internal purpose; thus “constitutive rules construct the social objects and events to which regulative rules are applied” (W. R. Scott, 2001: 64). These ideas travel “in and around organizations... often highly decoupled from actual organizational practice” (Meyer, 1996: 252). Meyer refers to this softer creation and transmission of ideas in the environment as “Otherhood”, which may denote movements, associational structures, and professions etc., that do not require direct action in order to be of influence. As will be seen below, the idea of constitutive rules further challenges the most basic definitions of evaluation, as well as the decisions that are made internally with regard to design and implementation.

#### *Evaluation from an institutional perspective*

Evaluation and particular derivatives of it like quality assurance have been seen as types of organisational recipes to be adopted for improvement, as a general “management” idea, but also as part of specific reform packages, which are then translated into institutional practice (Dahler-Larsen, 1998, 2006b, 2008). This idea has already been noted in Chapter 3, but will now be related to the institutional perspective. Dahler-Larsen (1998) favours an institutional organisational-theoretical perspective to understanding evaluation in organisations, arguing that it particularly challenges the limitations of currently accepted views of evaluation use<sup>125</sup> (1998: 163ff). Within this model, evaluations are rather understood as “abstract and ambiguous rituals”<sup>126</sup> that drive organisations as much as reporting on their progress (ibid) and have become obligatory (Dahler-Larsen & Krogstrup, 2000: 283). This builds upon the notion of isomorphism, outlined above. However, according to Dahler-Larsen evaluations are considered to construct reality as well as attempting to reconstruct it. Seen from the reflexive-modernity paradigm, Dahler-Larsen notes a changing understanding of evaluation’s purpose, which requires study of a much wider context and more varied interests and values. Such behaviour challenges the possibilities to rationally and resolutely gain clear and unequivocal results of an evaluation. The most one can seem to hope for are recommended reflections that can be seen against these developing values. This requires an evaluator to gain a grasp of how “programme philosophies, organizational formulae and routines are institutionalising, both in relation to the organisation’s own history and the current demands from the environment”<sup>127</sup> (1998: 165). In such a view it is often the formulae that ought to be under investigation, rather than specific programmes. According to Dahler-Larsen, the

---

<sup>125</sup> In the previous chapter these limitations were also linked to a lack of focus on decision processes.

<sup>126</sup> My translation

<sup>127</sup> My translation



whole evaluation process is a choice of values that further constructs and develops reality (1998: 167). This will require greater understanding of the dialogue that exists within loosely coupled systems.

While this approach appears helpful, especially when seeking a new definition of evaluation in relation to institutionalisation, it confirms the necessity of a decision based explanation within an organisational theoretical framework. Dahler-Larsen (1998: 162ff) argues that under reflexive modernization, evaluations can be seen as *arenas* or *identity dramas*, and even if they become mere rituals they can drive and direct future organisational goals as well as reviewing activity. When coupled with developments towards greater accountability, the rituals may become more directive. In responses from informants one should therefore look for descriptions amongst evaluators that suggest not just where their models are drawn from, but also where they perhaps refuse to draw models from. Dahler-Larsen (2005b) recognises that evaluation must be more than a survey, and should be related to values, criteria and standards of some description. This can partly be explained by the lack of competence within the individual organization (2005b: 369). However, as Dahler-Larsen also points out, research from Denmark has shown that evaluation results have little influence over the resources an organization may receive in the future and there is little evidence that such developments alone lead to less funding, closure or general ceasing of activities. On the contrary, poor evaluations can often lead to greater resources being set into 'failing' programmes. This is considered linked to the point that there is a lack of definition of what 'results' are. The nature of NPM should lead to focus on outcomes (2005b: 370-1)<sup>128</sup>. Demands from NPM for increased evaluation are observed to have contributed to a "considerable organizational battle for the design of evaluations, where considerable energy and attention from leadership is bound up in the necessary institutional defensiveness against forms of evaluations, and possible publication of data which can give a negative picture of the organization"<sup>129</sup> (Dahler-Larsen, 2005b: 372).

Evaluation appears to have become embedded as an organisational routine, especially when seen in relation to current "recipes" focused on organisational effectiveness (Dahler-Larsen, 1998: 25). Dahler-Larsen agrees with Sanders (2002, in Dahler-Larsen, 2006) that evaluation is drawn to the forefront of organisational thought and action, is mainstreamed and part of the "organizational structures, cultures and processes which regulate how organizations function" (2006a: 146). There is an increasingly greater absorption and integration of evaluation into routines and procedures. This, the author claims, leads to an "on-going [collaborative] process of reflection and...

---

<sup>128</sup> The debate over how results and the effect of programmes can be ascertained is not taken up here, but rather acknowledged as a factor influencing attitudes to evaluation. What is clear from the research literature is a debate over values.

<sup>129</sup> My translation

learning”. That being the case, the author also recognises that the absorption of such function has often developed alongside the demands raised by the greater introduction of NPM, affecting the public policy arena. This is due to the belief that evaluation becomes part of an organisational “relevance structure”<sup>130</sup>, which may subtly transform evaluation especially when organisations evaluate one another and becomes standardised both in terms of chosen criteria and methodology (Dahler-Larsen, 2006a: 147). Such behaviour does not, however, always fit the activity under study (Abma & Noordegraaf, 2003 in Ibid.).

However, Dahler-Larsen and Krogstrup also consider that looking at institutionalisation is not unproblematic. They outline how it will require some recognition from investigation that a “standard” has been accepted and adopted as the “correct” way to organise; unavoidable and developing via routine. Evaluation, however, has many forms and phenomena; from top -down to bottom up; summative-formative etc.; the process is not linear and complete and it can be practised differently even within an organisation, with different demands and will certainly vary across the different levels of analysis and organisational fields (2000: 285 -6). Additionally, the authors claim there are phenomena that appear similar to evaluation and can be confused for it, like Quality Assurance. One might also question in what way evaluation is synonymous or not with quality assurance in this context. At the same time it is to be remembered that there is not, however, sufficient enough data collection to consider this question across the HEIs in focus in this study. Furthermore, as the authors conclude, evaluation practice is not a “New Institutional dream situation”; it appears difficult to trace processes and approaches within and across organisations (2000: 286). What does, however, appear evident is that under NPM there have been increased moves towards institutionalisation of evaluation<sup>131</sup>. However, evaluation contexts are complex and it is difficult to ascertain the spread of standards, across fields and different levels of analysis, while isolating the processes involved (2000: 286). One of the authors’ conclusions is that although eventual effects of an evaluation on a programme will be difficult to discover for “users, controlling authorities, and the general public”, there is an understanding that the action of doing an evaluation is itself considered “appropriate”, a view apparently held widely across society in an age of reflexive modernity (2000: 287). As a result, evaluative activity is thought to produce and reproduce more and greater focus on evaluation. The underlying difference now is that an atmosphere of “doubt and criticism” overtakes the

---

<sup>130</sup> Relevance structures are a concept developed by Berger and Kellner (1982), which Dahler-Larsen describes as “a socially defined orientation which guides the selection and structuring of knowledge appropriate for particular purposes” (2006a: 141). Such an approach is counterbalanced by Popper’s reflection that knowledge is actively constructed, offering a choice between existing and newly defined structures.

<sup>131</sup> Interestingly Hellstern (1986) noted that this did not appear to lead to increased utilisation.

process, while “conventional norms” are replaced by “organisational decisions and environmental pressure” (ibid.)

Although an increased focus on evaluation across the organisation has been generally seen as a positive development, Dahler-Larsen (2006a) also raises questions to the effect of mainstreaming external standards on an organisation’s competence to evaluate activity itself. As Allison (Allison & Zelikow, 1999) noted, routines and regulations are built upon specific organizational values which in turn can limit decision making with regard to future organisational behaviour. Dahler-Larsen also reflects over the interaction of organisations with regard to evaluation. It appears that when interacting with or assessing one another, organisations focus more on structures, procedures and control systems than on more direct outcome measures (Power, 1997 in Dahler-Larsen, 2006a). At the same time issues are raised throughout the organisation as to how to proceed with the process at hand. Understanding the interpretation of these processes at the micro-level requires consideration of another linked perspective, that of the sensemaking paradigm. This will be dealt with in due course, but attention is first turned to the concept of constitutive effects viewed in relation to evaluation.

#### *Constitutive effects of evaluation*

Dahler-Larsen declares that evaluation should be seen in the light of cultural and institutional developments in society as a whole rather than just as the result of public policy developments (2006b: 11). In this way he sees it as a change of mind-set, where “an evaluation doesn’t just describe reality but to a great degree has an effect on it as well. Evaluations constitute something” (2006b: 12). Drawing on the work of Røvik outlined above, Dahler-Larsen considers that “the social and political significance of evaluation and performance data increases due to “the ideology of New Public Management”, noting that performance measurement is considered to contribute to “political decision making and resource allocation” and in so doing it will “enhance the rationality and accountability of each institution” (2007: 17). Dahler-Larsen notes that such processes appear to have led to the idea that evaluation should be adopted in a more encompassing manner as an on-going event throughout the whole organisation. It appears to become not just a political demand but is also passed on as a recipe for success; in this case that evaluation should be mainstreamed, all-encompassing and fully integrated; an institutionalising factor. While increased focus upon evaluation in organisations and greater participation in the process have, as has already been outlined, been goals across the evaluation field, there is a more discomfort with the closer linkage to performance management and rational decision making.

Despite the fact that there is greater standardisation across evaluation systems the particular consequences can be very diverse (Dahler-Larsen & Krogstrup, 2000: 297). The authors note that simple standards can detract from the complex

contexts under which human processing activities take place. They agree with the conceptualisation of institutional theory based on cultural models, recognising that evaluators will act as interpreters and new ideas might be institutionalised on that basis (ibid.). They also find evidence in concert with these newer approaches that there might not be such a marked division between “symbol and substance” rather than a more deterministic spread of systems and ideals that achieve symbolic status and can lead to decoupling within organisations. Standardisation can take place in different forms and at different levels of fields and organisations, which can affect “local” use, more incrementally than determinately. Although this doesn't explain everything, it does challenge “functional, rational and technical” mechanisms of explanation - opening for questions of values and identity to offer explanatory power (2000: 298).

Constitutive effects concerning how “evaluation as an institutionalised phenomenon co-constructs the social reality surrounding the evaluation”<sup>132</sup> are observable in “three aspects of social reality”- content / material, timing, social /identity (2000: 295). The first aspect, “material”, deals with how evaluation can frame interpretations, orientations and actions; where the system becomes the mental frame that work is considered through. With the demands placed by evaluation, across, through and within, it can add a new slant on the meaning of work already undertaken and can be retrospective and prospective; experiences are interpreted retrospectively, and prospectively one can anticipate what one is going to be evaluated on and influence work undertaken (ibid.). This idea is considered especially salient and challenging in this study. One might, for example, investigate whether groups consider how they will be judged in relation to the impact of their programmes and adjust their evaluations accordingly. The second aspect refers to “timing”, raising questions concerning at what point an evaluation will take place, the authors considering that different timing will produce different pictures (Dahler-Larsen & Krogstrup, 2000: 296). The authors also consider institutional timing be important, in relation to when other things take place in the institution. Data will be different according to the point at which it is collected (ibid.). When evaluation becomes institutionalised within a pattern, like QA, performance indicators etc., the pattern itself can have constitutive effects, and groups must adjust in order that the right effects will be visible at the right time. Important questions raised by this aspect include who controls the timing and how it is controlled. The third aspect concerns social relations and identity, asking questions about who has the right to be heard and how they are defined; i.e. customer, user, consumer, as well as how they are selected and who they should represent (Dahler-Larsen & Krogstrup, 2000: 296). Roles can be redefined by such processes; especially if the task of evaluation reduces some sense of autonomy of the professional or challenges their standpoint (Ibid.: 297). As a result professions might try to define their

---

<sup>132</sup> My translation from Danish

own standards. This raises further interesting questions for this study, with regard to a tension for decision makers.

Building upon the former points, Dahler-Larsen reflects how the public arena appears to have taken on a new rule giving role, where it reports, interprets, edits and presents data, becoming constitutive in the sense that many more effects than a simple presentation of data will become evident. In the author's examples "measurement of quality may lead to everything else but better quality", elsewhere described as "performance paradox"<sup>133</sup> (Dahler-Larsen, 2007: 19). This might also include the "reactions to evaluation". But these developments as observed from this perspective are not without further problems. As was seen in Chapter 3, Dahler-Larsen considers there to be a mismatch between NPM terminology, favouring "transparency, visibility, documentation and measurement", and the experience that evaluation processes are "long, complex and non-linear" (2007: 18). This was initially noted to relate especially to the impact of performance indicators, but applies more widely to evaluation. A question is raised as to whether the aforementioned goals of NPM are achievable or whether there is another impact of such behaviour.

This brings us back to the concept of use. Although this study does not follow the use of evaluation data, Dahler-Larsen outlines the centrality of use that shapes the evaluation definition and thus could be said to influence the process in its entirety, with all the subsequent complexities with defining such a loose term, and the importance of considering the way it influences evaluators (2007: 20-23). So here questions might be raised with subunit members as to their perception of how data are used, what feedback they receive and tentatively to explore the impact of this on their decision making. As was outlined in section 3.6, 'use' is an idea that has been uppermost in the mind of evaluation theorists. An additional question might concern how quality assurance systems and the action of self-evaluation affects this perception or is affected by it.

According to Dahler-Larsen (1998, 2007) a constitutive influence, that I will return to under the data section of the study, appears to be that QA becomes the "formal" but not necessarily "active" image of evaluation, that begins to impact the framework over time. Academics might appear to interpret programme feedback, outcomes and results based on their own professional value systems, while underplaying the importance of the focus of QA systems. At the same time the demands for reporting outlined within the organisation begin to direct their focus towards alternative ways of evaluating, causing them to adopt processes and approach issues in ways that appear contradictory to the value structures. These factors are combined with their perception of the way information required is subsequently processed, as well as the type of decision arena that underpins the process of designing an evaluation. Understanding how

---

<sup>133</sup> Van Thiel and Leeuw, (2002).

these processes develop requires further analysis relevant at the micro level. This is dealt with in the next section.

#### **5.5.4.1 Sensemaking – links to Institutional theory?**

Another approach to decision processes in organisations that is important to consider is the sensemaking perspective. Sensemaking in organisations primarily concerns understanding “[how] does something come to be an event for organizational members”, followed closely by interest in what an event might mean to those participating and how they respond by “bringing meaning into existence” (Weick, Sutcliffe, & Obstfeld, 2005: 410). Weick considers that as an interpretive perspective the sensemaking paradigm is not fully represented by any theory of organizations, despite recognising the “nature of organization” (1995: 69). By organising, this perspective considers how order is brought recurrently through agency, constituting rules and meaning, where organisation emerges from sensemaking (1995: 82). Weick (2001) considers sensemaking to be different from decision making, applying Daft and Macintosh’s (1981) view of decision making considering it to be about “strategic rationality”, where the aim is to “remove ignorance” by finding “clear answers” to “clear questions” (2001: 107-8). Sensemaking is focused on “contextual rationality”, “where vague questions, muddy answers, and negotiated agreements... attempt to reduce confusion” (2001: 108). Sensemaking is therefore considered to be the framework that enables decisions to be made (Weick, 2001: 460). Therefore less focus is placed on the decision as an event, rather as a sequential process moving from chaos to some kind of order, via notions of noticing, bracketing, labelling, retrospective reflection over current and prior events, culminating in agency based on presumption – occurring interactively across a social setting (Weick et al., 2005: 410-13). The idea of order does not however suggest “getting it right” or finding the truth, reiterating the recursive notion underlying the concept, thus moving away from the rational perspectives on decision behaviour and accepting the notion of plausibility rather than accuracy (2005: 415). Members will attempt to influence others’ sensemaking towards a particular meaning through a process of “sensegiving” (2005: 416). This is the attempt to influence others’ “sensemaking and meaning construction... towards a preferred redefinition of organizational reality” (Gioia & Chittipeddi, 1991: 442). Following up these reflections, sensegiving has been found to be “triggered”, at both leader and stakeholder levels, by “the perception or anticipation of a gap in organisational sensemaking” (Maitlis & Lawrence, 2007: 57). At stakeholder levels, members feel a sense of *bounded responsibility* to act upon issues thought important for the organisation and where leaders are thought to lack competence. The authors found sensegiving to be “enabled” by the presence of *discursive ability*, where actors “construct and articulate persuasive accounts” and by *process facilitators*; the “routines, practices and structures” that allow such behaviour (ibid). Collective sensemaking is therefore something more than “shared values”, highlighting the constructive process of enactment, whereby individuals adjust to their environment by acting upon it (Weick, 2003: 185). Weick also recognises that “shared” is an

ambiguous concept which can imply “domination or codetermination” (Weick, 1995: 136). This refers back to aspects of collegiality, recognising that investigation is required beyond the level of agreement, to the understanding and framing of decision alternatives. However, Weick considers that the incidental nature of decision making, and the collective process of enactment links sensemaking and the process of organising, noting that “people are in a complex *reciprocal* relationship with *their environments*” (Weick, 2003: 186). As such, enactment should be seen as a change mechanism (Jennings & Greenwood, 2003).

Despite not being represented in any particular theory of organisations, Weick sees links to Scott’s concept of organisation, outlined in the section above, where sensemaking would be most prevalent in organisations thought to be more loosely coupled, open systems (1995: 69 - 70). Weick et al. (2005) recognise that sensemaking perspectives have rarely been combined with Institutional theory, despite both reflecting on organisations as open, natural systems. This is mainly due to the focus upon different levels of analysis. Scott considers sensemaking research to be at the socio-psychological level, while Institutional theory has been more focused at the ecological level (W. R. Scott, 2003: 122). However, opportunity for combination exists, for example, studies applying the former should lead to the provision of “micromechanisms” that reveal “cognitive structures associated with mimetic processes, agency [etc]” and could link micro levels of analysis to macro levels normally focused upon within studies applying the latter (Weick et al., 2005: 417). Weick considers there to be no ontological difference between these levels (Weick, 2003: 190). However, Weick et al. in turn recognise that sensemaking perspectives on organisations can benefit from the understanding of wider “social and cultural context” as noted by Scott to be explored by Institutional theory (1995: 151, in *Ibid.*). In addition, in sensemaking perspectives agency may both be “over exaggerated” and in relation to institutional influence and enactment might be over-individualised (Weick, 2003; Weick et al., 2005). At the same time, micro studies focused more upon the cognitive measures of the degree of institutionalisation can hinder over-exaggeration of the “sanctioning capacity of the external environment”, avoiding confusion with resource dependence perspectives (Zucker, 1991: 104). Zucker also recognises that micro studies can help distinguish variation to external demands, revealing differentiation of responses rather than isomorphism (1991: 105). It would seem that the perspectives are complimentary rather than commensurate. Despite these issues, the sensemaking perspective and issues of enactment offer an interesting framework for understanding evaluation and the decision processes related to it. These are briefly discussed in the next section.

#### *Evaluation as sensemaking and assisted sensemaking*

Evaluation has been viewed as a sensemaking activity from the “simple and somewhat naïve argument” that it is the very purpose of it; to identify and assess

processes and assign value to them (Van der Meer, 2007: 169). However, despite such simplicity at the outset Van der Meer recognises the complexity that develops as different actors “attach meaning to the evaluation process and its outcomes” (ibid.). Van der Meer outlines how the “interplay” of these processes determines “the shape, outcome and the impact of evaluations” (2007: 170). The former, initiation stage, is of most interest in this study, despite restating the belief that the processes are perceived to be interlinked and recursive. During initiation substantive, methodological and organisational choices are made by both commissioners and evaluators, in what will form the evaluative “script”. Van der Meer considers that such choices are based on the existing meaning frames and practice patterns, or “repertoires”, of the different actors involved where the sensemaking interactions shape the evaluation. The intermingling of repertoires becomes more complex as the process develops and more actors become involved. Van der Meer recognises that these “third actors” challenge the principal-agent perception of evaluation decision processes that has prevailed, raising a question of who really initiates an evaluation and at what juncture in the course of action (2007: 172 -3). Understanding actors’ repertoires is therefore thought to be important here.

Dahler-Larsen (2005a, 2007) also considers evaluation processes from a sensemaking perspective. Following Mark et al (2000), he adopts the view, that evaluation is a form of “assisted sensemaking”, implying that the nature of the purpose of an evaluation can change the approach of the evaluators as they may reconsider the content and purpose of their programmes within the light of the indicators set (Dahler-Larsen, 2007: 25). Viewing evaluation as assisted sensemaking affords the possibility to consider how the process takes place, particularly the decisions made (Mark et al., 2000). The authors’ concept of sensemaking is drawn from the work of Weick, (1995) yet departing from his attention on organisational management and focusing more upon representational and valuative natural sensemaking (Mark et al., 2000). The authors consider that in Weick’s focus, aside from the everyday types of sensemaking that people engage in to make sense of the world around them, consideration is placed upon the process within organisations that seek to illuminate and overcome bias and improve judgement and decision making. Mark et al recognise that humans attempt to represent the world around them and make value-based judgments about quality (2000: 6). It is to these processes that evaluators attempt to offer models for explanation, but the authors see their approach as combining the two rather than focusing on one or the other. However, they recognise that there is always an underlying purpose to any evaluation, which in their case is the idea of social betterment. The different models of evaluation (as noted in Chapter 4) have influenced decisions about design (2000: 11), which will include understanding the purpose of them, and lead to different inquiry modes (2000: 12).

Evaluation therefore is considered to assist interested stakeholders in their making sense of programme design and implementation, by undertaking



focused, systematic inquiry. Mark et al therefore set out to offer a “sensible” model of evaluation that will enable “evaluators and others think through the most important decisions that must be made in planning an evaluation” (2000: viii - ix). This approach seeks to disentangle the various purposes of evaluation. Mark et al consider that such an approach offers a better insight into how evaluations will be used than judging the effectiveness of an evaluation upon its use (2000: 22). Their model though is based on a principle of utilising evaluation for the purposes of social betterment. This principle leads to a broader definition of evaluation, when compared with the often noted alternative of “evaluation as the determinant of merit and worth” (2000: 3). Mark et al see the latter as only one of the “legitimate” purposes of evaluation, the others being program and organisational improvement, oversight and compliance and knowledge development. According to the choices over which purpose is in focus will differ depending upon context, but should be implemented to further social betterment, or as they refer to it, “betterment-driven evaluation” (2000: 12). However, their model, appearing normative, pays only a little attention to other “motives” that might drive an evaluation, including requirement by legislation or an external mandator or even those “less than pure”, even though they maintain that focus on their model can overcome “political” tactics (2000: 50). But it is for these reasons that this study is considered important, understanding how evaluators respond in such situations and develop their model will hopefully enlighten the decision processes and enable groups to make sense of their role. In that regard it would seem to offer further assistance to the aims of Mark et al and enlighten another part of the process.

## **5.6 Combining the models**

One of the issues that arise when considering the different decision process models is the extent to which they are distinct from one another or how they may in some way interact or can be combined. Allison outlines how the models may “complement one another”, where the rational model outlines the “broader context, larger[...] patterns and shared images”, the organisational behaviour model reflects the “organizational routines that produce the information, options, and action, whereas the political model “focuses in greater detail on the individuals who constitute [the organisation] and the politics and procedures by which their competing perceptions and preferences are combined” (1999: 392). The models may also compete with one another, revealing contrary conclusions (1999: 394) although the level of complexity of Allison’s case study at macro government level does make it harder to catalogue the attitudes, preferences and opinions of the processes under investigation. The addition of an institutional model will only add to the complexity. Allison also recognises that not all choices or activities are observable, as well as accounting for the misinterpretations of actors involved in these processes about the behaviour of others. Invariably when attempting to explain the behaviour of an aggregate actor, no single understanding of what happened is considered likely to become evident.

Allison and Zelikow recognised that while their models could offer different interpretations of the same event, they also produce “different explanations of quite different occurrences” (1999: 387). They recognised that one’s paradigm of inquiry magnifies a set of factors over another when considering approaches to be divergent. The authors recognised two imperatives, first clarifying the “explanandum”, that which is to be explained, and then beginning to describe the phenomenon rather than pre-categorizing under a particular model. This might be seen at various levels of abstraction (1999: 388), although the authors also go as far as suggesting that causal relationships can be ascertained that account for the “difference between what actually happened, on the one hand, and some specified or assumed alternative states of the world, or the other” (1999: 388). This recognition of different logics observed from different perspectives is mirrored in the work of James Thompson (1967), who noted especially how the open systems and closed systems perspectives compliment one another, rather than offer alternative explanations.

Scott (2003) also considers how perspectives can be combined. While Scott deals mainly with organisational models, rather than decision process models per se, there is a great deal of overlap and influence between the models as they are presented here. A multi-paradigmatic approach is increasingly suggested, due to greater complexity and recognition of difficulties with generalisation and commensurability of organisational theories (W. R. Scott & Davis, 2007: 370). Reflecting on the use of multiple models, Scott (2003) outlines the contributions of Etzioni’s “Structuralist” model, Lawrence and Lorsch’s “Contingency” model and Thompson’s “Levels” model. Etzioni combined rational and natural systems perspectives highlighting the presence and importance of power and conflict and challenge within the formal and informal parts of an organisation (W. R. Scott, 2003: 103-4). In attempting integration, Lawrence and Lorsch further considered the variation on the level of formalisation and relationship to the environment, where rational and natural perspectives account for different subsets of organisation forms rather than differences within an organisation as Etzioni envisaged, and open systems was the framework for these subsets (W. R. Scott, 2003: 104-5). Thompson considered the perspectives to have equal efficacy for a particular organisation, but applied to different levels<sup>134</sup> within that organisation dependent on their openness to the environment and susceptibility to uncertainty (W. R. Scott, 2003: 105-6).

Thompson’s approach, including the author’s perceptions of its grounding, is of particular interest to this study. According to Thompson (2003: 4-5) the drive for greater organisational efficiency led to organisational commentaries like Taylor’s (1911) scientific management, Gulick and Urwick’s (1947)

---

<sup>134</sup> The technical level attempts to be as rational as possible, the managerial level focuses on the natural and the institutional level must be open to the environment (Scott, 2003: 106).

administrative management, and Weber's (1947) bureaucracy, which in turn led to greater acceptance for the rational model, which assumed a closed system and attempted to demonstrate clearly defined goals and efficient structures. Thompson declares the results of this model, the focus of which is on planning and control, to be a functional organisation where all action and allocation is appropriate (2003: 6). Thompson recognises this is challenged by open system strategists. But Thompson also challenged the natural system theorists, whom although recognised uncertainty in variables not under complete control, also took a functional view of the "interdependence of organization and environment" (2003: 7). Thompson's greater issue, however, appeared to be that the fields to all intents and purposes rejected one another, when despite adopting different logics, focused on different phenomena ignored by the other. Thompson's work builds upon the developments of the "Simon – March – Cyert stream of study" that recognised "the organization as a problem-facing and problem-solving phenomenon" (2003: 9). This view accepts the premise of bounded rationality, outlined earlier, developing a process of satisficing in organisational decision-making. However, Thompson also considered that over-focusing on uncertainty can lead to a rejection of earlier theories that will weaken analysis of organisational activity. Thompson therefore suggested the combined model, based on his propositions that closed and open system theories are complimentary. He suggested that his model conceived "complex organizations as open systems, hence indeterminate and faced with uncertainty, but at the same time as subject to criteria of rationality and hence needing determinateness and certainty" (Thompson, 2003: 10).

Scott, outlining a "layered" model, considers that the perspectives contrast, but do not disprove one another, and are "applicable to differing levels of analysis" (W. R. Scott, 2003: 121). The layered model addresses whether the focus of the organisation is means or values oriented, "self-sufficient" or "context-dependent" and at what level it functions. Scott and Davies (2003) also note how Open systems theory has led to the substitution of structure with process, with greater emphasis on organising rather than organisation. The authors note that one result of the adoption of such approaches has been the "gradual breakdown of the public / private-profit / non-profit distinctions" as organisations are viewed as "boundaryless" (2003: 388). Scott's reflections are helpful for this study, particularly if the data show a departure from the way that theory has been applied earlier. It will therefore be vital to account for the differences in organisational type, level of observation and analysis in relation to other works cited.

Adopting a multi-paradigmatic approach does not however mean a lack of integration. Scott and Davis refer to Lewis and Grimes' (1999) overview of such approaches. "Parallel studies" are a branch of multi-paradigmatic research where data is collected and analysed to "cultivate varied representations of a complex phenomenon" (1999: 675). Similar to Allison's (1971) studies, parallel studies "preserve theoretical conflicts by depicting organizational voices, images, and

interests magnified by opposing lenses”, although the authors recognise that subsequent authors have sought to ground their work “in more contrasting assumptions” (ibid.). In the next chapter I will consider how this study in some way attempts to emulate this approach. Scott and Davis, while recognising the complexity and often conflict between paradigms and theories believe there is more to be gained from integrating rather than suppressing the diversity between them (2003: 374). Huber and McDaniel also considered that within increasingly more complex environments, that are “hostile, complex, and turbulent” (1986: 572), earlier paradigms of organisational design did not fully embrace how decision making was changing from an emphasis on positional power to technical power, requiring more distributive forms of communication. The authors agreed with Herbert Simon that decision processes could be distinguished from production processes and had become central to organisational behaviour (1973, in 1986: 575). This however, is a more normative view, and contrasts as the authors affirm with the decision-making theorists focused more upon the actual processes taking place and how these affect outcomes, where the latter deals with “emergent” processes compared to the former’s focus on “intended” processes. Despite seeing these differences, Huber and McDaniel identified the importance of distinguishing “decision units” within the organisational structure, that is, those who make decisions on behalf of the organisation. Thus, they highlight the importance of investigating decision processes that within and across levels in an organisation, even though they are more focused on normative design rather than processes that unfold.

While many authors have begun to call for greater integration of models representing organisation and decision theory, there have been criticisms of such approaches. Pfeffer, for example, commenting on Allison’s studies disagrees to some extent that all the models can be applied to the “same situation”, considering that they will make different predictions and “an analyst must decide where to place his bets” (Pfeffer, 1981b: 29). Pfeffer recognises that multiple dimensions of each model must be applied in order to begin to understand which one best identifies the organisation and under study, and preferably within a comparative frame of reference (1981b: 30). In addition, he recognises that any process of trying to analyse which model fits a specific organisation is complex, and the analyst is affected and influenced by “accepted paradigms” (ibid.). Scott and Davis agree with the latter reflection, recognising that theorists “bring their disciplinary habits with them to the objects of their study” as well as ideas which they have often been socialised in through professional schools (2007: 370-1).

On reflection, Hardy (1990a) recognises three generations of research into academic decision making processes. In the 1970s 4 major models were recognised as descriptors of university governance: bureaucratic, collegial, political, and garbage can (1990a: 401). Building on this research, focus during the 1980s was upon mixed models, which in turn led to a new research programme in the 1990s, which focused upon cultural aspects of universities

(1990a: 393). In summarising this research Hardy outlines 5 overlapping models, adding a rational perspective, and recognising the necessity for investigation of underlying intentions as well as the complexity of the overlap. An example of this complexity is her suggestion that collegiality and politics are in fact “two sides of the same coin”. The problem has been that models of collegiality have focused upon the decision making structures but not the processes, examining mostly the degree of decentralisation, whereas collegiality as process should rely more upon understanding underlying attitudes and behaviour (1990a: 397).

Weaknesses in each individual model will also be apparent. Commenting on the organisational model in particular, Rosati (1981) noted that processes vary greatly according to context, structure and participant style and the attention and involvement of the top leader<sup>135</sup> and future monitoring of the implementation process. While this view into one of the models may lack the analytical power of the combination of alternates, and also will need adjustment when applied to lower level decisions, it does draw us back to the importance of power and position in the process. The focus in this research is, though, more on process than position. Hardy (1990) notes that structural investigations asking questions like who is involved in decision process have had a preeminent focus in organisational decision research. She goes on to recognise that questions of process further examine how decisions are perceived to take place and what the motivation behind them is. But assessing such processes is not easy. Such cautionary remarks and criticisms will be further considered in the methodology section outlined in Chapter 6, for example a response to Langley (1999) concerning how well the alternate templates strategy allows for a combination of the models that are said to overlap.

### **5.6.1 The basis of decisions and decision processes**

While it is considered important to outline the models of decision making, focus is also required upon how decisions are made. The different perspectives have been shown to present competing, but not always exclusive, frameworks for decision making processes taking place in organisations. As Scott reiterates, “bureaucratic-administrative” models reflect the rational system perspective, and “coalitional-bargaining” models reflect natural and open systems and these perspectives consist of frameworks for goal setting against which decisions are taken (W. R. Scott, 2003: 303). In these cases, as was seen earlier in this Chapter, decisions are based on purposive criteria, although the decision process varies by choice alternatives. As has also been seen, more complex models of organisation and decision making, present, as Scott puts it, different “classes of decision situations within...organisations” where no preference orderings are clear (2003: 304). These have been outlined within political and institutional

---

<sup>135</sup> In Rosati’s cases the President of the USA.

models. In outlining these models of decisions, questions were further raised concerning power and control (Rowe, 1989; Thompson & Tuden, 1987).

Thompson and Tuden hypothesised that both consensus and choice are required for effective implementation of organisational decisions, noting, for example, that if choice is taken before consensus is reached it will still need to be reached later (1987: 211)<sup>136</sup>. The authors present a typology of types and constraints, which builds on the proposition that “decision issues always involve two major dimensions: ...beliefs about cause / effect relations and... preferences regarding possible outcomes” (Thompson, 1967: 134). The authors recognise that both dimensions, their basic variables, are present even if not “consciously considered”. Thompson’s thesis is that organisations strive towards rationality, despite being both natural and open systems at the same time, in the drive for effectiveness and efficiency (W. R. Scott, 2003: 105 - 6). Where there is lack of certainty<sup>137</sup> about cause and effects and / or preferences then decision-making necessarily departs from a rational perspective (Rowe, 1989). Each of the different “decision issues” will require a different strategy (Thompson, 2003: 134). According to Scott the decisions relate to various organisational models, computation equating to bureaucracy, compromise to legislature, judgment to collegiums and inspiration to charismatic leadership (2003: 304). Thompson and Tuden’s (1987: 198) framework for decision strategies is outlined in the figure below.

		Preferences regarding possible outcomes	
		Certainty	Uncertainty
Beliefs about cause-effect relations	Certain	1. Computation / Calculation	2. Compromise
	Uncertain	3. Judgement	4. Inspiration

**Figure 8: Thompson and Tuden's framework of decision strategies**

Decision issue 1 requires certainty of both means and ends, creating a computational strategy, which Thompson considers essentially dichotomous,

<sup>136</sup> This work is a reprint of their original article published in 1959.

<sup>137</sup> Scott distinguishes between “agreement” and “disagreement” (2003: 304)

dependent on the volume of data (Thompson, 2003: 134). Thompson recognises, however, that there are often greater constraints on decision makers with regard to both sets of variables. Within the same column, decision issue type 3 recognises the uncertainty of means even when goals are considered to be clear (ibid: 135). Thompson uses a relevant example for this study, recognising that an educational programme hinges on many factors, like participant motivation, which will in turn affect outcomes even if these are ascertained and agreed upon. In these situations a judgemental strategy will be adopted, building upon the expertise of professionals and supported by “extensive discussion” (Scott, 2003: 304).

Within the next column focus is placed upon uncertainty concerning ends or outcomes. Decision issue 2 recognises an outcome uncertainty (Thompson, 2003), or as Scott interprets it disagreement over the preferable alternative (Scott, 2003: 304). This uncertainty will often arise when “dynamic human objects” are involved and will influence outcomes (Thompson, 2003: 137), in the case of this study, where evaluations will involve the input of programme participants who may have differing desires from the programme compared to providers. In such cases negotiation and bargaining will help determine decision, or if the disagreement is extreme enough then some type of representative body will be established, and procedural agreement will often be the basis of decision (Scott, 2003: 304-5). The final category is that of “inspiration”, so named because it is required in situations where there is neither agreement about means nor ends, if “any decision is forthcoming” (Thompson, 2003: 135). Scott, drawing on Thompson and Tuden, recognises that within these “crisis situations” where there is little agreement and a great deal of uncertainty, charismatic leaders will often arise, which will lead over time to a form of routine-like decision making (2003: 305). But Scott notes, as was also outlined in Section 5.5.2, that the various findings of March, Olsen and Cohen support the view that these situations are not “crises”, but rather the conditions of organised anarchy, especially characteristic of educational organisations (Cohen et al., 1972; W. R. Scott, 2003: 305). As was outlined earlier, in such “garbage can” decision situations of high uncertainty, solutions can be attached, or dumped on, varied solutions. Universities are considered as classic examples in this work, likewise in Thompson’s judgmental category. Interestingly, these different authors appear to focus upon different part of a continuum. In Thompson’s category there is a greater sense of unanimity combined with professional expertise, even though this can be challenged by the task environment and the resulting dependency of the organisation (2003: 139). For example, complexity and uncertainty are increased by greater heterogeneity within the environment, which increases the complexity within the organisation as groups must comply with multiple and often conflicting demands, and despite localisation of this interdependence require a degree of central coordination to get things done (Ibid: 140). On the other hand, Cohen et al (1972) outline examples of where leadership often appears more by default. There would seem to be a difference not only based on certainty of means and ends, which of

course can be challenged by the environment, but also on the ability to withstand demands. In this study focus is particularly on the perceived relative strength of the subunit to “interpret” such demands when making decisions about evaluations.

When comparing these propositions, it is important to point out that Thompson originally applied his propositions at the macro level. Pfeffer, following up this issue, proposes that it is “possible that either or both forms of agreement may define an organization, but that for a variety of reasons, consensus about cause-effect relations may be easier to attain and hence constitute a more fundamental property of organizations” (1981a: 13). While the author recognises that organisations might exhibit lack of consensus about goals, at the subunit level there will often be different ideologies, and possibly a greater degree of agreement. As Pfeffer goes on to say, “...it is important to recognize that organizations have subunits which may have their own ideologies, shared meanings and subparadigms”, where the internal boundaries are thought defined by “communication”, “extent of control” and the subunit characteristics by “commonalities in paradigm” and the degree of “shared definitions of the situation” (1981a: 13). This proposition is thought to be relevant to this study in which participants within different organisational subunits, consider the purpose and design of their evaluations, also within their wider organisational settings. While the wider organisation might search for consensus over these issues, the subunits are therefore thought to have their own preferences. Weick and McDaniel (1989) adapted Thompson and Tuden’s taxonomy which they felt to be limited, considering that sensemaking processes precede the decision strategies, determining “the extent of agreement on preferences and cause-effect relations” (Weick, 1995: 112). Weick considers this to be part of the underlying ideologies that consist of cause-effect beliefs, outcome preferences and “expectations of appropriate behaviors” (1995: 111). The complexity of the issue in these particular cases is that focus in the organisation is placed on the wider question of quality assurance of programmes within the general organisational structure, as well as particular focus on the impact of the academic area within question, that is the development of educational leadership. As has been pointed out this is further complicated by the fact that places within these programmes might be commissioned by an external mandator, or that the programmes might be arranged externally. Thompson and Tuden’s typology of decision strategies has been applied within the field of evaluation studies. Hellstern (1986) applied a similar model but decisions were based on knowledge needs, whilst Hansen (2005a) adapted it “freely” to apply to choice of evaluation models rather than looking at the decision process within evaluation per se.



## **5.7 Decision makers**

Within the frameworks for decision making questions have been raised concerning those who are involved, in this case concerning evaluation design and implementation. In this study the focus is on the decision making processes within HEI subunits, and understanding how decisions are made and what the content is with regard to the evaluation of their programmes, which are, of course, amongst other things dependent on the degree of coupling and relationship to wider groups. Drawing together the research outlined from the field of Higher Education as well as wider decision research, focus is placed on the academic professional group and the view of the members within it concerning evaluation. In Chapter 2 it has already been presented how these views are generally based on theoretically held perspectives of organisational effectiveness (W. R. Scott, 2003). Dahler-Larsen (1998: 146ff) commenting on Scott's (1977) research into organisational effectiveness agrees that varying definitions as such are normative, and therefore different interests place weight on different criteria. Dahler-Larsen extends this reflection to propose that any declaration that an evaluation is built on "valid" criteria is necessarily a "cultural definition". Understanding how such ideas are thought to take shape within the group is thought to be important. A presentation of these groups considered to be helpful is that of the "occupational community".

### **5.7.1 Occupational communities- link to professional judgement**

The concept of occupational community was applied by Dahler-Larsen to evaluation settings (1998: 141) and is drawn from the work of Van Maanen and Barley (1984). Van Maanen and Barley recognised groups who, within a phenomenological cultural perspective, are recognised as an occupational community when seeing themselves as engaged in similar work, identity and fellowship based on a set of shared norms, values and interpretations. Such an understanding is affected by individual identity and also influences the wider organisation developing identity. Cox (2005) notes that the idea of occupational communities may often be considered similar to the concept of "community of practice" developed by Lave and Wenger (1991). Within a concept of occupational communities, however, focus is rather on the "power of common work situations and structures – as opposed to directly joint practices – to create commonality... and immediate mutual understanding... and underpin social networks", which can be problematic for research in communities of practice (Cox, 2005: 530). These issues were additionally noted by Gronn (2003: 30-31) who recognised firstly the difficulty of defining "fluid" communities of practice as well as the accounting for likely problems of conflict when considering allegiances of members, particularly with regard to rival groups.

Dahler-Larsen (1998: 146) notes that "occupational communities" are observed to react negatively to the demand to evaluate but this reaction is tempered by, amongst other factors, their position in the wider organisation, their perceived relative degree of autonomy / heteronomy and degree of acceptance of the

evaluation criteria to be operationalised. These reactions to evaluations, which according to Dahler-Larsen, often come after results have been published, are thought based upon a perception of mismatch with norms, values and standards, and are typically more belligerent to more concrete criteria that appear not to reflect that which they consider intrinsically special with their programme, as well as that which is based on human contribution. Such reactions will call more particularly for an internally designed and led evaluation focused on processes rather than structures and outcomes<sup>138</sup>, which reflects the fact that an occupational community does not always know why its members act as the sum of their actions will also draw more widely on more ‘immeasurable’ societal norms, values and demands. Thus Dahler-Larsen constructs an image within his institutional perspective on evaluation, of greater demands from mandators for accountability over implementation of programme goals in terms of results, which draws a response consisting of ideological self-defence of the occupational community but which is complicated by the normative problem of attempting to assess the accomplishments in such a short space of time after delivery (1998: 149). The author therefore constructs a useful typology to inform how the occupational community’s reaction to evaluation is based upon its perception of relative autonomy/heteronomy and how well evaluation criteria match their own cultural understanding. This typology outlined in the table below is thought useful to illuminate the question of domains in which choices about evaluation are made. If occupational communities can be identified, they may be considered to reflect the characteristics presented in the table. Although based within an institutional perspective, this typology is also considered to be useful to illuminate the other models.

---

<sup>138</sup> Dahler-Larsen draws again here on the work of Scott (1977).

**Table 6: Occupational community reactions to evaluation demands (after Dahler-Larsen, 1998)**

	<b>Evaluation criteria are in agreement with the occupational community's own criteria for quality work</b>	<b>Evaluation criteria are not in agreement with the occupational community's own criteria for quality work</b>
<b>Occupational community considers itself to be relatively autonomous</b>	Occasion used to promote own interests, achieve favourable viewpoint and use evaluation as a lever to gain resources, partners etc. to achieve positive results. This is especially seen when poorer results than expected are attributed to other influencing factors than the underlying programme theory or members of the occupational community.	Use professional role to undermine the methodological logic of an evaluation, as they disagree with the criteria employed. Will decouple the evaluation activity from own work. Promoting one's own programme theory will also help prevent future negative situations. This can bring negative reactions if it is obvious that an organisation has 'rigged' the results.
<b>Occupational community considers itself to be relatively heteronomous</b>	A weaker academic / professional group 'lives up to' evaluation criteria matching the organisation's and group's own values <sup>139</sup> . This is especially the case when applying for extra resources, tied to certain criteria.	The academic / professional group is forced to accept criteria they do not agree with but do not have the autonomy to ignore <sup>140</sup> . They can therefore either live with the problem of discrepancy, which will lead to decreased influence over future processes, while experiencing a sense of 'role distance' or attempt to revise their own institutionalised criteria, which may focus merely on quantitatively measurable outcomes or data from an individual level, often limiting individual engagement.

<sup>139</sup> According to Dahler-Larsen, while this approach sounds 'attractive' in terms of logical consistence, it makes learning processes and searching for new goals difficult within the evaluation process.

<sup>140</sup> According to Dahler-Larsen, while this approach sounds 'attractive' from a leadership perspective, it can lead to "cynicism, irony, decreased engagement, and self-supporting patterns of grumpiness and shifting of responsibility" (2001: 92 my translation). Such behaviour can split the occupational community under evaluation (1998: 154).

Challenges associated with this model could relate to the degree of internal / external agreement over criteria. What if some in the organisation agree with the premises while others do not? This will necessarily affect, as we see above, the overall relationship to criteria, but it is unclear what kind of variation there will be. This is therefore likely affected by organisational position and role. For example how do those in internal higher positions with more control over the decision influence the process in relation to those implementing the evaluation? In addition, investigation should hopefully consider how the context affects evaluation design e.g. between different educational frameworks and their different demands for evaluation of programme input. As will be seen in the data chapters, I have tried to approach the issue of response to demands within the interviews with different providers. It is of course a sensitive issue and few strong conclusions are drawn, but ideas drawn together from the different groups spoken to.

The table below adapts the decision models of Dahler-Larsen (incorporating Thompson and Tuden) and Allison, while adding perceived type of evaluation utilisation. This offers a framework for understanding evaluation decision processes and practices within the subunits under study.

**Table 7: Occupational community reactions to evaluation demands related to responses<sup>141</sup>**

Perceived degree of academic/professional autonomy	Occupational Community's degree of internal agreement with evaluation criteria	
	High	Low
High	Occasion used to promote own interests, achieve favourable viewpoint and use evaluation as a lever to gain resources, partners etc. to achieve positive results. This is especially seen when poorer results than expected are attributed to other influencing factors than the underlying programme theory or members of the occupational community.	Use professional role to undermine the methodological logic of an evaluation, as they disagree with the criteria employed. Will decouple the evaluation activity from own work. Promoting one's own programme theory will also help prevent future negative situations. This can bring negative reactions if it is obvious that an organisation has 'rigged' the results.
	<i>Adoption (Bargaining/legitimizing)</i>	<i>Decoupling (Legitimizing)</i>
Low	A weaker academic / professional group 'lives up to' evaluation criteria matching the organisation's and group's own values <sup>142</sup> . This is especially the case when applying for extra resources, tied to certain criteria.	The academic / professional group is forced to accept criteria they do not agree with but do not have the autonomy to ignore <sup>143</sup> . They can therefore either live with the problem of discrepancy, which will lead to decreased influence over future processes, while experiencing a sense of 'role distance' or attempt to revise their own institutionalised criteria, which may focus merely on quantitatively measurable outcomes or data from an individual level, often limiting individual engagement.
	<i>Translation (Symbolic)</i>	<i>Colonising (Constitutive)</i>

<sup>141</sup> From Dahler-Larsen (1998: 149ff; 2001) (and drawing on Allison and Peterson)

<sup>142</sup> See footnote 139

<sup>143</sup> See footnote 140.

## 5.8 Discussion

In this chapter I have outlined briefly models of decision-making in organisations, with an emphasis on naturalistic approaches that attempt to explain how decisions are actually made. The intention is to use these models as analytical framework when investigating decision making in relation to programme evaluation. It is acknowledged that areas will be illuminated differently under different models of decision behaviour and policy formation; hence a multi lens framework for analysis is considered necessary.

### 5.8.1 Proposed implications of the models

It has often been assumed that improving the quality of evaluation procedures or increasing participation will produce better findings and enable greater utilization. However, this study suggests that the underlying ideological positions of evaluators will help offer significantly greater explanatory power for why programmes are evaluated as they are. In the previous sections, therefore, I drew on theories of programme evaluation, evaluation utilization and naturalistic decision-making<sup>144</sup> to offer a nuanced approach. As a result this study notes the importance of challenging “implicit conceptual models” (Allison & Zelikow, 1999). This will involve supplementing commonly used rational models with other “frames of reference”, exploring complexity and decision-making procedures from different perceptions. As has been noted, research connecting evaluation and decision-making has been very limited (Holton III & Naquin, 2005) with little distinction between the public and private sector, despite recognition of fundamental difference between the two (e.g. Lægreid et al.2004). This study recognises the complexity of the evaluation context, therefore, alternative models of decision behaviour are offered in order to illuminate the process, relying on participant reflection to develop a more robust understanding (March & Heath, 1994: 18).

I have outlined Alison’s (1969, 1971; Allison & Zelikow, 1999) three conceptual decision making models: “rational actor, organizational behavior, and Government politics”, noting that these were further adapted by Peterson (1976), who divided the final model into “ideological bargaining, and pluralist bargaining”. The models are developed from observations of macro level Governmental behaviour, but are recognised to be applicable at other levels including “local governments; nongovernmental organizations...schools, [and] universities” (Allison & Zelikow, 1999: 7)<sup>145</sup>. This approach was also recognised as useful by Scott (2003) and Pfeffer (1981b) and influenced research on decision processes within Higher Education (Hardy, 1990a, 1990b, 1991; Hardy et al., 1983). In addition, an institutional approach was adopted which has built particularly upon the work of Dahler-Larsen (Dahler-Larsen,

---

<sup>144</sup> To reiterate, naturalistic models focus upon how decisions are made in reality.

<sup>145</sup> The models have also framed to some extent the work of Lægreid *et al.* (2004).

1998, 2001, 2006b). It is thought that these models will function as alternate templates (Langley, 1999), within the strategy to be outlined in the next chapter.

### **5.8.2 Initial implications of the rational actor model for this study**

Under this analytical model, explanation of the observed process at hand can be said to be complete when the evaluation performed by an organisation is considered to be a reasonable response given the specified objectives of the programme (Allison & Zelikow, 1999: 5). Rational decision-making might, for example, build upon the attempt to discover a model capable of isolating various variables that could provide evidence of the programme's impact. Such a course of action is unlikely, but would be reliant on clarity of underlying rationale, aims and objectives of the programme as well as understanding of the causal connections and measurements likely to provide evidence of impact. Within this model, it might be expected that the purpose in evaluating might be to collect information that would subsequently be utilized conceptually and instrumentally. Dahler-Larsen (1998: 121) also applies this model, considering the corresponding organisational model to be one of a system, "loyally applying plans decided after mapping goals and weighing up alternatives"<sup>146</sup>.

### **5.8.3 Initial implications of the organizational behaviour model for this study**

This model focuses on behaviour that is characterised in terms of *outputs* of standard patterns of behaviour, or rules, rather than deliberate *choice*. We can question, therefore, how much evaluation takes place as the result of standard procedures or because a ready alternative more or less matches the goals of the activity. In the context of this study organisations are often smaller but commonly demand a broad range of tasks of their workers, of which evaluation is only one. This model also suggests such groups to be loosely coupled, tied to previous solutions and routines, and divided across task in such a way that creates difficulty for top leaders to exert a total form of control over activity, creating more a 'bottom-up' type of response. Although, as Allison and Zelikow reflect, leaders still have the potential to "substantially disturb...specific behavior" (Ibid.). In contrast to model one, it is assumed that the search for the optimum form of evaluation would be ignored; regardless of the intensity of demands both internally and externally. Programme providers and internal evaluators in particular are thus equally unlikely to develop optimum devices or calculate the possibility for doing so when demanded to provide evidence that their programme works. Following this rationale (Allison & Zelikow, 1999: 6), explanation of the evaluation process at hand might be said to be complete when the organizational decision routine and implementation are acknowledged and patterns of behaviour and action outcomes are identified. Within this model, it might be expected that the purpose in evaluating might be to collect information that would be utilized instrumentally and conceptually. Dahler-Larsen also

---

<sup>146</sup> My translation.

applies this model, albeit under the idea of the learning organisation, considering the corresponding organisational model to be one of a “knowledge based system, correcting itself through feedback following action”<sup>147</sup> recognising the limitations of decision making and operating under conditions of bounded rationality (1998: 121). Dahler-Larsen also considers that use in this perspective will be focused on “enlightenment” (1998: 162).

#### **5.8.4 Initial implications of the political model (bargaining) model for this study**

One would also assume that under this model discovery of the optimum form of evaluation would be ignored. But within this model the individual ideology and underlying values of those involved in decisions about the evaluation process are important, even if there is an observed division on pluralist and ideological grounds. Pluralist forms might be exemplified in evaluation activity aimed at securing programme survival, against for example a demand from mandators to provide evidence of impact against goals. Providers might struggle to develop useful information to support the livelihood of their programmes, resorting to a pluralist bargaining for what kind of information could be made available. Resulting evaluations might provide some information of programme impact, but would likely do little more than give the perception that a programme was functioning as it was intended. Scott recognises that within political frameworks there will be divergent thought and interest conflict but that resolution is expected through processes (2003: 355). Within this model, it might be expected that the purpose in evaluating might be to collect information that would be utilized symbolically and possibly to legitimate certain positions. Dahler-Larsen (1998: 121) again considers the corresponding organisational model to be one of a system, but one where “different groupings fight over resources based on interests and power bases”<sup>148</sup> challenging concepts of consensus and the order of the more rational, closed system models.

Where ideological bargaining takes place one might expect to discover a programme provider at odds with their mandator or at least reliant upon developing a model in a collegial, bottom up fashion, where normative underlying values compete and require information that will support a general standpoint. The ‘belief’ that the impact of school leadership training and development upon pupil outcomes can be measured is one such controversy, as is the debate over the overall aims of the programmes, whether they should be generic, skills based or theory focused conceptual development. Programme providers may seek to maintain their standpoint of ‘evaluating’ that which they believe can be ‘measured’, defending their stance to mandators. Such behaviour would most likely lead to the collection of information that would be utilized to legitimise programme survival. It may be harder to find evidence for the

---

<sup>147</sup> My translation.

<sup>148</sup> My translation.



bargaining model in this study, but interview respondents will be invited to reflect upon tensions when designing the evaluations for their programmes. In this case the perception of programme leaders and chief administrators will be illuminating. In this model much is reliant upon participant understanding, where it is recognised that each knows only a small part of the story and memories fade (Allison & Zelikow, 1999: 312). Within this model, it might be expected that the purpose in evaluating might be to collect information that would be utilized to legitimate programmes.

### **5.8.5 Initial implications of the institutional model for this study**

The Institutional model recognises the open system, interdependence and influence of the environment on organisational decision making. Evaluation is normally seen as symbolic or ritual event, as organisations respond to external demands and frameworks to implement policy. While normally a macro perspective, often considering the behaviour of organisations within a field, sensemaking perspectives have allowed a micro-view of organisational behaviour, where decisions are influenced by external demands and members attempt to make sense of these demands in their design and implementation, enacting a plausible response. There is a difficulty in transferring external policy to the “inner life” of organisations, which is especially notable within the HE sector (Westerheijden, 2007).

Dahler-Larsen (1998: 121) divides this model into two sections, the former focused on loosely coupled systems and the latter formed through shifts in organisational identity. In the former, the corresponding organisational model is considered to be a “loosely coupled system of values, knowledge forms, methods, organisational recipes and routines, where imitation and the taken as given provides legitimacy, each having its own logic”<sup>149</sup>. Dahler-Larsen agrees that this logic challenges the logic of consequentiality. In these models Dahler-Larsen expects use to be ritual, taking the form of institutional revision, developing through constitutive effects that can construct reality. In the second model, concepts of identity development mean that the organisation “unfolds whilst interacting actor locally translate in a reflexive fashion in loosely coupled institutional elements with reference to creation and maintenance of identity” (1998: 121)<sup>150</sup>. This model is focused in local translation rather than diffusion leading to isomorphism. Use in such approaches will contribute to this development of identity.

Ideas of anarchy and loose coupling do not, as was seen earlier, equate with chaos. Interestingly Scott recognises that within such frameworks organisations have been considered as “anarchies” and loosely coupled, which can mean that internal subunits can dichotomously be highly ordered and autonomous in

---

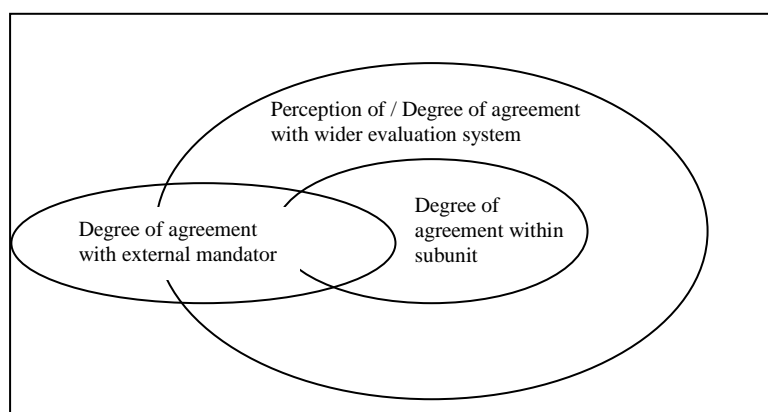
<sup>149</sup> My translation.

<sup>150</sup> My translation.

relation to the wider organisation under investigation (2003: 355). When the overall political goals for a programme are unclear, there will necessarily be a local interpretation and adjustment in order that a concrete implementation can take place. This is thought to result in a “gap” between the original political intentions and goals and the criteria for evaluation that are subsequently applied, a factor more clearly presented in institutional theories than in accounts of rational or learning organisations (Dahler-Larsen, 1998: 87). This idea is drawn from Scott’s (1977) recognition that goals and evaluation criteria are often different (in Dahler-Larsen, 1998: 86). From an institutional perspective we should expect some degree of difference between the environmental political demands (either internal or external) and the criteria set up for an evaluation. What remains unanswered though is the degree of dissonance in the intervening process and what the organisational causes for this are. Dahler-Larsen helpfully recognises that concretisation of criteria is necessary when evaluating in a way that is unnecessary when merely setting out overall goals and values. The ability to know what the successful product will look like may therefore be decided post event rather than pre event. In many instances this may explain why programmes adopt an air of “action research”, which will allow a fluidity and retrospectively non-linear rationality to speak for the activity that has been engaged in.

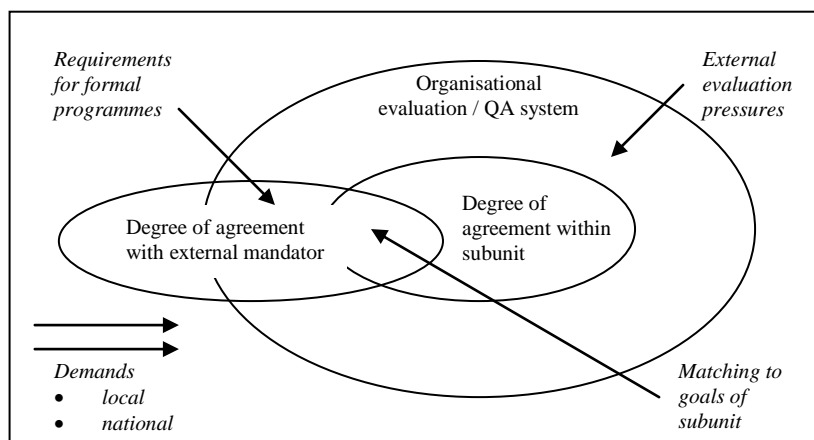
### 5.9 Investigating decisions about evaluation

In this current study when investigating the perceptions of decision behaviour with regard to evaluation and assessment within the subunits, it is considered necessary to explore how members consider the goals of evaluation at different levels. It will therefore be important to discuss the perception of degree of agreement within subunit, the degree of perceived agreement with wider organisational demands and the degree of perceived agreement with direct external demands. These intertwined relationships are presented in a model of goal agreement in figure 9 below:



**Figure 9: Evaluative goal agreement across organisational levels**

When applying these categories to the current context, the subunit under investigation can be understood to be the programme unit offering varying forms of postgraduate programme in school leadership development within the wider Higher Education Institution. The wider evaluation system will generally refer and apply to quality assurance systems that have been set up across the institution, but may also include other assessment strategies within specific institutes and/or faculties. It is recognised that these will be influenced by external pressures for evaluation at different levels. Influence from the subject field will also be considered important. As was seen in both chapters 2 and 3 one area of the field has tried to come to grips with is that of evaluation of output, more recently distinguished as impact. Despite striving towards greater understanding of cause and effect in this area (e.g. Guskey, 2000; Leithwood & Levin, 2005) there does not appear yet a recognition that this has been ascertained. Many groups would challenge the validity of such search. It will be interesting to observe whether providers perceive a different reality that guides their action. Do they believe that cause and effect of their programmes is certain, and if not what guides their actions? As many programmes are offered to external mandators investigation will also be required as to how their goals for the programme and subsequent requirements for evaluation will influence the subunit and to what degree they are in agreement. Discussion should focus on how these processes develop. The interaction of these relationships is outlined in the figure below.



**Figure 10: Evaluative goal agreement within context**

For the purpose of this study, the categories of Stufflebeam and colleagues, outlined in section 5.1 have been adapted to take into account the application of broader decision models and more recent research into the design and implementation of evaluations outlined in this chapter. The new categories form the basis of research questions and areas for investigation with the various

subunit members invited to take part in this study. The operationalisation and methodology associated with this process are outlined in the next chapter.

**Table 8: Reapplication of Stufflebeam et al.'s categories of evaluation problems**

<b>Stufflebeam's category of problems</b>	<b>Focus</b>	<b>Category in this study</b>	<b>Focus</b>
Evaluation definition	How does the underlying perception of definition of evaluation define the model chosen to be implemented?	Definition	How do subunit members understand the concept of evaluation and how does this influence the process?
Decision making	How can knowledge about the decision making process improve evaluations and their utilisation?	Decision making	What responses are there to the demand and pressures for evaluation and what can application of mixed models of decision making tell us about evaluation processes?
Values and Criteria	What are the criteria by which evaluation data will be interpreted, and whose values weigh heaviest?	Demands <sup>151</sup>	What demands are placed upon the subunit and how are they interpreted?
Administrative levels	What is the point of focus and level of analysis of an evaluation?	Decision makers	Who is involved in the decision making concerning evaluations?
The research model	How is evaluation methodology different to research methodology?	Designs	What is the purpose of the design and what degree of agreement is there about models?

<sup>151</sup> Due to being focused upon the perceptions of programme providers rather than explaining the wider values of all stakeholders.

In summary, the different conceptual models build on a “cluster of assumptions and categories” that influence analysis (Allison & Zelikow, 1999: 379). The models are not thought of as the main form of explanation of the decision making process at hand, but rather an analytical framework for understanding the participants’ view of their organisational decision-making process. The models are seen as complimentary rather than mutually exclusive, that can offer competing conclusions as a result of their assumptions and propositions. Allison and Zelikow recognise that while Model 1 paints the broader picture of the decision that is made searching for an understanding of optimal choice and is a “powerful, first approximation”, Model 2 focuses on the organisational routines that produce the information, options and action. Model 3 is more detailed in understanding the individuals within the decision framework and how perceptions and preferences are combined (1999: 392ff). Together they should enable broader analysis of the evaluation process. Allison, however, opens for the possibility that alongside providing different answers, the models probably ask different questions. Recognizing the latter point would appear to partly allay concerns of whether it is possible to accept duplicate approaches concurrently (Pfeffer, 1981b). Model 4 recognises the complexity and ambiguity of organising; particularly the influence of the environment and the nature of how these demands might collectively be appraised within organisations that are governed by formal and informal pressures. The intention is to use these models to better understand the design processes that underlie the formation and implementation of evaluations. I now turn to the methodology for this study. The data collected related to demands will be dealt with in chapter 7, while that concerning definitions and designs will be considered in chapter 8. Chapter 9 will include the data collected related to decision makers and decision making. The next chapter deals with methodology for the study.

## **6. Knowledge claims and methodology**

### **6.1 Purpose of the study and strategy of research**

The background for this study is the desire to explore the decision processes surrounding the evaluation of postgraduate programmes for school leadership development offered by HEIs. From an initial study of programme frameworks and content, interest was placed on how programmes were evaluated. As was recognised in Chapter 2, pressures and demands on providers to supply evidence of programme impact have been increasing; in England there was noted to be a particular pressure to relate impact upon school outcomes. A question was therefore raised concerning evaluation models that were used to assess postgraduate programmes. Following from this Chapter 3 dealt with the concept of evaluation, reviewing theory, development of evaluation designs and models and different traditions. Additionally, the review further outlined the connection between these developments and wider public policy approaches and related initiatives. In Chapter 4 attention was focused upon the development of evaluation structures within higher education institutions and how these related to external demands for quality assurance and internal pressures for professional improvement. Reflection within regard to the developments in these three chapters led to focus on the decision processes that took place within the subunits concerning models of evaluation to be implemented. Chapter 5 therefore outlined different theories and models of decision processes within organisations.

The purpose of this chapter is to present the methodological framework of the study. It includes a discussion of the underlying ontological and epistemological assumptions that frame the choices taken for methods used and form of data analysis. Firstly I review the purpose of the study, relating it to the review of literature outlined in the preceding chapters. This is followed by the discussion with regard to a critical realist approach to research which addresses the strategy of inquiry chosen for data collection related to naturalistic decision making processes. The next section outlines the qualitative methods chosen, followed by a discussion of the analytical approach. The analytical approach of template analysis (TA) is chosen which is linked to the use of a priori theoretically linked frameworks, or templates, as the basis for empirical data collection. These templates are developed from themes in the literature review and develop through exploratory interviews and documentary review. The theoretical models outlined in chapters 3 to 5 are combined within an alternate templates strategy. Discussion also takes place with regard to how TA is pertinent to a critical realist approach. Focus is also placed upon discussion of assuring quality of the overall study. In outlining these various parts, the overall structure of the study is presented.

The study therefore attempts to explore how decisions are perceived to be made concerning evaluation models to be designed and implemented within

programme groups offering postgraduate studies in school leadership; namely whom and what influences these decisions and how groups attend to different demands placed upon them. A qualitative interview study was chosen to investigate subunit perceptions of internal decision making. Four subunits were chosen, two in Norway and two in England. In the ensuing sections I will outline the reasoning behind the choices made for this study. First I will deal with the ontological and epistemological basis informing my approach.

## **6.2 A pragmatic approach to critical realism**

The purpose of this study is to interpret the attitudes of members to the decision-making processes in their organisation when enacting evaluations. The literature review combined with analysis of theoretical positioning, suggests that such an approach departs to some extent with the main body of evaluation research. An interpretive approach is required, focused on attempting to understand the meaning and context in which events take place (Maxwell, 1996). This approach has often been related to pragmatic research where focus is placed upon gathering qualitative data for analysis and inference (Teddlie and Tashakkori, 2003). Such an approach is derived from the research problem at hand, underlined by the fact that a pragmatic approach is considered to look epistemologically at the destination of an idea rather than its origin (Maxcy, 2003: 75). The research problem thus becomes of primary importance (Tashakkori & Teddlie, 1998). This raises ontological questions over the nature of reality and its construction, while warning of an end to the ‘philosophy of formalism’ (Maxcy, 2003; Teddlie & Tashakkori, 2003).

A pragmatic approach recognises that objectivity of truth and the ability to grasp, explain or interpret it must vary with type of study. As this study seeks to gain a more detailed and richer grasp of the concepts under investigation at first glance a pragmatic based qualitative paradigm of inquiry would seem wise to adopt (Creswell, 2003; Teddlie & Tashakkori, 2003). This would suggest acceptance of a more open ontological view, built on the belief that methods must match purpose. In such cases it is important for the researcher to reflect over presuppositions of the external world, rather than accepting some kind “ontological asymmetry” (Baert, 2005: 152). Following on from a Deweyan idea, such views are espoused by those rejecting an instrumentalist view of science. Baert exemplifies Giddens and Bhaskar holding to these approaches that “implicitly assume that an ontologically grounded social theory provides the necessary conceptual apparatus to make the portrayal of the social possible” (Baert, 2005: 151). Although Baert’s view offers useful criticism of approaches to philosophy of science, his loosely based neo-pragmatic view can appear overly normative at times. Teddlie and Tashakkori offer a seemingly more ‘pragmatic’ approach while recognising the importance of the debate that Baert raises. As such, a ‘conversational’ approach can help the researcher reflect on their objectives and the best way of achieving them, whereby methodological questions do not ultimately become merely “reduced to matters of ontology”

(*Ibid.*: 154). This “methodological pluralism” can however lead to difficulties in application and analysis (Ackroyd, 2004: 137).

The pragmatic approach bears some resemblance to critical realism, where existence of an objective reality is accepted, but the two fields differ to some extent over the possibility to explicate truth. According to Smith (2007) one of the difficulties with pragmatism is the proposition that ontology and epistemology can be separated from choice of methods and strategy. Part of this problem is “epistemic relativism”, whereby judgements about “aptness” of method are related to the current vogue rather than use of philosophical criteria (2007: 5-6). A critical realist approach, however, adopts an ontological view of objective existence while also recognising the “transitive nature of knowledge” (D. Scott, 2007: 14).

### ***Critical realist approach***

Critical realism (CR) is considered to be an ontologically focused meta-theory (Hesketh & Fleetwood, 2006: 658). Initially the CR field in the USA, exemplified best by the work of Drake et al.(1921), sought to draw on the reactions against realism and pragmatism (Rennie, 2009). Drake and others sought to find balance between objective and subjective views of knowledge; between dualism and monism (Drake, 1921). It was later that focus shifted from epistemological reflection to ontology (Rennie, 2009). Modern application of CR is drawn mainly from the early work of Bhaskar<sup>152</sup> (1978, 1979, 1984, 1986) in addition to later commentary by Archer (2000), where CR continues to challenge the main stream of methodological approaches (Sayer, 2004). Burgoyne describes CR as possibly a “synthesis to the thesis and antithesis” of positivism and constructionism<sup>153</sup>, agreeing with Drake (1921) that it is a third way (2008: 65). The author argues that rather than building on an “epistemological assertion” like positivistic and constructivist based positions, CR begins with an ontological proposition of general event regularity – a sense of stability in the world, which may not always occur varying by situation and context (*ibid.*). This assumes a difference between the natural and social worlds, where recognition of social construction in the latter does not exclude the sense of underlying order and reality, which may be poorly or little understood (Fairclough, 2005: 922). Rather than the *closed* stability of positivism or the perceived absence of any stability beyond the constructed meaning given, CR proposes “an open system with emergent properties”, where the world is not considered “mechanically predictable” but the observable is a “manifestation of the real”, that which is possible (Burgoyne, 2008: 65). Therefore CR research,

---

<sup>152</sup> Bhaskar refers rather to the concepts of scientific and transcendental realism, calling positivism an “illusion” (Bhaskar, 1986).

<sup>153</sup> The author intertwines terms like construction, constructive and constructivism, recognising the origins of terminological difference between fields such as psychology and sociology (Burgoyne, 2008).



according to Burgoyne, focuses upon “understanding what *stimuli*... have triggered what *processes* and how these are affected by the *context* leading to what *outcomes*” (ibid). Fairclough notes that critical realism, where ontology becomes more clearly distinguished from epistemology, thus highlights the contingent nature of agency and its contextual interaction, differentiating the ‘real’ (structures), ‘actual’ (events and processes) and ‘empirical’ (where the real and actual are experienced and acted upon by social actors). (Fairclough, 2005: 922). The social world is transformational in nature, “agents draw upon social structures (etc.) and, in doing so, reproduce and transform these same structures” (Hesketh & Fleetwood, 2006: 658). This ‘stratified’ and ‘transformational’ approach to ontology thus attempts to “avoid the ‘epistemic fallacy’ of confusing the nature of reality with our knowledge of reality” (Fairclough, 2005: 922). Agents “recreate, reproduce and /or transform” pre-existing structures rather than creating them (Ackroyd & Fleetwood, 2000: 14). A point of importance, is then, that neither of the “polarities” regarding subject and object is “privileged”, as in other approaches, but rather are recognised as internally related (Fleetwood, 2005: 216).

#### ***Application within this study***

Drawing on the work of Ackroyd (2004: 156ff), implications that follow from building on the CR perspective for this study include the importance of understanding and applying theory, association of it with data findings which may be independent of it, where the research process is seen as interpretive and creative, accepting and embracing iteration, whilst recognising that data are purposely constructed. Ackroyd maintains that all research is to some extent theoretical and guided by “prior conceptualisation” making theory indispensable (2004: 156). Kuzel agrees that there is always some prior understanding or theory, “no investigator is a blank slate” (1999: 35). Theory does not, however, “determine what is seen and taken to be significant”, and observation challenges preconception where recognition of context and structures is important (Ackroyd, 2004: 156). The research process is interpretive and data must be analysed and “made sense of”, where the experience of groups under study will be important<sup>154</sup> (2004: 158). At the same time the process is iterative, processes identified are not “established once and for all” and interpretations also develop (ibid.). Additionally, data collected are “constructed for particular purposes and with particular ends in view”, requiring the researcher to address to issue that knowledge is often self-serving and therefore limited (2004: 159). I now turn to address these issues as they arise with regard to the identification of empirical investigation and further choice of methods.

Analysis therefore takes place within a process of continual design, taking an iterative (Rubin & Rubin, 2005) or looping (Richards, 2005) form. Within such

---

<sup>154</sup> The author recommends participant observation, but in this study it was not available as a method.

a process it is important that these loops are ‘purposive’ (Ibid.), which are iteratively and intellectually logical rather than linear (Newman et al.2003). Rennie (2009) considers that an underlying problem is the lack of a meta-methodology within qualitative research. This involves recognising a “hermeneutic operating field” which will include the critical realist perceptions applied within pragmatist thought, theories of inference and reflexivity (2009: 18). The author adopts the CR proposition, agreeing with the hermeneutic approach that interpretation is both realist and relativist, where researchers disclose their reflexivity (2009: 19-20). This builds further on Pierce’s theory of inference, which recognises the conduct of science as inductive, whilst incorporating abduction and non-formal deduction whereby a hypothesis is constructed to “explain a surprising finding” and an investigation deduced for a chain testing that will lead to an inductive revision of assumptions (2009: 13). Rennie describes Pierce’s view of scientific progression as the “interplay of induction and abduction, mediated by non-formal deduction” (ibid.). This also draws to some extent upon the hermeneutic turn associated with Gadamer and Ricoeur (Langdrige, 2007). Abduction as a continuous process starts from “an unmet expectation and works backward to invent a plausible world or a theory that would make the surprise meaningful”, where there is an interplay between the “observational and conceptual” built on rich data (Van Maanen, Sørensen, & Mitchell, 2007: 1149). Van Maanen et al. recognise that it is useful for the movement “back and forth from data-based theorizing to intuition resting on experience, habits of mind and research context” (Van Maanen et al., 2007: 1148). Interest, however, is still focused upon meaning rather than frequency of phenomena (Van Maanen, 1979b: 520).

With regard to main focus of the study, decision making perspectives are normally thought to be more subjective and arbitrary in this classification (Van de Ven & Astley, 1981: 436), where drawing on the work of Berger and Luckman (1967), meaning is assumed to be socially constructed and retrospectively imposed. This dialectical view is criticised by the critical realist approach, considering that individuals *recreate* social structures rather than create them (Cruickshank, 2003: 103ff). Turner also recognises that CR opens for the possibility of multiple perspectives, but considers them as “theoretical ways of framing reality”, and therefore not requiring any single representation of reality (2006: 417).

### **6.3 Alternate templates strategy**

This section considers the alternate templates strategy that is applied in relation to the theoretical background and literature review, data collection and analysis. It therefore draws together the findings from Chapters 2 to 5, but particularly the organisational decision models outlined in Chapter 5. Although the thesis begins without specific hypothesis and is based on general questions (Dooley, 1990: 282), alternate templates are applied to the processes under investigation (Langley, 1999). This is because this study is conducted towards a process

research framework, rather than being focused upon variance, as the attention is placed upon understanding the meaning of a process event (Maxwell, 1996) seen through “several alternative interpretations” while attempting to understand what is “going on in people’s heads” (Langley, 1999). Here the purpose of using of process research is aimed at identifying and attempting to understand the events and actions included in the subunit decision-making processes that lead to the choice and application of evaluation model. Collecting process data raises questions for identifying the unit of analysis, as it is difficult to define a fluid, protracted decision-making process where context is important. Adopting a qualitative process approach to account for and illuminate this continuum “leads inevitably, to the consideration of multiple levels of analysis that are sometimes difficult to separate from one another” (Langley, 1999: 692). In such instances defining the strategy is important.

The alternate templates strategy, is based on analysing and interpreting the same processes through “different but internally coherent sets of a priori theoretical premises”, which are then assessed to the extent “to which each theoretical template contributes to a satisfactory explanation” (Langley, 1999: 698). When applying this strategy it is important to keep the lenses separate, but each alone will be *insufficient* despite its relevance. The models’ explanatory power and accuracy is considered to be increased when they are applied in tandem. According to Langley, this application of the different, but complimentary models can lead to data interpretation that may reveal “contributions and gaps in each”. Langley sees this strategy as similar to Allison’s multi model approach and additionally drawing on Weick, describes it as a process of sensemaking. There are also links to the parallel studies models of Lewis and Grimes (1999) referred to in section 5.6.

This has particular relevance for this study. Weick (1976) recognised that critical analysis is required of language and communication that facilitates the decision process, and in order to do this different theoretical perspectives should be held. Such research opens for a combined strategy of deductive use of theory and inductive use of data (Langley, 1999). This is similar to Ragin’s retroduction (1994), as well as the interactive research process (Maxwell, 1996). A similar strategy also appears to have been applied by Peterson (1976), that led to his nuanced view of Allison’s third model. Instead of attempting to generalise, the intention is to develop propositions and limited theory by “[r]efining partial paradigms, and specifying the classes of actions for which they are relevant”, (Allison, 1971 in Langley, 1999: 699). This method also has a similar rationale to that used by Maitlis and Ozelik (2004) in their study of toxic decision making processes. The authors analysed aspects of three different organisational contexts that shaped the decision processes, noting that the decision process itself, and not just the issue under discussion, affected a certain outcome (Maitlis & Ozelik, 2004: 377). This strategy also appears similar to transformative approaches, which adopt a theoretical lens “as an overarching

perspective within a design<sup>155</sup> ... [where the] lens provides a framework for topics of interest, methods for collecting data, and outcomes or changes anticipated by the study” (Creswell, 2003: 16)<sup>156</sup>. Despite the fact that these models are often based on mixed methods, primarily qualitative methods will be used due to the richness and complexity of data to be obtained. These methods are outlined further in the next section.

## 6.4 Methods

In this section I outline the methods adopted and used within the data collection. I discuss the choices made when considering models in relation to the overall purpose and research problem. While interviews are the main methods used in this research, they are supported by documentary analysis and background research with regard to the subunits and the members within them. Analysis of programmes’ theoretical and ideological underpinning is discovered through use of literature, goals of course, policies and actual evaluation reports etc. This informs the interviews with course leaders, teachers and administrators (Denzin & Lincoln, 2000; Rubin & Rubin, 1995).

### 6.4.1 Sampling

This section deals briefly with the choices made about sampling and how these are linked to research strategy. An overview is also given with regard to number of informants and interviews that took place. The issue of sampling is always challenged in qualitative studies. Purposive sampling, better known as non-probability sampling, is based on the operationalisation of a given “criterion or purpose to replace the principle of cancelled random errors” (Teddlie & Tashakkori, 2003: 279). In qualitative studies purposive sampling is often selected for investigating “information-rich cases”, required for a more detailed study of the research focus (Patton, 1990 in Teddlie & Tashakkori, 2003). Samples should however be assessed as to whether they are appropriate in choice of paradigm and research model and adequate in terms of selection of unit of analysis and adjusted continuously until saturation, whilst searching for alternative explanations (Kuzel, 1999). Here research will need to ascertain information about the site and participants, including *setting*, *actors*, *events* that are under study and *processes* that unfold and evolve as the research takes place (Miles and Huberman, 1994 in Creswell, 2003). Here the sample size will often be minimized, non-randomly, to select the best illuminative examples (Teddlie & Tashakkori, 2003). In this instance the results of a de facto pilot study helped ascertain the focus of purposeful sampling.

Following Maxwell’s goals for purposive sampling, institutions are drawn from England and Norway in order to capture a semblance of heterogeneity in the

---

<sup>155</sup> “that contains both quantitative and qualitative data”

<sup>156</sup> While such an approach would normally involve mixed methods, here it is important to acknowledge the potential for their use rather their actual use.

theoretical population (1996: 71ff). Purposeful sampling is applied to ensure understanding of variation in the phenomena under study, while testing ideas about the setting through phenomena “crucial to the validity of those ideas” (Maxwell, 2002: 53). The purpose here is to attempt to avoid capturing only some typical members of a subset and defining the dimensions of variation considered to be most salient, selecting those who will represent them. As was pointed out in earlier chapters, the context of evaluation traditions is considered an important factor, alongside ideological positioning concerning programme content, to the decisions making process. While this can easily raise the complexity of a study, particularly with regard to analysing internal and external difference, it will inevitably also result in less available data about particular settings.

Initial analysis of 3 sites in Norway was formulated during data collection for the HEAD project<sup>157</sup> which took place from the autumn of 2005 to the winter of 2006. Respondents were approached as to their willingness to further investigate issues regarding the evaluation of programmes within this study. All were in agreement. On reflection two of the sites offering postgraduate programmes in school leadership as well as modular programmes, were chosen and approached for further focus with regard to interviews for this study. These also had a distance learning focus. One of the providers, NOR1, was chosen and approached, in what would become a de facto pilot study, where the interview guide was progressively assessed. Where adjustments were made, the questions were relayed back to those already interviewed for further reflection and responses. These were forthcoming. The interviews in NOR1 took place in the summer and autumn of 2007. The process of transcription is outlined below. NOR2 had also been approached and interviews took place during the late autumn / early winter of 2007. One particular member of the programme group had left prior to the interviews taking place and was no longer available as a respondent.

The choice of sites in England was based upon initial, informal discussions with members of programme groups, followed up by analysis of programme frameworks and materials, and institutional documentation. This documentary review narrowed the focus to 4 sites. To reflect some of the characteristics in Norway, the sites were chosen due to their offering of postgraduate and modular programmes, and application of distance learning. Informal discussion and further documentary analysis led to 2 sites being approached. One of these sites declined so a further site was chosen where after an initial response from the central members of subunit core team was positive. These interviews took place in the autumn / winter of 2007. In both ENG1 and ENG2 there were additional

---

<sup>157</sup> As was outlined in Chapter 1 analysis was based upon a comparative review of school programmes curriculum and programme documentation and supported by interviews with aim, purpose and theoretical basis of programmes. The Norwegian interviews were undertaken with Professor Arild Tjeldvoll.

members of the programme teams who did not respond to requests for an interview. One further, but more peripheral, potential interviewee also failed to participate after initially agreeing to do so. These issues will be further dealt with in the final section of this chapter.

The framework for data collection is outlined in the table below, while the implementation is discussed in the ensuing subsections.

**Table 9: Framework and timetable of data collection and sources**

Period	Country	
	Norway	England
2005	22 Interviews across 3 institutions concerning postgraduate programmes for school leadership <sup>158</sup>	Preliminary, informal discussions with key informants and background information related to 3 institutions <sup>159</sup>
2006	Identification of final institutions (3 outlined; 2 chosen). Secondary data collection: policy documents, protocols, academic articles and other writing.	Identification of final institutions; (reassignment after decline of request <sup>160</sup> ) Secondary data collection: policy documents, protocols, academic articles and other writing.
2007	Interviews with NOR1 and NOR2. Review of transcriptions; respondent feedback.	Interviews with ENG1, ENG2 and ENG3
2008	Follow up finalised	Final interviews arranged and follow up. Review of transcriptions; respondent feedback.

The overview of respondents is outlined in the table below. These are ordered by country in the interest of anonymity, mainly to avoid “counting” the numbers of members in the central core teams.

<sup>158</sup> Data Collection as part of Head Project (see footnote in Chapter 1).

<sup>159</sup> See footnote 118

<sup>160</sup> One institution initially approached decided on reflection that too few members of staff could participate due to pressures of work.

**Table 10: Overview of respondents**

Country	Number of respondents in the final data presentation	Number of interviews (including follow up)
Norway (NOR1 + NOR2)	15 respondents <sup>161</sup>	21
England (ENG1 + ENG2 + ENG3)	9 respondents	12

This research therefore builds on in depth interviews, with a few providers, within one field of study but across multiple sites. This raises the question over the extent to which one can generalise from the results. Gomm *et al.* (2000)<sup>162</sup> note that problems surrounding the issue of bias and measurement error can be limited by adopting theoretical ideas and information concerning case and population with the analysis<sup>163</sup> from previous knowledge or greater contextual understanding. This is achieved by systematically selecting cases upon this basis, recognising that no case will preserve all the features of the population. Rather than generalizability, the focus is to attempt to consider new challenges to theory (Andersen, 1997). The approach taken is idiographic, without attempt to generalise beyond the sample, but offer a description of the processes, building on shared experience (Langdrige, 2007: 58). The homogeneity of the sample reflects the need to find a “closely defined group” with significance for the problem statement and research questions (Smith & Osborn, 2008: 56). Langdrige recognises that this approach suits Template Analysis, which will be outlined in the next section. I also return briefly to this point in the section concerning limitations, delimitations and validity.

#### **6.4.2 Levels and unit of analysis**

Although it was noted earlier that units of analysis are often hard to distinguish in process research, the main investigation is focused upon the organisational decision making processes of school leadership training programme providers. The unit of observation is at the level of individual organisational members. Organisational behaviour is analysed with the help of the alternate templates strategy outlined above and through template analysis, outlined in the next section.

Van de Ven and Astley (1981) consider the problem of addressing the on-going issue of levels of analysis, particularly the distinctions between macro and micro level (1981: 458ff). This issue involves both analysis of data and application of theory. The authors also recognise, like Hardy (1990a), the division between

---

<sup>161</sup> In Norway access had been possible with staff that had recently been involved in a particular programme and part of the evaluation decision processes.

<sup>162</sup> The authors focus mainly on case study research, but address the problem of the qualitative method more generally as well.

<sup>163</sup> Especially with regard to heterogeneity and assessment of representativeness.

deterministic and voluntaristic orientations of human nature (1981:429) and its impact on research. Deterministic research focuses more on position and structure than the perceptual focus of voluntaristic based research. Within the classification of research raised by Van de Ven and Astley, this study is more closely related to the “strategic choice view”, centred on investigating interaction, process and meaning at the micro-level rather than position, interrelationships and functional behaviour, more associated with a “system-structural view” (1981: 437). This study emphasises a “micro-level / voluntaristic” methodological orientation, focused upon individuals’ perceptions of sub-unit processes. However, as the authors point out, in order to enable greater understanding of the contexts of these processes some theories of wider structures and understanding of the macro-organisation will be necessary (1981: 458-9). As a result, reference in this study is made, for example, to the macro-contingency perspective of decision making and assessment in organisations, outlined by Thompson (1967).

While the level of analysis is subunit decision making, that is, subsets within HEIs, the units of observation are the individual members within these subunits. The purpose for this study is to gain understating of the experience of decision-making process, in two complimentary contexts. As has been outlined in the previous chapters, literature and policy review suggests tighter control through, for example, nationally mandated standards in the case of the England and looser control grounded on intentions rather than formal standards in the case of Norway. Recognition of dissimilar or divergent practice in evaluation stems from the different national cultures and traditions, whilst the field also exhibits similarity related to “task uncertainty and development of international networks” (Hansen, 2009: 74).

#### **6.4.3 Secondary data collection and contextualisation**

Data collection began with analysis of organisational presentations, evaluation frameworks and documentation concerning the evaluations that have taken place within the organisations, for example meeting agendas, minutes, policy documents and guidelines, internal evaluation reports and externally available documents. These were also supplemented during the rounds of interviews and during transcription where possible and necessary. Analysis of documents is one further part of building up a picture of the research setting. But more than just providing “background” information, documents as secondary data can be assimilated into the wider research. Rather than merely considering documents as “accurate” portrayals of reality, they ought to be considered as “texts” that “construct their own kinds of reality” (Atkinson & Coffey, 2004: 73). Atkinson and Coffey note that focus ought to be placed on their form and function as much as their “truth” or “validity”. Prior discusses how texts have these “dual relations”, they can be considered as “receptacles”, with regard to what they contain, but also as “agents”, due to the effects that they have in their own right (2004: 76). Researchers might therefore additionally focus on how they are produced and also how they “function” in particular circumstances and are



“consumed” (Prior, 2004: 91). So while these provide useful contextual information for understanding the particular background of the respondents, they also inform the interview guides and subsequent analysis (C. Marshall & Rossman, 2006: 107).

Analysing secondary data also informs the subject area of this study. Miller argues that greater sociological understanding of how institutions function can be gained by analysing texts, while “combining an empirical focus with an analytical attitude” (1997: 77). This requires analysing the texts in combination with analysis of the organization and its members or some part of their activity, building on the linkage between text and its social context, or ‘interpretive domain’, that structures how the text will be “assembled and interpreted”. Rather than a deterministic model, Miller suggests that institutional settings will *encourage, privilege or prefer* certain interpretations of texts by providing “categories and procedures” for classification (1997: 79). These differences across “decision horizons” are noted within a micro-political perspective, in this instance reinforced by interview techniques. Miller noted that the significance of institutional texts to decision making will contrast from institution to institution, reflecting on instances from varied observational research where personal experience, or “social cues”, appeared to be given greater weighting, while other situations where the texts, or “technological cues” were consistently considered of greater importance (Miller, 1997: 82).

When analysing texts it is considered therefore important to follow up the perceptions of their importance within the organisation through self-reporting. Although this does not involve observational analysis, it gives greater insight to the purpose, place and use of documentation as well as an understanding of the institution under investigation. Each of the rounds of interviews was preceded and then further facilitated by gathering data in the form of organisational documentation, protocols, agendas and minutes, policy documents and evaluation forms and discussion papers relevant both to the subgroup in question as well as the wider institution that they were placed in. These were also framed within policy documents, both national and from the supranational level, as well as information gleaned from accreditation bodies. While the texts informed the interview guides, they were also used as discussion bases, asking respondents to consider their formation, implications and use. Due to a number of informants requesting institutional anonymity, as well as others who would have been too easily identifiable from their responses given the institutional name, the documentary analysis of this part of the data collection has been excluded from the final presentation of the study. Key informants were asked to comment upon processes where written evidence was lacking.

#### **6.4.4 Interviews**

The purpose of the interviews was to investigate further the research problem of this study concerning the decision making process with regard to evaluation within organisations. As has already been outlined, the frameworks are codified

to a greater or lesser extent within the organisational documents. However, the interviews were focused upon how these frameworks came into being, who was involved and how the processes developed. The interviews therefore required the participants' impressions of how these unfold. In such cases, the conceptual framework should make explicit the focus on uncovering and describing participants' "subjective" perspectives (C. Marshall & Rossman, 2006: 102). Semi-structured interviewing offers a "trade-off between consistency and flexibility" (Langdrige, 2007: 65). Systematisation will however be necessary on a multi-site study with many participants (C. Marshall & Rossman, 2006: 101). With the study being across two different countries the interview guide required preparation in both English and Norwegian. Translation of the interview guide had been checked in terms of content and language use by peer review. Issues of translation like connotation and meaning from "source to target" language (C. Marshall & Rossman, 2006: 111) were also taken into account at this initial stage. The English translation of the interview guide is presented in the appendix.

Studies into naturalistic decision-making may be helpful, as they move the focus away from the 'decision event' and attempt to discover what decision-makers actually do (Orasanu & Connolly, 1993). This has led to an increased use of qualitative field studies within decision-making research studying phenomena in context, although the majority of these have drawn on methods from anthropology, ethnography etc. (Lipshitz et al., 2001). The authors also recognised that "field observations are critical to NDM research because real-world decisions are embedded in and contribute to ongoing tasks. Researchers must understand the environments that demand decisions, the affordances and constraints of those environments, and the kinds of knowledge and skills needed to respond to those demands" (Ibid.: 343). While observations were not used in this study it is interesting to note that the authors recognise that interviewing is also a useful method to draw out such information. Focusing on the underlying values influencing decision-making requires an understanding of the decision-making process within the organisation. This involves discovering the core function, basic tasks and internal relationships within the organisation as perceived by organisational members, as well as understanding external relationships with the environment. Interviews were, then, chosen as the core method.

Interviews were conducted by attempting to balance *responsive* (Rubin & Rubin, 1995, 2005) and *active* interviewing (Andersen, 2006) strategies. Rubin and Rubin's (1995, 2005) model of qualitative interviewing builds particularly on a constructivist philosophy within a pragmatic model and reflecting many models from that persuasion. The authors view qualitative interviews as shared social experiences or an in depth partnership, where the process is partner led. The purpose of the interview is to "tap lived experience", balanced between control by the interviewer and approximated normal conversation (Madill & Gough, 2008: 256). There is a tension as to how interviews should progress. A

common view is that respondents' views should emerge in an *emic* form rather than the *etic* perspective of the researcher (C. Marshall & Rossman, 2006: 101). Shaw more precisely suggests that an emic perspective can be used to illuminate official definitions (1999: 14); in this case it will be appropriate to understand underlying perspectives of subunit members with regard to the formal evaluation and quality assurance frameworks of their institutions, as well as wider codified and non-codified demands.

*Responsive interviewing* highlights the importance of the context, while exploring themes and concepts arising during the process and searches for the *subtle* as well as the *obvious* within an ongoing analysis. Therefore design, data gathering and analysis will be linked *intimately* and modified as the project develops. The model is constructed around building *main questions* to address the overall research problem; *probes* that manage the conversation and provoke new detail; and *follow-up questions* that explore ideas further and develop new lines of inquiry.

Andersen (2006) concurs with the general pattern for interviews laid out by Rubin and Rubin, but gives the researcher a greater role in directing the focus of the interview. *Active* interviewing places an "active perspective on the interview situation, but utilises this understanding within a more conventional, sociological perspective where subjective perceptions of reality and active data construction represent empirical patterns that can be generated and tested in relation to analytical assumptions" (2006: 295). Respondents are chosen because they are well informed about the area of investigation; they are considered to have something particular to say on the subject at hand. Andersen drawing on the dualistic approach of Weber, considers an important distinction with mainstream constructive approaches to be the development of an analytical structure, however loose or tight the frame of the interview is (2006: 285). This approach is considered to give greater analytical control. In such cases it is important for the research to attempt to distinguish between the factual and interpretative parts of respondents' accounts (2006: 284). In such cases the "facts" are considered to be "constructed" data that are correct within a given frame, as one recognises the cognitive limitations regarding recall of events. Even though the researcher has a more leading role, this does not mean that the "leading" questions should be asked (2006: 290). What it does mean is that the researcher is prepared and has some prior knowledge of the area under focus, including documentary research, as well as understanding of theoretical perspectives that can be explored and challenged. This balance of positions appears well suited to a critical realist approach and also the use of template analysis, to be outlined in the next section.

The vast majority of interviews were face to face interviews, but due to issues of scheduling and availability 3 were telephone interviews. The data from the interviews was recorded and then transcribed. The data preparation (Gibbs, 2007) began after the first interview was completed. Two major issues need to be balanced, ethics and precision; involving "issues of accuracy, fidelity and

interpretation” (Gibbs, 2007: 11). Transcription is considered the first stage of analysis (Langdrige, 2004: 261) and the processing of data (C. Marshall & Rossman, 2006: 110) and is additionally a “change of medium” that transforms data (Gibbs, 2007: 11). Within a phenomenological study to be analysed by TA transcription is verbatim, but at a simpler level of detail than those performing discourse analysis, as content is of most interest (Langdrige, 2007: 73-4). Poland (2003) reflects that the possibility of considering an interview transcript as verbatim is generally associated with realist approaches. At best they are considered “partial accounts”. Nevertheless, he recognises that they can still provide rich, quality data. Poland therefore considers the main issue of concern to be assuring the quality of transcript and being aware of possible threats. Threats to quality are discrepancy between the spoken and transcribed account, the “interview-tape-transcript interface”, in relation to misinterpretation of word use, difficulties with reconstruction of sentences, respondents quoting others, omissions and lost data during recording difficulties and replaying of recordings (Poland, 2003: 270-1). Following Poland’s ensuing advice, periodical checks of recording levels were undertaken during recording, post interview these were listened to while the content of the interview was fresh in mind and compared to field notes taken. As seen below respondents were also encouraged to quality assure transcripts. As Poland points out however, respondents will also be subject to the same difficulties recollecting events (2003: 282).

### ***Transcription***

This study involved both the transcription of data as well as the subsequent translation of that collected in Norway<sup>164</sup>. Following the advice of Strauss and Corbin (1998: 285ff), translation of the Norwegian transcripts was only of key passages used in the final presentation of data. In both cases all transcription was undertaken by me. Due to ethical issues the transcriptions were made anonymous. In the first instance transcriptions were given codes and personal names were omitted during the first round. Each transcription was then reread. This was part of quality assuring the transcript as well as in order to edit out words and phrases thought to ease identification of persons. Revised transcriptions were returned to respondents for comments. Problematic phrases, areas for misunderstanding or issues of clarity from the recordings were discussed with respondents as were suitable translations into English where necessary. Where any difficulties of translation occur, the original phrase is included in italics with a footnote to explain that the closest possible translation has been used to get a sense of what is said. Some code switching is inevitable which ever language is being used (Welch & Piekkari, 2006). As Marshall and Rossman note, the overall purpose is to generate “accurate and meaningful data” (2006: 111). In addition, respondents were asked to indicate any phrases or references they wished to be omitted from the study, in addition to any sections where they considered they could be recognised. Three transcriptions were

---

<sup>164</sup> All of the interviews in Norway were conducted in Norwegian.

adjusted with regard to these reflections. Grammatical errors were also rectified for the data used in the final presentation related to comments made by respondents.

## **6.5 Analysis: reflection, coding and revision**

In this section I look at the analytical processes adopted in the study, including the reasoning for their choice and examples of how they were operationalised. I will also return to their use in the data chapters. After fully transcribing the first interview accounts I began the process of coding the transcripts. Miles describes the process of analysing qualitative data as “a mysterious, half-formulated art” (1979: 593).

Van Maanen (1979a), in a more detailed exposition of ethnography, recognises principles for analysis of qualitative data. Van Maanen recognises the importance of distinguishing first order informant conceptions (“facts”) from the second order conceptions of the researcher (“theories”) about the processes taking place<sup>165</sup> (1979a: 540). The researcher must further be aware of the difference and separation between presentational and operational data with first order concepts. The author outlines the process of using second order concepts, or “interpretations of interpretations” to explain patterns within the first order data (1979a: 541). The categories are, however, those of the researcher. Van Maanen further notes though that it is the respondents that are socialised within their normalised, natural setting and these backgrounds and process formed observations inform the categories chosen (1979a: 542). In this current study the data to some extent are presentational in form. While this is mainly an ethnographic construction assessing the relationship between informant and researcher, there also seems to be a degree of similarity in interview based studies, where the researcher must gauge the degree to which data are “ideological, normative and abstract” (ibid.). Data were, however, also assessed through member checking. This also avoids what Van Maanen considers as “taking for granted” respondents interjections and sensemaking of events and processes. While observation of meetings regarding evaluation processes, for example, might have offered greater depth the limitations of working across four sites was considered to preclude this. The analysis of first order concepts and development of second order constructions are outlined in the data chapters.

### **6.5.1 Coding**

Coding is the descriptive and sensemaking process of systematic data categorisation or labelling of textual data (Langdrige, 2004: 262). Coding is used to reflect over the meaning of text, to develop questions in relation to other emergent ideas, to align all material, to develop and blend categories and search for patterns (Richards, 2005: 87). Textual data can be linked to theoretical ideas

---

<sup>165</sup> Within ethnographic studies based upon participant observation, as Van Maanen is most focused upon, this factor becomes increasingly more important.

(Gibbs, 2002: 57). Working up from data retrieved is considered important, where theory refinement cannot be expected to *emerge* but must rather be ‘goaded’ out as a human construct (Turner, in Richards, 2005: 67ff). However, where there are already theories that might be thought to be applicable to an area of study an analytical approach can be made whereby the theory and data interact.

The coding process is eased by using computer assisted qualitative data analysis software (CAQDAS). In this study NVivo 7 has been used. NVivo is useful for qualitative analysis, particularly where focus is on meaning and perception within the data. As such it is useful for “fine-grained and intensive analysis” and “encourages an exploratory approach” (Gibbs, 2002: xxiii). The action of coding produces codes from actual text, known as “nodes” in NVivo. This is the centre of analytical thinking in the process, within and across data sources (Gibbs, 2002: 58). Some nodes also function structurally, as within a node “tree”, as framework to aid analysis and control over data (2002: 59). Gibbs further outlines how this process produces a conceptual schema, which as in the case of this study, can be directed from literature review and background research. Such an a priori approach to coding fits in with an analytical procedure such as template analysis which is outlined below.

The use of CAQDAS related to grounded theory has been criticised for being over mechanical (Holton, 2007) but these fears would appear to be offset by the increased control over data that are available. While Langdridge recognises that the coding process needs to be “creative”, he notes that it should additionally be “consistent and rigorous” (2004: 267). Richards (2005) also calls for focus on coder reliability and consistency checks, which should be iterative. These are supported my multilevel coding strategies. Langdridge (2004) outlines the three levels, or orders, of codes in forms of thematic analysis, where different levels of interpretation are applied at each. The first order consists of descriptive coding, categorizing and ordering textual data, and revising codes as the process continues. The second order introduces a greater degree of interpretation, introducing “super-ordinate” constructs of the developed codes. The third order sees the super-ordinate constructs develop into patterns, which introduces the theoretical perspectives into the analysis while attempting to remain data grounded. Within Thematic Analysis the next step is to draw out the major themes from the patterns in the data, which subsume the lower levels (2004: 270ff). These themes then form the basis of the data presentation chapters 7-9. As Langdridge notes, in qualitative studies results and discussion are usually combined (2004: 271). These ideas are further extended below with regard to the procedures for template analysis, a specific application of the principles of coding.

### **6.5.2 Template analysis: basis and procedures used**

Madill and Gough outline the “procedural categorization” of thematic analysis, considering the associated methods to focus upon on coding qualitative data to

produce “clusters of text with similar meaning”, while searching for concepts that will illuminate, or “capture the essence” of that which is investigated (Madill & Gough, 2008: 258). Within this overall structure the authors place the methods of analytic induction, framework analysis, grounded theory, thematic analysis (in its own right), theory-led thematic analysis, template analysis (TA) and interpretive phenomenological analysis (IPA). In this study I refer to the latter two. TA and IPA are more commonly linked to phenomenological philosophy and psychology (King, 1998; Langdrige, 2007; Willig, 2001), but are gaining wider interest within the field of organisation studies. An interpretive phenomenological approach considers that direct access to respondents’ “life worlds” is not possible, and therefore what is produced is rather an interpretation on behalf of the researcher (Willig, 2001: 53). The focus of the data collection therefore requires methods allowing for naturalistic description and interpretation based upon the meanings of those experiencing them (Langdrige, 2007: 2). At the same time template analysis is described as more flexible and less prescriptive than, for example, grounded theory (King, 2004: 257). There would appear to be cross-over with some tenets of CR here, recognising the complexities of accounting for the natural world, and the contextual specificity of the social world. Methods and analytical tools will be required that allow for perceptual and narrative accounts of respondents’ experiences of events and processes, as well as the intention of the researcher to attempt to “bracket” presuppositions about the context and phenomena under focus (Langdrige, 2007).

There is a great deal of similarity between TA and IPA, with their interpretive focus on “producing a thematic analysis of experience”, the main differences though being that TA begins with an a priori template for coding (Langdrige, 2007: 56) and that within IPA individual cases are analysed in “greater depth” before integration of the case set (King, 2004: 257). Sample size also differs, IPA studies generating 10 or fewer responses, and TA studies between 20 and 30 (*ibid.*). The template allows for theoretical exploration as well as emergence of meaning within the process of data collection and analysis (*ibid.*). It is considered to follow a “double hermeneutic” (Langdrige, 2007: 108), where sensemaking activities take place by and between respondent and interviewer (Smith & Osborn, 2008: 54). TA is the process of analysing qualitative data through developing a coding template, a summary of themes in a data set, highlighted by the researcher (King, 2008i). This template outlines the relationships between themes, usually in a hierarchical structure (King, 2004). Within this structure “broad themes... encompass successively narrower, more specific ones”, where “themes are features of participants’ accounts characterising particular perceptions and/or experiences that the researcher sees as relevant to the research question” (King, 2008i). In addition, “‘coding’ is [outlined as] the process of identifying themes in accounts and attaching labels (codes) to index them” (King, 2008i). TA is considered especially relevant for ascertaining the perspectives of different groups within organisational research (King, 2004: 268).

King identifies and outlines 7 steps within TA: defining a priori themes; transcribing interviews; an initial coding of the data; producing a preliminary template; developing the template; and presenting the final template within the writing up phase of the research (King, 2008f). The final step recognises that throughout the whole process the researcher should also address issues of reflexivity, assuring quality control (ibid.). Following King's suggestion, this study uses the interview topic guide as the basis for the construction of the initial template, which has in turn drawn on the philosophical orientation, literature review and exploratory data; the main questions forming the "higher-order codes", and the sub-questions and probes forming the "lower-order codes" (2004: 259). King sees this applying well to studies where the majority of topics for investigation are defined beforehand by the researcher. With regard to issues of transcription King considers that while demands are more limited related to discourse or conversation analysis, a full verbatim transcription is necessary as "it is not usually possible to be sure which parts of an interview are relevant to the research question until quite some time into the analysis" (King, 2008h). Parallel coding, or classifying the same portion of text under different themes, is also possible within TA, depending on the ontological and epistemological approach (King, 2004: 258). It was considered an appropriate procedure in this study and was enabled by the use of NVivo. TA is well served by the use of NVivo, allowing for manipulation of data and sophisticated coding procedures and within and cross-case analysis (Langdridge, 2007: 83).

It should not be considered that themes "hide" in the data, and therefore the when defining themes and codes, the focus is related to the research problem, phenomenological point of interest and existing theory (King, 2008g). Care should be taken to avoid overlooking areas falling outside of the a priori themes, which should be considered "tentative" and subject to both "redefinition and removal" (ibid.). King does not regard a template as a theoretical model, but rather a "representation of the way you have gone about coding the data to identify themes in it", which may develop into a model (King, 2008b). Parallel coding, or classifying the same portion of text under different themes, is also possible within TA, depending on the ontological and epistemological approach (King, 2004: 258). It was considered an appropriate procedure in this study and was enabled by the use of NVivo. The initial template, while reflecting the research focus, should not interfere with a thorough analysis of the data, and should recognise that not all data will fit into the framework (King, 2008d). The template is then further adjusted according to the progress of the research.

The template is developed in order to find the best representation of the themes within the data (King, 2008a). As King notes, this process can always be repeated but there are some important approaches to take. Firstly, the template is applied to each transcript, coding the parts that are relevant while also modifying as new themes are discovered or existing themes are developed. As each adjustment takes place each transcript is rechecked and the coding adapted.



This is made easier within NVivo where codes can be viewed across all transcripts and changed for all documents or split up further. King recognises how NVivo can aid discovering relationships among themes building also on the template structure (2004: 263). When revising the template King suggests a fourfold process (2004: 261ff), which is outlined in the table below including examples from this study. These points are further discussed within the ensuing data chapters.

**Table 11: Template revision (after King, 2004: 261ff)**

<b>Template Revision process</b>	<b>Description</b>	<b>Example from current study</b>
<i>Insertion</i>	Issues found to be relevant during analysis but not included in initial template form new categories.	One sub-unit reflected a different type of internal decision process to the others that came to be called “collegial construction”.
<i>Deletion</i>	Redundant codes from the initial phases are removed	Overtly descriptive categories were removed, as were those recognised to be parallel or synonymous.
<i>Changing scope</i>	Codes are redefined at different levels if thought too narrowly or broadly defined.	The code “response to commissioners” became a higher order theme in response to reflection over the transcripts.
<i>Changing higher-order classification</i>	Sub category codes are redefined from one ‘tree’ to another.	The category “designs” moved from merely a descriptive category to a more process focused code.

While following these procedures it is, at the same time, important to avoid presenting merely descriptive accounts, which can develop from just providing a summary or index of themes in the transcripts (King, 2008c). The template is a tool to aid a rich interpretation of data. King suggests a threefold strategy, which underlies the revision process outline above, to improve and assure the whole process: listing themes, prioritising and openness. Iteratively listing themes provides an overview as well as elucidating patterns. As King points out though, this does not suggest variables to compare sub-groups nor does frequency indicate salience of themes. Rather, prioritising is the assessment of salience. Firstly, themes are looked in the context of an individual account, and then later across the accounts, to determine which themes best interpret the stories told. King suggests this avoids focusing too closely on what is common rather than understanding the context through which themes emerge. The final part of the strategy is to avoid narrowing the analysis too soon; avoiding creating a self-fulfilling prophecy. It is therefore important to maintain an open stance, to search for themes that were initially missed during the research design process

and theoretical review. These perceptions can be added to the research framework. These processes continue until saturation point is reached, or as King puts it, the law of diminishing returns is fulfilled (King, 2008a). At this point all of the transcription text are coded, as well as the codes being revised and the data analysed further, “at least twice” but usually three or four times (King, 2004: 263). The final template may look like a linear, hierarchical system and so care must be taken to present the data in such a way that reflects the interconnectedness and integration of themes (2004: 267). These processes were followed in this project. The initial and final templates are presented in the appendix.

## **6.6 Quality, delimitations and limitations**

In terms of ethics this study was reported to, assessed by and accepted by the Norwegian Social Science Data Services (NSD)<sup>166</sup>. The respondents were informed of this, which included their anonymity, right to respond to data presented and their unreserved right of withdrawal from the project at any time. Respondents were asked to sign an agreement of notification and acceptance in regard to this.

Within debates about methodology and methods used issues of quality and validity in qualitative research have been central. Realist focused qualitative studies are based on detailed interpretive work, with the aim of producing results “not wildly idiosyncratic” (Madill, Jordan, & Shirley, 2000: 9). The authors note in realist based studies that there ought to be some element of broad reproducibility, despite the flexible and less structured approach taken. They recognise however that critical realism “admits an inherent subjectivity in the production of knowledge” and therefore is linked to constructive approaches (2000: 3). Therefore issues of reproducibility must be weighed against interest in “permeability”, where propositions change by new encounters and knowledge is considered more situation dependent, related to the varied understandings and interpretations of participants, the researcher, and the influencing meaning and scientific systems (Madill et al., 2000: 9). To Maxwell, from a realist perspective, it is understanding from inquiry rather than validity that should be in focus in qualitative studies (Maxwell, 2002: 39). Adopting a critical realist position, Maxwell considers validity as relative to purposes and circumstances; referring to “accounts” within a particular perspective rather than “data or methods” (2002: 42). Maxwell outlines how researchers should focus on three main types of validity with regard to a particular phenomenon. The first is descriptive validity, focused on the accuracy of account, the second interpretive validity, with the aim of ensuring emic exposure of data, the third theoretical validity, concerns the quality of the researcher’s explanation and concepts used to define some phenomenon.

---

<sup>166</sup> Norsk Samfunnsvitenskapelig Datatjeneste

Maxwell also raises the issues of generalizability and evaluative validity (2002: 52ff). Generalizability deals with the possibility to extend findings beyond the situation that has been purposely sampled, to make sense for other situations, which can either be internal or external. The focus however is placed more heavily on the internal. This means rather than an exact replication, “insights” derived might be useful for similar contexts (Yardley, 2008: 238). Maxwell recognises that interviewing reduces this capacity due to the brief encounter with the participant’s world (2002: 54). Evaluative validity reflects the interpretations and judgements that the researcher makes about the subject at hand (2002: 55), seemingly linked to reflexivity. Maxwell recognises that these processes are usually applied *posteriori* than *a priori*. Richards (2004) argues that validity can be demonstrated through increasing the “scope” of the data, “interrogating interpretations” by thorough coding, “establishing saturation” and developing audit and log trails. These processes are aided by use of NVivo, as outlined earlier. As has already been noted, NVivo is considered to support the approach of template analysis and I turn now to discuss how these questions are dealt with.

In this study validity refers more to determination of whether findings are “accurate from the standpoint of the researcher, the participant, or the readers of an account” (Creswell, 2003: 195ff). In this research reliance upon depth of description, combined with follow up interviewing and possible *member checking* (or respondent validation) is thought helpful to check the findings’ accuracy, particularly when faced with key informant bias (Maxwell, 1996). It is also important that the context or natural setting is presented faithfully (Creswell, 2003; Richards, 2005). Questions of validity within this research may not be the same as is ‘traditionally’ understood; focus is rather placed upon “on interpretations and definitions of situations by expert decision makers, and the impact of those interpretations on task performance” (Lipshitz et al.2001). Within such interpretive studies there will be different requirements concerning ‘reliability, falsifiability and objectivity’. Following Mishler’s research, Lipshitz *et al.* look for credibility in findings and conclusions, which demands detailed explanation of the formation of research questions, methods, plausibility of answers and reasonableness of the assumptions. The authors recognise this is often a “judgement call.” Liptshitz *et al.* also note how Mishler focused on *transferability*, where question is raised as to how findings and conclusions apply in other contexts, which may be seen via “case-to-case translation” by feature.

Template analysis offers a supportive framework for dealing with these issues, where addressing issues of quality and reflexivity are considered to be important throughout the process of analysis (King, 2008e). King considers 4 important areas in TA. The first is independent scrutiny of analysis. In this study it takes the form of supervision, reflection with other faculty members and discussion within the academic group. The second is respondent feedback. King prefers the term “feedback” rather than “validation”, considering the latter to create

problems in relation to validity. Although King recognises that respondent feedback is important, this can also be limited by unknown motives for the way the interviewee wishes the research to be framed. It is important to note however that these comments relate to the “template” and not to their responses or comments on data provided by other interviewees. On a less suspicious note, it can also be difficult for respondents to take a detached view and “objectively” reassess their own “lifeworld” to consider whether their interpretation is correct. King does not doubt the role feedback can play rather than the expectations of what the response can provide. In line with King’s third suggestion, an audit trail was followed with regard to recording the steps taken and noting decisions made in the process from raw data to final interpretation (King, 2008e). NVivo provides the opportunity to save summaries of the research template at the different stages of the process, as well as in writing memos to inform the reason for these choices. In addition codes, or nodes in NVivo terminology, that become redundant can be saved as “free nodes”, remaining linked to the transcript data they were first applied to but separate from the main framework, or “tree”. Due to issues of anonymity I have not included a coded transcript for purposes of example, as King suggests as a possibility. Finally, King also raises the issue of reflexivity, and how the researcher reflects over their work throughout the process. This is important considering that the template is developed a priori. King considers that the processes described above aid reflexivity, but also that TA as already outlined generally requires explicit reasoning for decisions made as the template develops.

Delimitations narrow the scope of any research (Creswell, 2003: 148). This project is delimited to the central phenomenon of decision-making about the evaluation processes to be enacted in relation to postgraduate programmes. It is further delimited to a study of those responsible for such decisions within school leadership programmes at the provider level. This takes place within an interpretive design (Maxwell, 1996) that is pragmatic in nature (Creswell, 2003; Teddlie & Tashakkori, 2003). Sites are chosen among those providing postgraduate programmes, and are delimited to a purposive sample of institutions in Norway and England expressing heterogeneity within the field.

At the same time it is important to recognise the limitations of the study. The extent to which this research can understand evaluation as decision-making processes is a key question. The aim of this research at one level is to contribute to a widened understanding existing evaluation theory with regard to understanding the impact of decision-making processes upon approach to evaluation within the organisation. On a more specific level this research aims to better understand how such choices are made in terms of evaluation impact and justify the underlying rationale of programmes provided. Such a design is, according to critics, reliant first and foremost upon the ‘quality of information’ about social processes provided by interviewees and conversation partners. While qualitative interviewing is thought to strengthen the likelihood of describing such social processes (Rubin & Rubin, 2005) the underlying

constructivist interpretive rationale is frequently challenged to the extent to which anything can be explained. However, this research must be presented in the words of the actors involved. The flexibility and iteratively oriented activity required for such an approach produces a problem of information overload into which any study can be overwhelmed (*Ibid.*). In order to gain the more in depth and rich description of the decision making processes the sample involved is limited to four institutions across two countries. Although this limits any attempt to generalise results to a wider population, it must be reiterated that the purpose of this research is to discover elements within the decision making process that may challenge existing theory and further it substantively rather than formally (Richards, 2005: 129). As such this research is far from purely descriptive and is thought to work towards a significant challenge to existing understanding.

## **6.7 Writing up**

Langdridge considers that the account of the research needs to be “persuasive”, demonstrating the rigour of analysis and presentation (2007: 80). King (2004, 2008j) outlines suggestions of different approaches in terms of writing up. A basic framework requires that TA studies focus upon rich direct quotes from respondents, where “longer quotes... give the reader the flavour of the original accounts” (King, 2008j). The writing up is considered to be a “continuation of the interpretative process”, which might lead to refinements in the template. Wolcott considers that this “nexus between description and analysis” is “dialectic” (2001: 112). Awareness of this interplay enables the writer to focus during the “tightening phase”, as well as avoiding over interpretation of data. It builds upon retroductive and iterative ideas outlined above, but drawing them in to this final phase.

In this study I have adopted a presentation framework similar to King’s synthesis model (2004: 268; 2008j). It mainly offers a thematic presentation of findings but is supported by cases described by members from the different subunits involved. Additionally, this allows for looking at the issues across the subunit. The tension is to present the data clearly and succinctly, while avoiding the temptation to over-generalise. The difficulty can be to make accurate choices about which data to use. This, however, is approached by following the procedures outlined above. In this study the data chapters are presented around the major themes in the template: demands, designs and decisions. These issues were identified early on in the process but as will be seen were developed, enriched and adjusted as the data collection and analysis took place. This is discussed further in chapters 7 to 10.

## **7. Pressures and demands for evaluation?**

In this chapter focus is placed upon the varying pressures and demands that members of the subunits delivering postgraduate programmes in school leadership perceived were placed on them concerning evaluation. Respondents discussed the demands placed upon them as programme providers. The focus was placed upon demands related to the evaluation of their programmes which involved wider discussion and reflection over their genesis and implementation. As is fitting with semi-structured interviewing, and outlined in chapter 6, overall themes were presented to respondents as well as probes and follow up questions. As is also appropriate with active informant techniques, the focus of the discussion was directed back to reflections concerning evaluation. Focus was therefore placed upon the content of these demands as well as those experienced to be making them. With regard to the latter, first order analysis revealed and confirmed two broad contexts from which demands and pressures were placed, within-institution and from the task environment, which is further defined throughout the chapter. It was recognised from analysis of the data that these pressures and demands might be direct and indirect. Further clarification concerning these categories will be given in this chapter.

In earlier chapters of this thesis various potential sources of pressure and demand have been identified. In chapter 2 it was recognised that there has been pressure for improved results in schools, where school leadership has been considered to be a contributing factor to this improvement. As a result educational policies were seen to have been focused upon improving the training and development of current and future school leaders, with increasing discussion concerning how to ascertain the impact of programmes implemented to meet these needs. In chapter 2 outlines were made of how policies concerning school leadership training and development are implemented within a context of increasing focus upon and institutionalisation of evaluation in wider society. The subunits under study are all part of HEIs and influenced by changes in their evaluative frameworks as a result of the Bologna Process. In chapter 4 the major facets and themes surrounding the development of evaluation and quality assurance within Higher education systems were outlined, in this case as they affected England and Norway. Demands placed upon subunits are shaped by these developments. These can be mediated through the wider organisation that the subunits operate within, but also appear to be shaped by and shape the demands of the programme participants themselves.

As was outlined in Chapter 6, the interview guide was developed along with a preliminary coding template based on the literature review and formed within the reapplication of Stufflebeam et al.'s evaluation problems, as outlined in section 5.9. The interview guide and preliminary coding template are included in the appendix. During the phases of analysis this template was developed, as outlined in section 6.5.2, before reaching the final template. This template is also

included in the appendix. Each of the sections presented in the next three chapters includes an outtake presenting the codes drawn from the respective section in the final coding template<sup>167</sup>.

This current chapter addresses the issue of demands related to evaluation frameworks perceived by subunit members. Of key interest was the direction and focus of these demands, considering who was making them and what their substance was. These areas are presented more descriptively and are developed further in subsequent chapters with regard to their influence on chosen design and consideration of the way that decisions are made.

Different terms were explored to discover the diversity of pressures felt by members of the subunits, attempting to consider any nuance between demand and influence. In discussing demands for evaluation there was often overlap with consideration for programme content. As will be seen, respondents saw these often to be interlinked; what was been asked of them to deliver and what they felt bound to demonstrate and in what manner they should demonstrate this and be accountable for them. While these are considered to overlap they are emphasised differently at different levels and across boundaries. This is discussed further in the subsections of this chapter. With regard to programme participants it was considered particularly important to attempt to distinguish between demands for the programme and concerning the evaluation of the programme. Respondents from the four different subunits under study highlighted different demands and pressures from inside the organisation and from the environment. In this chapter I deal first with those demands perceived to come from the task environment before dealing with those considered internal.

## **7.1 Perceived pressure from the task environment**

In this section focus is placed on those groups that are perceived by sub-unit members to place demands or pressures upon them with regard to evaluation of programmes from outside of the organisation they are situated within. As was outlined in chapter 5, a problem often develops as one tries to define the environment of an organisation. In this study I have chosen to use the concept of the “Task Environment” as understood by James Thompson (1967, 2003). Thompson draws on Dill’s concept of Task Environment in order to deal with the problem of having a residual category, as the environment is often conceived to include “everything else” outside of the organisation under investigation (Thompson, 2003: 27). Thompson recognises that this delimitation focuses attention on the parts of the environment “relevant or potentially relevant to goal setting and goal attainment” (ibid.) and will encompass customers/clients, suppliers, competitors and regulatory groups. While such categories might appear more applicable to manufacturing industry, Thompson also studied

---

<sup>167</sup> A similar form for presentation was employed by Tiplic (2008).

knowledge based organisations like universities. This should not suggest that task environments are easy to define, for example Thompson recognises that they will vary from organisation to organisation (Thompson, 2003: 28). As Thompson also recognises there is also the necessity for “domain consensus”, which amounts to the expectation set of “members of an organisation (in this case within the subunits under study) and for those with whom they interact, about what the organization will and will not do” which will guide decision making (Thompson, 2003: 29). But as Thompson further notes, one must look beyond formal statements to discover “choices of action alternatives” as well as beyond “individual goals and motives” when ascertaining this consensus. This study has not been about matching the formal statements of policy to actual behaviour, but rather an attempt at investigating from the subunit level how the process of designing evaluations is thought to occur and what influences it. As we have already seen and will explore in subsequent sections, this is thought to partly be moderated by the level of cooperation, operational proximity and degree of agreement over task and how to achieve it within the subunit. In dealing with the varied “constraints and contingencies”, an organisation under rational conditions should attempt to manage its dependencies (Thompson, 2003: 30). While Thompson offers varying propositions for how organisations might attempt to do this, this chapter deals rather specifically with the perceptions of these pressures. These points are further revisited in the subsequent chapters related to designs chosen and processes of decision making.

Returning to the main focus of this chapter and drawing on the responses given, three broad categories of groups within the task environment were considered to place demands and pressures for evaluation upon the subunit: policy makers and associated agencies, commissioning bodies and programme participants / students. While there were some more isolated reflections concerning the impact of competitors, most of the reflections concerning the task environment were focused upon these customers/clients, policy makers and regulatory groups. In this study the concept of commissioning bodies applies more particularly to responses from the Norwegian context, referring to the reflections over local and regional authorities who commission programmes for specific groups of school leaders.



Demands and pressures		Underlying issues
Perceived pressure from task environment	Policy makers and agencies	Systematisation
		Regulation
		Information
		Impact focus
	Commissioner Pressure	Programme availability
		Variability
		Commissioner competence
		Memory
	Participant Expectations and Demands	Implementation
		Prior experience
Voluntary/involuntary		

### 7.1.1 Perception of pressures and demands from policy makers and agencies

In this subsection consideration is made of the pressures and demands from policy makers and associated agencies upon the different subunits under study. As has already been stated, the purpose of this and subsequent sections is not to demonstrate generality, but rather consider how different perceptions of external demands might be thought to influence the decision process surrounding the implementation of programmes. At the same time, as there are similarities concerning the evaluation and quality assurance frameworks the different organisations must adhere to, there is considered to be likelihood that experiences will be analogous.

As was recognised in the introductory section of this chapter, and as was outlined in chapters 4 and 2, there are general demands with regard to evaluation and assessment within HEIs in addition to any subject specific requirements that might be in place.

#### *Pressures for systematisation, demands for impact...*

There was a perception across the subunits of increased pressure for evaluative systematisation and an increase in results / impact focus, although in England these pressures were considered to have developed more strongly. It was noted by a respondent in NOR 1 that a system change had taken place in Norway with regard to higher education policy, going from control by rule and of content within a process where focus was on the individual student and their final results, to what amounted to a redefinition of competence, which differed greatly from the traditional definition of quality in the HEI setting. The changing system and new demands were perceived to have led to new concepts of competence, particularly within the field of educational leadership. The shifting nature of demands was considered to place pressure on organisations to shift

their conceptual understanding of quality and competence, and how they might go about ascertaining it. The respondent noted that:

*If you think that previously we only had rule-based management and control of the content of study plans and curricula. ...There was a concept of results but it was individually based; that which you achieved in an exam and not as the basis for an evaluation of the programme. This is... process steering. What happens now is that definitions are created of what is good competence through a questionnaire, thinking that it will have a retroactive effect on the programme. The content therefore becomes the concept of competence, which is different to that which we have traditionally used, or related to within the university system. As a result the academic requirements are developed with new concepts of competence. I think that educational leadership studies are partly characterised by this. [NOR1i]*

In addition the respondent considered that academic institutions were under greater pressure as a result of reform focused upon attempting to break down the barriers between academic demands and practical needs which were considered to have traditionally separated HEIs from other educational and training institutions. This division visible through the twentieth century had seen HEIs developing and “encasing” their own theories and methods of developing knowledge through an “academic approach” compared with the “practical” approach often exhibited by schools. The respondent considered the fact that this approach was now disintegrating to be important, especially with relation to the assessment of programmes and activities within the so-called “knowledge society”. This had led to a new demand, focused upon greater convergence of profession, research activity and policy leading to systematisation:

*There is a much greater demand that schooling should be research based and profession oriented such that researchers should be out in school promoting change, without going via all of these levels, and the government should use research for so-called “evidence based” policy. In this way researchers enter into public administration to a much greater degree and the educational system should use research methods to evaluate themselves or be evaluated by others. So what you see is dissolution, a disintegration of the typical boundaries between that which is academia, that which is practical enterprise, that which is government, and professional decision domains. Institutional boundaries disintegrate and there is a mingling of tasks and who does what, which is a consequence of research being partly politicised or that politics become depoliticised. Those that succeed are those who have the “knowledge” and know how to filter it, where assessment becomes very important. [NOR1i]*

It was further recognised that changes came as a result of supra-national demands associated with the Bologna process, as outlined in chapter 4. A rich web of demands was drawn, that were both interlocking and at times contradictory. The respondent went on to reflect over this:

*It's clear that there are so many expectations and demands, and not just as result of user surveys, but how one adjusts a programme along the way is dependent upon a number of factors; some of them developing from the Bologna process as well as a number of change processes demanded centrally within [the organisation] to match a particular form.*

Similar reflections were evident in each of the subunits under study, with the recognition that additionally the Bologna process had affected policy process and created a framework for programme development that challenged previous methods of working.

Despite these reflections over a broad framework and development of an accreditation and assessment regime there was variance in terms of how this regulatory framework was implemented or experienced at the subunit level. In the subunit at NOR1 there was not felt to be any specific demands for evaluation of programmes coming from the Governmental level, one respondent giving the example from a project proposal:

*So with regard to responding to that type of demand arising externally, I haven't really experienced that we get so many demands from the Directorate<sup>168</sup> or the Ministry<sup>169</sup>. Well, in the [not named] project terms of reference, some years ago, there were some targets, but these were expressed so broadly that it was something you would have planned to do anyway...[NOR1f].*

There were elements within the other respondents' reflections noting that there was thought to be a fundamental change of approach to the focus and assessment of HEIs. But, while the changing pressures resulting from policy reform appeared to redefine the focus of HEI based conceptions of competence, to some extent the impact of these changes had been slower than the respondents first anticipated. When it did come, there was a suggestion that it did not call for a great deal of change, particularly at the subunit level. For example, members of the subunit at NOR1 already felt that their frameworks more than met any new demands for evaluation that came as a result of the HEI Reform, and the demands did not appear at once anyway:

*they actually came a bit later, because the Quality Reform and pressure from that came a bit later again. So I remember that we discussed in our group that we actually had already established these procedures and routines. So for us it wasn't just the quality assurance but also most of the other parts in relation to supervision and follow up of students were in place. It wasn't really much of a change process for us. [NOR1f].*

---

<sup>168</sup> Utdanningsdirektoratet – The Norwegian Directorate for Education

<sup>169</sup> Utdanningsdepartementet – The Ministry of Education

This point is also relevant for and will be discussed further in chapter 9 in relation to the decision making processes concerning the evaluation models to be implemented on the programme. Another respondent from NOR1 agreed with this perception that the new focus upon quality assurance from the national accreditation body was little different from the evaluation the subunit already undertook. The purpose was felt rather to be a control mechanism, implemented to confirm that the processes were undertaken rather than so much interest in their details and remits. The respondent did not experience much conflict with subunit aims:

*No, it's pretty much in line with our own agenda. But it is such now that it isn't first and foremost what we do; it is actually more about that we undertake evaluations that is in focus... [NOR1d].*

Another respondent agreed with this point, recognising the basic difference to be that of disclosure; the importance of evaluation information being made available in the public domain. When asked whether there was any feedback related to these demands a respondent reflected that previously assessment data from evaluations was only available to programme participants. This was however changing:

*But now in a way it has been made more visible with NOKUT coming into the picture. So there has come a demand for us to make more visible what we have been doing, an assessment of it, and furthermore there has been a demand that it needs to be accessible to all on the internet. So there is another demand that has started to come. [NOR1a].*

Interestingly then, the pressures upon the subunit appeared to be more about improving information flow rather than any increased demands or changes in the practice of evaluation. While members accepted that in some way demands appeared to have increased, these had little impact at the programme level and again didn't impinge on what was already being done:

*I think it's pretty much that what we do, we would have done anyway even if nobody had said anything, particularly at programme level, and we don't think the other demands are particularly taxing... for me it's just about passing on information, so I don't really feel that there are any demands at all in relation to evaluation... [NOR1j].*

This "passing on of information" was perceived by another respondent to be rather a part of a general policy demand felt across the education system and wider, using the tools of control as a way to achieve improved output.

There were recognised to be both positive and negative influences upon the subunit at NOR1 in relation to the design and implementation of the evaluation

of their programmes. While these factors will be mainly dealt with in the subsequent chapters, there were interesting views of the wider policy demands. Another respondent concurred that demands and pressures had had little impact upon the general activity of the subunit, where there was much greater focus upon the micro-process of programme delivery than there was on meeting wider policy requirements. Any system that was developed should be internally beneficial and improvement focused:

Despite the disagreements with the publication of results and standardisation of focus, another respondent noted that one of the spinoffs of this policy did appear to have contributed to an improvement in the quality of data produced and available for use within the subunit:

*But what you can say has happened over the last few years within schools, universities, colleges and the rest, has been that the demands for more formal assessment, at least clearer in written form, has moved us from an educated guess to knowing. That's the positive side. [NOR1e].*

While the responses presented so far have all been related to the subunit at NOR1, there was also a notable similarity within NOR2, an especially referenced in relation to the external accreditation bodies. It was recognised that the formal demands were met in relation to both constituency of the programme as well as in terms of the quality assurance processes, and the subunit attempted to meet all of the demands placed upon them externally. Despite this an additional point of interest was raised concerning the quality assurance systems that had to be in place, which were considered to be challenging. One respondent noted that the field of educational leadership provision had become more of a market, balancing the demands had become difficult as selling the programmes to different commissioners in addition to running a more traditional school based postgraduate programme meant more complex issues in terms of who the participants were. In this case the perceived control logic of the formal demands of programmes appeared to inhibit the development of the market that the reform had put into place, but this also meant changes from the traditional practice of the wider organisation. These developments also challenged the routines and practices of the QA system within the organisation:

*We are more creative because we need to relate to a demanding group of users and we are going into new markets that the organisation hasn't gone into before: and that is also creative. We use more resources in operation, for better or worse, and that impacts our profitability and breaks with the logic of the quality assurance systems.*

*How?*

*We break with the logic of the quality assurance system by the virtue of having too little time; we can't sit and wait for codes and all that nonsense. [NOR2e].*

The respondent considered that by using HEIs as programme providers an awkward system had been introduced. Alongside the tightening of quality systems there was the intention of postgraduate education for current and future leaders. Demand for quick response did not sit easily with demand for control in higher education. The demands experienced by the subunits in Norway from policy makers, appeared to be more focused upon ensuring the quality assurance frameworks were adhered to. There was not perceived to be any particular demand from the Government concerning the output of the programmes, summed up by the comments of a respondent from NOR1:

*Up to now we haven't had that type of demand, but then we haven't had demands for school leadership preparation either. Maybe we need to start somewhere; internationally there are demands... [NOR1i].*

The respondent reflected particularly over perceptions of the system. As will be seen further below, evaluation of the programme output appeared left to commissioning bodies in Norway.

### ***Linking programmes, outcomes and funding***

In England which was considered to be far more directive there was similar recognition that the accreditation systems had increased the codified demands upon HEIs, but furthermore that processes were already felt to satisfactorily be in place. But an interesting set of reflections concerned changes in the source of pressures and demands. In ENG1 one respondent remarked that there was now a much clearer link to funding. The funding bodies appeared much more to be fronting the demands for quality control, even though there was not perceived to be such great change to the act of evaluation at institutional level. The respondent considered that current QA processes were already rigorous enough, at least in terms of meeting the demands that had been set and that the increasing demands did not take into account that most HEIs were performing well and evaluating effectively:

*So, again, that's another form of quality control, quality assurance, but again it's kind of rules that are laid down by the HEFCE, Higher Education Funding Council for England. So there are some things as Universities that we have to do following those National Guidelines, but I think that most universities... look at issues of quality in a very serious way. [ENG1g].*

Similar comments were also made by respondents at the subunit at ENG2, where it was additionally recognised that the nature of focus upon evaluation had shifted from programme control to fiscal accountability. One respondent was asked to explain how these changes had been felt:

*Well, there are the pressures that come from the Quality Assurance Agency and from the Higher Education Funding Council, [with the expectation that they will] be able to see very careful analysis and evaluation of practice and quality. I don't think that the QAA is doing the kind of big, large scale evaluations of Institutions that they used to do, and we're not liable for Ofsted for this particular programme... so we're not under that kind of inspection. So the pressures I think are there from the funding agencies, [ENG2m].*

There had, then, previously appeared to be more focus on the subject quality of programmes, and while respondents recognised that it could still potentially come they also considered that the major focus now appeared more related to funding and attempting to satisfy these demands. One respondent compared how discussion from a previous evaluation of the programme had led to much more internal evaluation of programme content than was currently under focus:

*as result of [that evaluation] the two issues that came out of it were, the lack of progression, well the lack of ability for students to really demonstrate progression, and the lack of any extended writing, and one of the things that [became a concern] was how to actually bring those things in, [ENG1p].*

But another part of this focus in the UK, from outside of the institution, was also directed towards the increasing demands for impact evidence with regard to programme achievements and output. It was, as noted in chapter 2, tied heavily to the concept of “value for money” outlined above. A respondent from ENG1 was asked whether these growing requirements created any difficulties. The respondent felt that this shifting focus at national policy level towards financial accountability based on assessment of evidence for impact was becoming almost the sole basis upon which programmes were commissioned, but this was perceived also to be in response to the previous lack of programme evaluation and assessment:

*Well I think that a lot of Government initiatives, education initiatives, have been driven by financial factors, and it is often said that the Treasury has a leading voice now, and basically the Treasury wants to know what the payoff is; is it worth us investing so many million pounds in this particular initiative when we don't really know[what it leads to]? And things like professional development and leadership development in a sense are kind of a black hole where lots of money is poured in, but we don't actually know what the effects of it all are. So government agencies are asking us to be much more clued up really. And they are also asking schools and others who are buying into professional development to be a little more aware of value for money and impact evaluation; ... what difference it makes, if any. And seeing professional development not as a day-off or an excuse to have a good lunch in a hotel or whatever, rather what difference is it making to the quality of teaching and learning. [ENG1g].*

The impact of these demands at the provider level as felt to be increasing. The respondent recognised that this, in turn, had led to a tension of evaluatory demands for the subunit as a programme provider:

*Well I think the tension comes because, ... there is an expectation on the part of whatever we do that we can demonstrate that we're having some kind of impact, whereas a decade ago I think that wouldn't have been the case either in schools or universities. Now it is incumbent upon us to demonstrate that, and that has led to some kind of creative thinking about ways in which you can do that but it comes back to what I would argue is a very kind of naïve and unsophisticated view about any particular programme or activity... and Government Agencies, Inspection, you know Ofsted and others are all sort of banging on the door saying that we need more of that, you need to get better at doing that. [ENG1g].*

Respondents from the subunit at ENG2 had perceived in a similar way the increasing demands to demonstrate impact. But one respondent recognised that while the initial focus had been on collecting and providing evidence of impact on pupil outcomes, there was a growing recognition of the difficulties of ascertaining impact upon pupil outcomes:

*I think particularly earlier on the external pressures were perhaps a bit, what's the word, perhaps asking for rather simplistic lines of causality between professional development activities and direct effects on pupil performance, and I think that has come from our Department for Education and Skills as was, and other Government bodies, but I think more recently there has been an appreciation, both externally and internally, of the complexities and subtleties of tracing the effects of [continuous professional development] on practice and on pupil's thinking and their organisation. [ENG2k].*

But whether or not certain demands had eased or not, it appeared clear from responses that attention had to be paid to the whole area of impact. It had, in most cases, become a prerequisite to gain external funding, which was becoming increasingly more necessary to ensure programme survival. Another respondent from ENG2 outlined how preeminent discussion about evaluation for impact and how embedded it had become in activity had become:

*I'm putting forward an alternative proposal for trying to move forward with this award. And one of the reasons for doing that is to do inevitably with funding and the need to grab hold of more funding from a body in the UK called the TDA, across all our courses, and your point about evaluation is very pertinent because in fact the first meeting that I went to this morning was about an evaluation report we have to complete on 3 of the courses which have this particular sort of funding. Alongside getting funding goes the filling in of an evaluation report...*



But as the respondent further noted, the competing demands meant balancing the focus of evaluation was becoming a much more complex activity, with differing agendas becoming even more polarised. While the external demands focused upon on programme impact, the subunit members were also interested in engagement with programme and willingness to return. The respondent noted that these competing demands concurrently:

*will impact on the evaluation, won't it, because if the TDA's evaluation criteria are 'will schools be more successful because the teachers have undergone effective professional learning experiences', ours will be the bottom line- will those students come back to us? [ENG2n].*

It was also interesting to note that on the rare occasions where programmes did not primarily rely on external funding, the perception of external demands decreased. One respondent at ENG1 outlined how one of the programmes had adopted a distance learning hybrid form and drew participants from across the world. In addition these participants were generally self-funded. The respondent outlined the freedom that brought to the programme providers from external frameworks, and further noted that:

*I think because our particular course, our Masters, is so global in the type of people it covers etc., [that] I don't think that we've had to take into account anything else in particular when we we've been doing any of that [evaluation]... and there is a freedom in that, that you don't get in some other programmes. [ENG1a].*

Here was some glimpse of a similar situation to that described by the Norwegian subunit members. There was an interesting difference at the time data collection whereby the Norwegian subunit members mainly talked about policy pressures in terms of quality assurance, mainly higher education policy, whereas those in England placed their focus more notably on financial accountability and issues of impact.

### ***An extra level, the NCSL***

In England there was also much focus on the National College for School Leadership (NCSL). While this study does not focus on the organisation and provision of the NCSL, there are points of interest that were raised by subunit members in England concerning the impact of the NCSL on their own programmes. Part of the increased pressures and demands upon HEIs in England were perceived to have stemmed from the role played by the NCSL. The role of the NCSL was outlined briefly in Chapter 2. Respondents referred to the NCSL in more general terms and its perceived influence on their programmes, noting the obvious effect on the market and numbers of participants that were on their programmes particularly seeing reductions in England and the other home nations in UK, but they also reflected how there had been an influence on the

evaluation of programmes. These comments will also be returned to and discussed further in the chapters concerning design and decision making.

There were three major themes discernable in the discussions with respondents in England with regard to the impact of the NCSL and subsequent demands and pressures they experienced with regard to evaluation. Firstly, as has already been outlined, respondents discussed the increasingly greater focus upon evaluating for programme impact and this was also evident in relation to the remit many had had while working as external evaluators on NCSL programmes. Secondly, the role of the NCSL had influenced change in the debate about evaluation of school leadership development. Thirdly, some respondents spoke of the conflict over the commensurability of the NPQH and Master programmes, grounded in an “evaluation” of programme quality. I turn now to these major points drawing on data collected at the NCSL as well as from within the subunits.

### ***Impact and evaluation focus***

Chapter 2 considered the role and mandate of the NCSL within a policy framework that had embraced NPM and modernisation agendas. Within this framework the importance of assessment and evaluation was highlighted, with increasing demands for impact studies and “evidence” of what worked, and often described as “best practice”. As was outlined in chapter 2 various projects had been commissioned to explore this area of impact appraisal further, particularly pertinent to this study in relation to the impact of school leadership training and development upon pupil and student outcomes. As was recognised above, evidence of programme efficacy was increasingly a stipulation for funding. One respondent from the evaluation unit at the NCSL (ENG3) confirmed the link between funding and impact, observing the existence of the College as evidence itself of the policy approach:

*Now also one of the aspirations for the focus on leadership, and it is really one of the major justifications for the government investment in this, is it adding any improvement? [ENG3p].*

As time had developed, this concept had also become progressively more and more visible. The framework for College’s programmes had developed in such a way that the concept of value for money had been introduced as a result of governmental demands. This concept was added to the frameworks for external evaluation, often undertaken by members of HEIs. The respondent was aware that the concept still needed to be better understood and developed, even though it was currently being applied to NCSL programmes:

*In many of our evaluation projects we ask evaluators to report to us against those criteria, so they’re always embedded in the tender document, responding to those processes. We’ve also added another one recently which is about value*

*for money, to what would you say as an evaluation team that this particular programme is offering value for money and we're exploring different ways in which that might be interpreted actually at this moment. [ENG3p].*

As was noted above, many of the evaluators working for the NCSL are members of HEIs providing postgraduate programmes in school leadership, and respondents in this study were or had been amongst that number. The demands and criteria for the NCSL evaluations were perceptibly different than those related to the postgraduate programmes. One respondent at ENG2 recognised that the criteria set out by NCSL were more formalised and explicit than might be used within the evaluation of their HEI provision, but at the same time considerations of impact were in some way comparable to those used on the programme:

*It's a bit different, in that the evaluation study that I did for the National College had quite specific criteria that they asked us to look at, so we used those criteria. But those criteria would not be dissimilar to ones that we would use for evaluating our own provision. There were similar sorts of issues, in terms of short, medium and long term effects and so on. [ENG2k].*

But there was noted to be development in the way impact was being looked at across NCSL programmes. A respondent within the subunit at ENG1 considered that the NCSL had improved their tender documents for external evaluation, particularly with regard to ascertaining impact:

*But they themselves have become increasingly sophisticated in their tender specifications. So, for example, the National College has done some quite interesting work looking at impact and seeing ways in which impact evaluation can be at a number of different levels but in a slightly different approach to say Guskey or some of the other writers. [ENG1g].*

But although the NCSL was seen to be consistently evaluating its own programmes there was an expressed concern over the focus on evaluation that was being taken. One respondent from ENG2 exemplified this issue by outlining a discussion that took place concerning a proposal that had been submitted to the NCSL for an evaluation of one of its programmes, which had been rejected. The respondent noted that the proposal included the intention to look for exemplars of successful practice of the programme under investigation and compare them with unsuccessful practice:

*...they said no, we don't want any of that we just want to know what works, we want to explore the reasons why things work. And the argument says well yes but you often only find out why things work under these circumstances by looking at things that aren't working in order to find out if the circumstances make any commonalities between them. And they are not interested in that. [ENG2m].*

The respondent used this as an example of the perception of intention for evaluation at the NCSL, mirroring the evidence informed approaches that were outlined and discussed in Chapter 2.

But there was also additional concern over the quality of evaluations undertaken by NCSL programme participants. In the Improving School Leadership country report for England prepared for the OECD (Higham et al., 2007), there was suggestion that less is known about what happens from the University programmes from evaluation compared with National College programmes. This discussion was raised with one respondent from ENG1, who while agreeing there was a certain amount of truth in this statement, noted that the report also challenged the focus and rigour of many of the evaluations implemented by NCSL. The respondent did however also suggest that not all the criticism the NCSL had received was well founded:

*...I'm not all that convinced that the evaluations the National College conducts of their own programmes are credible anyway; [asking questions like] did you have a nice time on your programme, if so could you tick this box. So I am not all that convinced that the standardized evaluations and conventional assessments tell you a great deal. Having said that, there's been a lot of criticism about the NPQH in the country but also a lot of positive comments on some of the College programmes too, like LPSH and so on. [ENG1b].*

As such issues concerning the quality, rather than just the frequency of evaluations came into question, in addition to questions about the extent of impact that one could ascertain from current models implemented across different programmes. These reflections are interesting as they reflect the conundrum for subunit members, having to interpret perceptibly similar demands and expectations for evaluation in different ways across different arenas. This had led to an unease amongst field members, which, as might be ascertained in the next sub-section concerning agendas and debates, they felt was difficult to respond to.

While this subsection deals primarily with the role of the NCSL in England, members of the subunits in Norway were observed to be knowledgeable about these developments. In the next chapter I will outline respondents' discussions concerning the within unit discussions and reactions to the concept of impact in relation to evaluation, but in this section I briefly consider if there is perceived to be any wider demands in relation to ascertaining the impact of the HEI programmes.

Focus upon impact was not currently considered to be evident in Norway at the national policy level or through related non-departmental bodies. Respondents across NOR1 shared the view that there was little external demand on the wider political level with regard to effect of programmes, but that changes were

mooted with regard to discussion concerning mandatory requirements for school leaders:

*No, there isn't pressure for that here, but there could be, especially if becomes like we think it might, that you'll have to have some leadership qualification and possibly 30 study points to be able to apply for a leadership position. The Conservative Party<sup>170</sup> has suggested this and it's on the cards that we'll move closer to other countries that have demands for leadership training ... I would imagine that in connection with that, if it happens, there will be closer follow up and greater demands about what the results of such an education should be. I think it would be great to have increased demands as then we could also negotiate with someone about what they want and why they want it and what we could offer and what is valuable etc. So, I think it's also legitimate to be able to ask questions, but I don't see that there are any questions about it now. [NOR1h].*

The respondent reflected further that this lack of demands and control was symptomatic of the general lack of demands across the education field, and especially within higher education:

*And I think that there is weak control, there are not any great demands placed on us to achieve anything. There is so much money used on Higher Education in Norway; should we just be allowed to do exactly as we like? I think it's fine if someone makes some demands of us and expects something. [NOR1h].*

In NOR2 there was also reflection over the possibility for greater demands for school leadership training but it was noted that an idea similar to the NCSL was unlikely. This in turn meant that it was unlikely that national standards would be adopted and thereby the basis of demands, exemplified by this response:

*They are probably searching for a form of minimum demands as to what a school leadership programme should consist of, possibly not a national curriculum but a conception of some kind or other of what it should look like in order to justify the label school leadership training. There isn't any feeling in Norway to create a national college; that would be in opposition to Norwegian thinking. [NOR2e].*

### ***Debates and agendas***

Another area of tension concerns the development of the debate concerning the field of evaluation. In many respects the tension in the English setting seemed to centre more upon a perceived division between the NCSL and the HEI community. Respondents from the HEI subunits and the NCSL referred to this issue, where the subject of evaluation was part of the on-going debate. While it

---

<sup>170</sup> Høyre, in Norwegian

was recognised that the Universities Partnership Group had been formed in order to facilitate discussion between leading HEIs and NCSL, there was also suggestion of disproportion in the relationship, with the NCSL perceived to be sitting with the resources and mandate over the field. In discussing the development of evaluation processes at the NCSL a respondent outlined the unease that had developed with some HEI groups, while considering that the purposes for evaluation at the College had generally been made clear:

*So for the most part much of our evaluation is formative, so it is about improvement of own practice, whereas you see, [the Universities], feel that we ought to be more engaged in the evaluation community contributing to the improvement of the evaluation process. Now whilst we want to improve our evaluation processes, our major purpose is to generate evidence about the effectiveness of our programmes to improve our practice. [ENG3p].*

The respondent went on to discuss how this debate had focused upon the way these approaches had been perceived by members of the academic community who had undertaken evaluation for the NCSL. In answer to the questions raised, the respondent reiterated that the NCSL had a different “agenda” when it came to evaluation of programmes, and was focused on internally improving practice compared to the more academic reflection over purposes and process that might be associated with HEIs.

While the NCSL was considered to be most focused upon its own development and delivery, respondents from the HEI subunits also considered that this approach within the framework that was set had also had an impact upon their programmes and delivery. One respondent considered that the control over resources the NCSL had, alongside the role it had been given, appeared to shape the wider understanding of the role and focus of school leadership. More specifically the direction in which NCSL had developed had certainly influenced the focus of HEI programmes, particularly a shift from training to development, and also the way they are evaluated. The respondent noted that:

*I think that the driver in terms of our understanding of headship and the way in which we view it, headteacher preparation, is very much the National College. It was created for a particular purpose, it has delivered on that purpose and one or two people who have been significant players within it, at one time or another, are back in academe and are developing ideas that are popular within the College...*

*And the fact that they've got that drive and the resources to do it, I think that is pushing us towards a particular understanding of how you evaluate practice and how you evaluate training, and it is training now, it's not development. It's not leadership development, it's leadership training, or [rather] leadership preparation is leadership training. [ENG2m].*

Another respondent from ENG2 agreed with this perception of NCSL impact on the programme being run by the subunit:

*well, [the NCSL] has certainly impacted on provision in that, not just the National College but [also] the advent of National Standards and the framework of expectations for Headteachers and other teachers within the school... has had an impact on provision, and I suppose, yes, on evaluation. [ENG2k].*

There was a sense, then, in which the debate about evaluation of school leadership development was very much being driven by the NCSL, even if they were from their own perspective not considered to be particularly involved in the debate.

### ***Degrees and Qualifications***

The third area of reaction from the HEIs to the development of the NCSL concerns an issue which might be considered more indirectly linked to evaluation, namely accreditation for study on NCSL programmes on master awards. This relates to the commensurability of the NPQH and master level degrees. Part of this debate centres on the issue of how many study points or credits can be awarded to programme participants embarking on a master degree who are in possession of an NPQH. This debate will be reflected further in Chapter 9 when discussing decision making.

From the perspective of the NCSL it was recognised by one respondent that while their programmes are different to master degrees and that the NCSL is not a University, programme participants should be made aware of “what masterly level work means”. The respondent further noted however that there continued to be an issue in getting NPQH and other programme accreditation, even though, for example, the Universities Partnership Group had in discussion with the College “suggested that the NPQH should be 60 Masters level credits, Leading from the Middle should be 30 masters level credits”. In addition the Universities Council for Educational Training of Teachers (UCET), had also raised the question about accreditation and equivalence, and “[suggested] that we establish a protocol, a memorandum of understanding that UCET members understand”. The respondent went on to say that:

*We've had this negotiation and we've agreed that all our programmes are worth this but it is still the case, regardless of whether there's a written agreement, that each university will interpret it in their own way. So, you know, [a university] might be a signatory to it, but then say, well, that's all very well we'll be delighted to recognise 60 m level credits but, you've done NPQH, but what we'd like you to do now is just write a 3,000 – 4,000 word essay which synthesises your learning from the process, to provide us with evidence of your capability of writing at masters level and your evidence of engagement with the research and literature about NPQH. And do you know when I was at [NN –*

*HEI] that's the kind of thing I would have done, and people would come to me with I've got this and this and this, will you accredit the prior learning and I would say, I'd be delighted to but before I give you exemption from X, Y and Z, I want you to do this exercise. And that did 2 things. One, it said that how serious are you about wanting to go on to masters level work and secondly, how capable are you of writing at masters level? Because if it transpired that I would be slightly nervous about your capabilities, I might say, well I won't give you exemption but I think that you need to start the programme from square one. [ENG3p].*

These reflections were recognised in the responses given by the various members of the subunits in England. As one academic at ENG1 noted:

*I am quite a strong believer in university standards and I think I would struggle to see the NPQH worth 60 credits in a master programme. If they had to write a sort of reflective essay on what they had got out of the NPQH and they used the available literature in a constructive way, that would be fine, but just to give 60 credits on the nod I think would be a bit of a problem. [ENG1b].*

This view was also shared by respondents at ENG2. A further issue was also raised concerning the level of work that was undertaken as part of the NPQH:

*I think, that although [NCSL] support research and although they encourage headteachers to undertake work as research associates, these are small scale, very limited activities and the projects and the courses that they run aren't in my view master degree standard. ...*

The respondent went on to outline how a masters module needed to be differentiable from programmes like the NPQH, where constructs and concepts like the “National Standards for Headteachers” should be questioned theoretically and empirically rather than being “taken as read”. The respondent therefore considered there to be something essentially different about the evaluation of programmes and subsequent attempts to discern equivalence:

*So there are valuations there that are different, and I think those have a bearing upon the way that the NCSL has addressed the business of headteacher development, and the way in which traditional masters degrees do so, and it's interesting that they've looked continuously and consistently for getting master degree, m level points as their called, for their NPQH and Leading from the Middle courses, and very few, well they've given them but they've very rarely given them for more than one sixth, more than 30 points at m level. [ENG2m].*

These responses appear to show a degree of evaluative “gatekeeping” with regard to the pressures and demands that have arisen from the task environment, where subunit members have used the academic frameworks in place to defend their standpoint. In England the developing requirement that NPQH be



mandatory for new school leaders has drawn many potential participants away from postgraduate programmes within HEIs. At the same time there has been much discussion as to the standard and level of the NPQH and other programmes developed by NCSL. My intention is not to imply that this accounts for all the reasons for these reactions from HEIs with regard to accreditation, but rather to show how the process is an evaluative one, based on a different value structure and professional perception. This point is dealt with further in chapters 8 and 9 when considering academic anchoring.

In Norway the respondents outlined how the pressures and demands for evaluation from national policy level had been most felt in terms increasing demands for the QA systems, but that these demands were experienced as controlling that the activity was done rather than the content of the evaluation. In England there was generally perceived to be a greater central demand for subunits to evaluate their own programmes in terms of the impact they were thought to have on the school development. It was noted that funding bodies had taken a more central role and that support for programmes often relied on a plan for evaluating subsequent impact.

This does not imply that issues of funding and impact are not important demands in Norway, but rather that they are not fully felt from the policy making level. Due to the national policy developments in Norway another group of stakeholders is important when considering demands on programme providers for the evaluation of their programmes, the local and regional authorities operating in a capacity as commissioning bodies.

### **7.1.2 Perceived commissioner pressure**

One impact of educational policy reform in Norway was that finance had been made available to local and regional authorities to fund programmes in school leadership training and development. As was related at the end of the previous section, no national standards were set and as was also outlined in Chapter 2 the criteria for decision making were to all intents and purposes left to the individual authority, as employer, to decide what was best for the development of their employees. The interaction between programme providers and commissioners will be dealt with further in subsequent chapters, but one of the pressures that arose immediately for providers was the number of different commissioners simultaneously tendering for programmes. This created a challenge for programme availability:

*The external demands [came] more from the “Knowledge Promotion”<sup>171</sup> reform, and there came a rush for our external further education programmes, as local authorities and networks of local authorities, and regional authorities all wanted school leadership programmes and I think most of the time we had a list*

---

<sup>171</sup> Kunnskapsløftet, in Norwegian

*of about 15 to 20 that wanted some kind of programme from us and we didn't have the capacity at all. We had a great deal of students on our external and internal programmes and there weren't many of us.*

With a sudden surge in numbers requiring programmes, subunits were faced with logistical pressures. But more than that it was also an issue for evaluation as respondents outlined that it was not always clear what the intentions of the commissioners were, what they expected as outcomes and how they thought this could be demonstrated. As the respondent from NOR1 noted the external demands that had come as a result of the reform had created more complexity at this level and became the essence of the demands experienced:

*With the arrival of the Knowledge Promotion everyone wanted an educational programme, and yes they all wanted it to be adapted in a way to their own needs. So really it has been demands from the field of practice that have been the external demands, at least as I've experienced them. [NOR1f].*

This reflection also challenges any issues of generality concerning this data. As there is a great deal of variation between the different commissioning bodies, and with only two institutions in Norway studied, I do not attempt to generalise concerning the nature of demands. The purpose throughout this study is rather to investigate how commissioning bodies demands, roles and involvement might be seen to influence the decision making process concerning evaluation and becomes a factor to be addressed. In this section focus upon the commissioner role is discussed in relation to the perceived demands, while other matters arising are followed up in subsequent chapters.

### **One reform, varied responses**

As is seen in the previous section, discussion concerning the impact of the "knowledge promotion" reform in Norway thus raised a further issue that the programme providers had to deal with, that of the variation in demands in relation to complexity of what was being required. Examples are given here both of commissioners perceived to have clear and unclear demands for the programmes. Those thought to be clear in their demands often additionally have a clear perception of the evaluation they expect from the programme. As will be seen though, this might not always be in concert with the subunit responding to the tender.

Examples of the great variability in what commissioners wanted from the programmes were noted in both subunits at NOR1 and NOR2. One respondent from NOR2 recognised that not only could commissioners be unclear in their demands but that they also were unclear what implications choosing study at master level had:

*It is often the case that a commissioner doesn't know or understand what they really want. And so, after winning a tender, we enter into a dialogue, a*

*discussion with them: “now listen, this is after all a programme at master level, you need to raise the bar higher than this”.* [NOR2e].

Another respondent at NOR1 remarked how despite having limited experience with different local authorities this issue of commissioner competence appeared as an important topic:

*My experiences are quite few really, but you can see that the concept of “commissioner competence”<sup>172</sup> recognises great variation in relation to the different groups, such that some are very clear in their demands and clear about what they want, whilst others are searching more. So it needs a process to ascertain what can be done in relation to where they are.* [NOR1d].

This issue of commissioner competence and its impact upon the decision making processes is dealt with further in Chapter 9, along subunit members’ perceptions of how they responded to these external demands. Despite this variation there were examples in both NOR1 and NOR2 of working with commissioners with very clear demands. In NOR2 groups who had chosen to accept the subunits bids to tender were considered to be sure of their intentions for the programme they were commissioning, one respondent commenting that:

*I can see that our customers are quite explicit about what they want and what they don’t want.* [NOR2e].

In addition, in this case there was a major focus from the commissioner that the programme would improve the quality of school leadership within the local authority and improve the chances of recruitment and employability, undergirded by an expectation of evaluable results:

*The demands that I hear most are those connected to effectiveness, focus on learning outcomes, and creating better schools. Other demands that have registered from those with expectations are that we should make the participants better leaders, such that they can lead more operatively, developing tools and methods for this.* [NOR2g].

But this was connected to a more widely experienced predicament of succession planning, noted as a problematic issue in both Norway and England, related to the number of school leaders that will retire from post during the next five to ten years. One respondent from NOR2 had recognised this need in the demands of one of the commissioning bodies:

*It’s partly because commissioners see the enormous need in the future for leaders, not that they are thinking that this is a way of creating teachers who are prepared leadership that can be recruited at some point later.* [NOR2h].

---

<sup>172</sup> Bestillerkompetanse, in Norwegian

In NOR1 an example was given by a number of respondents of a regional authority who had clear demands based on the perception of development that was required in their area, and the desire to focus on practice change, with the expectation to see what response the process would bring. This raised an issue in terms of what kind of evaluation could be done in relation to the programme, the respondent noting that:

*...whilst [NN] are distinctive from the other programmes that we run in that they linked it up to obligatory development work over a 2 year period, because they wanted to focus on the long-term. This creates a completely different level of reflection around the programme. [NOR1h].*

One of the respondents who had worked most closely with the project noted the specificity of the demands and expectation of definable results:

*Yes, they had very clear demands. They wanted their school leaders to complete a programme that was equivalent to the first year of our master programme. At the same time they had the idea to select the leadership team from their schools, so they were many from each school, and they wanted to use this process to commit the schools to change. So as commissioners that demanded to see change, concrete changes. [NOR1e].*

Although there were notable differences from place to place, in focusing more on developmental process work there were great challenges in evaluating and discovering results of the programmes' impact. Some commissioners, however, were aware of these limitations and had set their focus on trying to ascertain longitudinal change:

*There has been a great deal of difference. You can see that in some situations there have commitments to development work in the work place... and they are most focused upon that the change processes are long term, where it is not likely to see wonderful results after a year. [NOR1j].*

And this was a major difference in impact on programme survival, because while reputation based upon student reflections was a previous challenge that subunits had to face, now there was the added pressure of the commissioner "memory" and the way that the sub-units needed to prioritise their responses in relation to commissioner evaluation, however well-grounded or not. Although the circumstances were different, there was a sense in which one respondent from NOR2 summed up the changing demands:

*Students come and go and next autumn there are new people, aren't there, so people can just carry on unchecked, can't they. But we can't do that; those of us working on company-specific programmes are working with customers with a memory. [NOR2e].*

Interaction and responses with regard to this theme of commissioner competence will be dealt with in subsequent chapters; particularly with regard to the decision making process in response to these pressures and demands. What is interesting though at this point is the impact that the new educational reform in Norway initially had. Despite having worked in different ways with local and regional authorities over many years, the subunits now faced a widespread demand for their programmes as a result of funding being made available. As has been seen, however, there was great variation in the types of tenders being prepared as well in the competence of the commissioners and the demands that they made. It is interesting how the involvement of this extra level in the task environment has led to more focus upon how programmes should be evaluated, even though the demands for it do not appear to be any clearer.

### **7.1.3 Perceived participant expectations and demands**

Another group within the Task Environment was noted to be the programme participants, or students. Once again it is worth reiterating that the data is based on the perception of their demands as perceived by the members of the various subunits. Additionally there is expected to be a great deal of variation across four different institutions as well as the fact that the programmes themselves are not completely comparable. However, the purpose here is to explore how and demands from this grouping appear to influence the decision processes at the subunits under investigation, that is, what they might need to take into account and how they interpret the demands.

#### ***Variation as the norm, preferences and experiences***

The demands of participants were perceived to vary greatly, with the only norm being that there was great variation, which was both cross-site and internal. In NOR1 it was recognised that demands from programme participants varied greatly by group and their contexts. Despite this the subunit still attempted to interpret the different demands, evaluating them formatively as much as possible:

*There aren't any particularly special [demands], they are very context bound and group dependent really... at the same time we try to associate them to their intensions, so we might re-evaluate or adjust what we do... [NOR1d].*

Despite this recognised variation, some factors appeared to be mentioned more regularly and these are outlined below supported by brief examples.

#### ***Implementation***

The topic of implementation highlights an issue across the subunits, where issues with the day to day running of the programme, workload, quality and relevance of course materials and preferences for particular teaching staff were

mentioned. Again these demands from students varied both within and across programmes. A good example of this was provided by a respondent from NOR2, noting how students' preferences for a particular pedagogical approach concerning programme implementation could easily clash with how the subunit wished to operate. With such variation of expectation and preferred approach combined with resource limitations and increasing class size, a dilemma for the subunit was how to deliver better and more appropriate teaching:

*It's clear that they also have special demands related to implementation, to the educational approach taken, but it goes two ways. In some cases I can just leaning against the whiteboard and talk to them for 4 hours or so, I do that, just stand there and talk to them for 4 hours. And they say that these are the best times they've had, they can sit and reflect and gain an insight into things they've never thought of; some say it's the crème de la crème. But you can't just do that, you've got to switch things around and that's not our strongest point. But it's a question of resources. Right now I have a load of papers to mark, and I don't have the capacity to meet these demands for individual feedback. It's not good. [NOR2g].*

A respondent from ENG1 agreed with this point of different student demands linked to their varied perceptions of quality, recognising the impact of participants' personalities and preferences and different appreciation of teaching forms:

*Yes, I used to run a [programme] and my opening tutor every year, 50 % of the evaluations said things along the lines of "this is the most stimulating talk I've ever encountered and it's made me go back to my school and do X, Y and Z", but the rest of them were "get him off, he's useless". And that's what brought it home to me that your own emotional and personal position wherever you are is as important as what the speaker actually does or says, which I think is fascinating in itself. [ENG1a].*

Coping with such wide variation was considered to be a challenge, and this also related to participants' prior experience. This of course is not a static variable, with attitudes changing relative to increased programme experience.

### ***Prior Experience***

In this study the main consideration is postgraduate programmes, where the majority of participants study over an extended period of between three and four years. Even in relation to the issue of commissioned programmes in Norway that might begin as one year modules, participants may often continue with the programmes that they have been enrolled in. Thus it was recognised that participants' demands concerning the programme, and evaluation in particular, changed as they gained more experience on the programme. One respondent from NOR1 reflected over how demands changed in this way:

*It's clear that students at the first and second levels have a greater need to meet us and come in and discuss things. When they get [further along] they know a great deal more of what is expected of them and what they have on their hearts, and get opportunities to present everything in a written fashion. [NOR1i].*

Another example from ENG2 showed how past experience with the institution also played a part in framing the demands of students. One respondent noted how the institution experienced that many participants returned to take further courses:

*quite a lot of our students... tend to be committed to [the Institution], in fact I've supervised people at doctoral level, who've certainly done their masters and [earlier qualifications], but some who have done a teaching qualification and have worked through [the Institution], so you do get a good following in that respect I think. Other people come to [the Institution] at Masters level because they are working full time and perhaps that's the first qualification they study on a part time basis and it fits in with what they want to do. [ENG2p].*

This led to a further reflection that with increased experience of returning to study further, participants were often positively disposed to the organisation and therefore their demands for type of evaluation of programme changed and developed:

*In terms of whether they get the opportunity to evaluate courses, they don't really get the opportunity to evaluate the whole programme as such... but of course it always falls into that you've got people who like what they know, and in fact, in the main, people, if you're well into studying for an award, you like what you know, don't you, you put up with what you know, because otherwise you'd have got out earlier.*

An additional point concerning experience builds upon the type of demands participants on these programmes had and their expectations for evaluation, related particularly to expected outcome. An example came from the subunit at NOR1, where one respondent had recognised an underlying demand from students that formative evaluation would be undertaken, noting that students, particularly at the earlier stages:

*have a very clear voice. They are mature students with a great deal of experience, the average age is about 40 at level one, and they might be a subject leader at Upper Secondary level, or more and more are middle leaders, and they demand quality when they use their time to study. So another reason that we put evaluation on the agenda is because there are actors in the field who challenge us to do so. [NOR1h].*

This comment highlighted the challenges and demands of working with professionals. Across the subunits there had been a notable increase in

participants focusing on the relevance of the programmes to their workplace, and that evaluations of the programmes should somehow reflect relevance to practice. This appeared to have especially developed in England in what seemed to be an increasingly more competitive market of providers, but it was also relevant to the Norwegian setting. One respondent from ENG2 had recognised a growing demand over time that had developed from a situation where:

*“people wanted the ability to get a masters in the shortest period of time, as a sort of a career development...this creates a contradiction between progression and really masters level work and I’m afraid I’m of the view that masters work should be very much at masters level... But I think all of that reflects the pressures on teachers these days, you know, to get qualifications. I mean, there is the career structure for teachers now and I think although academic qualifications don’t come into that, by the very nature of being teachers, they sort of think that everything needs to be academic, and so there is a sort of a tension there isn’t there. [ENG1p].*

Another respondent from ENG2 also reflected over these demands, but considered a necessity to change the structure of the programme:

*I suppose I feel quite passionately about teachers engaging in masters level study, but I do feel that for a lot of teachers we have to make it pertinent and relevant. And we have to recognise the environment that our students as full time professionals, most of them, are operating in. And we have to make things flexible so that they can pursue lines of inquiry that are pertinent to their own situations, because, I think that’s what makes us very different... [ENG2n].*

### **Voluntary versus involuntary**

Along similar lines in Norway, it was also recognised that participants who had voluntarily chosen to study came with different demands to those who had been directed to follow a study programme. One example came from a respondent in NOR1:

*I read a student reflection report from one of our external programmes where it was mandatory to attend... I got the impression that it more difficult for some of them to complete, exactly because they hadn’t chosen to do it. They thought it was particularly tough to do alongside their jobs, so you get the voice really of what it is like to complete so many study points when you don’t really have the capacity to do so. So, of course there is a difference in that way. [NOR1c].*

Although many of the demands referred to in this section appear at first glance to relate more to programme structures and implementation, there are some important factors to recognise. Programme providers have to deal with the great variations in student expectations concerning focus and implementation of the courses. Respondents outlined the expectations of students for formative



evaluation, which might take a more informal format leading to a further expectation that programmes would constantly be updated to reflect participant requirements at a particular time. At the same time these issues were often moderated by experience, in terms of the relationship to the wider institution but also with regard to length of study on the particular programme. They also had to cope with varying reasons for being on the programme, in particular whether voluntarily or involuntarily. In subsequent sections I will explore how these factors influence the design of evaluation models and the responses subunits make. As we will see in the next section, there is also pressure at the Institutional level towards increased focus on student voice.

## **7.2 Perceived institutional pressure**

Respondents across all the subunits spoke at length concerning the institutional frameworks for quality assurance (QA) and how these framed the demands and pressures upon them with regard to evaluation. The general policies and frameworks for quality assurance systems are outlined in Chapter 4. In depth documentary review of the various policies and frameworks across the four institutions under investigation revealed as was anticipated a great deal of similarity in the overall themes to be covered, mirroring those frameworks initiated at supranational level in the years after the Bologna Declaration. Despite a great deal of similarity in codified form, the responses showed internal practice to be somewhat different. With that in mind it was interesting to gain the perceptions of the subunit members with regard to the pressures and demands placed on them within their own institutions. These perceptions are considered important when attempting to understand the process of decision making concerning evaluation.

### ***Influence from wider frameworks***

Whilst it is likely that interviews with members at higher levels of the organisational systems where the four subunits were located might have delineated these demands as stemming more from the task environment, at the subunit level focus was placed on the frameworks and policies that had been internally developed in response to external demands for evaluation and quality assurance. So, across the different subunits there was a strong awareness that many of the demands reflected the requirements within the wider external systems introduced for accreditation and quality assurance of HEIs. There was a perception among some respondents in England that the internal systems were generally the same across all HEIs. This was backed up by the fact that most of the respondents operated as external examiners for other HEIs. The same was evident in Norway. The respondents also noted that these systems lacked a degree of sophistication. One respondent from ENG1 noted how even institution wide frameworks appeared to be implemented intermittently, even though they were framed as a requirement:

*So, I think that you'll find that that is quite a common pattern for British Universities, so it's not overly elaborate. Some universities, indeed here in some modules, they do an evaluation sort of halfway through the course to get some sort of formative feedback and so on. Personally we don't do that within in our modules, but we have been kind of encouraged to, but I think there's a limit [ENG1g].*

Just as the programme providers in England had reflected over how funding had become more tied to external financial frameworks, a similar pressure was perceived evident upon the institutions with regard to meeting the requirements for quality assurance. The same respondent summed up a general point of why the Institution appeared to be acting to meet the wider needs:

*I think because they are obliged to; their funding is dependent upon it and we are working within a national framework. Whether they take it seriously or not I think is another matter. [ENG1g].*

There was an impression across the four different subunits that the wider institution in framing evaluation requirements was focused upon responding to the demands from policy makers. There were five main areas of demand discussed within the interviews that were thought to come from the Institution with regard to evaluation: programme validation, systematisation, operative control, throughput and student satisfaction.

Demands and pressures	Underlying issues
Institutional	Validation
	Systematisation
	Operative control
	Throughput
	Student satisfaction

### ***Programme validation***

Programme validation as a category stands out compared with the others, being much to do with the genesis of a particular programme, and gaining the institutional approval to begin. This requires meeting the academic demands associated with the particular HEI. As the systems were very similar to the basic framework outlined in chapter 4, the specific frameworks are not outlined here. There were general comments to how this influenced the programme formation more widely, but with regard to evaluation it mainly meant that certain formalities had to be outlined from the beginning, this was the first stage of evaluation. An exemplifying comment was made by a respondent at ENG1, who noted that:

*Well, we are responsible for designing the programmes, but as with all universities we have to have them approved through a kind of validation panel and there are certain rules and regulations, obviously certain things that we have to agree with, and concur with. [ENG1g].*

This process of validation generally caused little reaction, and the respondents recognised the need for academic standards and the importance of having evaluation systems in place for accountability. But, at the same time, rigidity could be a problem when programme providers sought to work outside of the normal framework. And as was seen earlier, developing programmes for a rapidly changing market, with in turn concrete areas of demand to be covered, was challenging for HEIs. At the same time HEIs were being forced to prioritise and streamline more than ever before. One respondent at ENG1 saw this as limiting programme development:

*it seems to be now that there's a pressure to move towards generic degree programmes, where there are specialisms inside an overall degree structure and the tolerance for unique or individual types of programmes has become sort of less... they're less enthusiastic about it. And I understand the commercial reasons why, but I think that it's also rather sad that we're not actually really customising programmes more acutely for different sorts of clients and different sorts of audiences. [ENG1b].*

The development of the commissioning system in Norway had also led to particular issues that created some difficulties for the subunits, who had to manage a tension between complying with institutional frameworks and being flexible in relation to external demands. One respondent at NOR1 encountered this tension when the programme was in its design phase:

*We have to relate to the demands placed upon us by the Faculty, that is, what a master programme 'is'. When we made the decision to have an experience based programme we discussed initially having 90 study points, but the Faculty wouldn't approve it so we had to have 120... and they also demanded that the level covering research methods, philosophy of science and statistics should [be part of ] a common study plan within the Faculty. So there are the formal demands. You can say that external demands influence us, as an academic group we don't have a free hand; it's the Institution that ratifies. [NOR1k].*

This was also an issue for the subunit at NOR2, and this meant that the commissioners needed to accept the internal demands when negotiating for the programme:

*And so there are also demands from the academic committee for standards; demands concerning curriculum, implementation form, enrolment. This is at master level, and not everyone makes it through the needle's eye, etc., etc. It has to pass the academic committee. Our customers need to understand that a*

*programme needs to be as high a level as our standard programmes and we can [only make small adjustments] to a certain form. [NOR2e].*

But, as another respondent outlined, this also went beyond formal requirements and issues of Institutional demands that were based on frameworks to be followed, and touched the academic identity of the programme providers and their perception of what the basis of the studies should be. As will also be seen in subsequent chapters this impacted decision processes:

*Yes there's clearly an area of conflict here, because we are an academic institution. There is clearly some premise here for why we do it. And you can say that the same applies to the staff themselves, we have some academic foundations, if we are going to talk about leadership, public sector leadership, and in this case school leadership, then we have a particular way of doing it and the underlying reason for doing it as well. [NOR2g].*

This example also leads into discussions in section 7.3 concerning perceived within unit demands as well as being revisited in chapter 9 when considering the decision process and responses to external demands.

### ***Systematisation***

Respondents also considered how the demands for evaluation had required greater systematisation. With external accreditation and quality assurance bodies there was a sense that systematised information needed to be in place, as well as increased demands for evidence of output in certain areas. Respondents discussed how this had led to greater systematisation across the academic year, with clearer frameworks in place, and this in turn required greater internal systematisation on the subunits' programmes. One respondent from NOR1 exemplified this:

*We have the demands that there should be a midpoint evaluation, and a summative evaluation, so it happens a couple of times in the year and we have now been working at developing our surveys to students, such that there is parity across the different levels. [NOR1c].*

Another respondent from NOR1 reflected over the fact that for the institution it appeared to have clearly become important to have the information that was required by external accountability systems:

*I think that for the institution, in relation to meeting points with NOKUT it's about having the right information as required. [NOR1f].*

And this systematisation was interestingly considered by another respondent to have contributed to the success of the programme:

*And this has been one of the reasons I think why we have succeeded with our study, that we have a large institution behind us who have placed demands upon us to develop plans and tell u show to evaluate our study programme etc., such that we have been continually aware which criteria and what applies, and how we should handle the problems of developing a programme. I think it must be harder for other colleges to manage this because they haven't got the apparatus to deal with it. Amongst other things with the development of the master programme the Faculty initiated a project for the Quality Reform at once and everyone had the same tasks. Even though the demands changed along the way and there was a lot of back and forth, we did get some good help from them. And it also applies to evaluation, now that NOKUT have demanded that we follow the same criteria and systems common to all colleges and universities. We as an academic group have noted this [NOR1i].*

However, there was also the consideration that the systematisation experienced was more of a bureaucratic response to external demands than an attempt to control:

*Yes, I also feel that the organisation as a system is more focused upon the structural side, that is, getting systems into place to put it in that way, because it is quite premature for higher education to think in terms of evaluation and systems and quality assurance. There's more of a bureaucratic demand tied to it, more than leadership and governance. [NOR1h].*

In ENG1 respondents also described the institutional framework and demands associated with it as bureaucratic, but perceived more of a control emphasis. One comment exemplified this formality of the system, as the respondent very simply summed up the demands as experienced:

*well, there's the institutional purposes, which are very bureaucratic and they require various things under certain columns [ENG1a].*

Respondents referred to an increasingly greater rigidity with regard to systematisation, but such developments were also seen to reinforce demands and pressures to follow models currently in fashion. This in turn challenged the adoption of any varied ways to assess and evaluate the programme:

*the assessment criteria which we impose on the programme are very much dictated by the [NN-organisation]. So, the assessment criteria are far more in line with a rather traditional master degree, than an innovative degree. And I actually personally think that's problematic [ENG1b].*

Respondents in NOR2 also referred to the process in a similar way, recognising that the central system imposed a framework that the subunit needed to employ. However, there was a sense in which the respondent only adhered to the minimum requirement:

*But it's also, when I say bureaucracy, it's not my bureaucracy but the admissions office, examination office; all the rigid systems that live their own lives and that I need to relate to. But I don't sit and write long reports, luckily, no. [NOR2e].*

The next three concepts deal more in relation to what the evaluation system was perceived to focus upon, and thus outlines the basis for the main institutional demands. One respondent from NOR1 summed up these three categories, there were fairly common responses across the subunits:

*What do they want to know? I think what they want to know is something about the quality of education, and I think they want to know about the results and throughput. And they probably want to know something about resources, how much we use, and to see the connection between effort and result. [NOR1c].*

As will be seen, gauges of educational quality had often been reduced to that which was generally considered to be based on that which was ascertained by the students, and this appeared across the subunits as a main form of focus. Building further on the comment above, the other main foci were what could be described as operative control and throughput.

#### ***Operative control (cost effectiveness)***

The concept of operative control also includes the concept of cost effectiveness. Respondents spoke of the increased demands of having control over the implementation of the programme and that evaluation was performed to the institutional framework. A good example of this was described by a respondent from ENG1, who considered these processes to be quite basic:

*Well, we have within every faculty we have a kind of quality assurance group and through the University as a whole we have a Quality Assurance / Quality Enhancement Committee, which kind of, you know it's quite a bureaucratic process where they will require of us evidence and some of that evidence is kind of form filling, going through the motions, have you got minutes of meetings and so on, but also sort of student feedback, module evaluations, course evaluations, all of that. But it's at that kind of level, it's not very sophisticated I think. [ENG1g].*

A similar situation was retailed in NOR1, where focus on the degree of operative control had increased. Alongside evaluation being about development and programme control these increasing pressures were considered to be about legitimating the very existence of the programme related to cost. This required greater control over the implementation of the programme and a quantitative assessment of the activities of the programme staff, which was not

unproblematic when balanced against the increased demands the HEI reform had introduced:

*And so you can say that the need to be externally considered legitimate has grown stronger in more recent times, and it's also about economic relationships, you have to legitimate that there is a cost to running a study programme and you need to be able to justify what you use the money on. For example we have had a number of reports up to board level where we have documented how great a need we have in terms of personnel in order to deliver such a programme with so many students. So you go in and calculate how many hours you have to disposition based on the employment contracts, compared to the rights the students on a master programme have in relation to teaching... [NOR1k].*

Another respondent at ENG2 also recognised this shift towards cost effectiveness, noting the complexities for those organising programmes at the postgraduate level, particularly with regard to participant numbers:

*Well, in more recent years I think it's more and more about being cost effective. At the bottom line it's become more and more important and because nearly all my work is at the postgraduate level, that creates a tension there, because you're not going to have such big... courses in terms of [student numbers] [ENG2p].*

It was further reflected that while many of these demands appeared to develop externally, there was also a sense in which the Institution had also been part of the process as well, ensuring that demands would be met:

*So the pressures I think are there from the funding agencies, but a lot of them I think are probably internally generated from within the University to make certain that the standards are being met. [ENG2m].*

The increasing demands for the evaluation of operative control were generally perceived to be problematic and time consuming, with greater focus upon legitimated forms of assessment. This had highlighted another area that had developed to become a quantitative measure of programme quality, throughput.

### ***Throughput***

One of the implications of highlighting operative control and cost effectiveness was noted to be the challenge of achieving greater throughput. While this was noted to be the result of wider HEI reform policy, the recognition across the subunits was that it had become a complex evaluation measure and was difficult to achieve, especially when these programmes dealt mainly with part time participants usually in full time employment in demanding leadership positions.

In ENG1 a respondent recognised that this had become a kind of performance indicator for programme effectiveness:

*And you know for example from an institutional point of view, where our course is effective or not is that they actually they look at it in terms of dropouts and, you know, how many people pass the examinations etc. And that's another way of course of how institutions evaluate courses, which is quite interesting. [ENG1a].*

In NOR1 the subunit had experienced this as the set demand that had come as a result of HEI reform, despite an otherwise general freedom from other pressures:

*There aren't any demands upon us now. It's more, well if one should talk about it then it is throughput, in relation to the Quality Reform, that results in relation to throughput have come more into focus. [NOR1d].*

And here was another form of funding pressure, but rather than expectation of programme impact, HEI budgets were more greatly affected in proportion to the numbers of students completing. In this case it had been a successful measure as the subunit had seen many students completing their studies on time relative to other programmes at the institution:

*And [the organisation is] of course focused upon throughput, that is, are we getting our students through. That is, the quantitative aspect, how many students started, how many completed, how many take the masters; it's actually about having the economy to continue a study programme... it's about quality assuring that we are able to run a leadership programme that is good, and it's first and foremost about getting students through, because then the organization gets funding about 2 years later. And on our programme we have a great deal of students and we get a lot through so we are a bit of a goldmine for the other master specialisations with few students. So that's one side of it, the economic and quantitative. [NOR1h].*

And another respondent summed up the situation in an interesting way, but reflecting that the demands did not relate to evaluation as the subunit interpreted, despite being part of the evaluation system. This was another kind of control:

*Well I think it like this, that we would have done these things even if no one had asked, at student level that is, and the other things are not particularly demanding. It's about communicating... for me it's only about communicating information so I don't feel any particular demands related to evaluation. But that which is there as a constant demand is related to throughput, and to attract enough students and earn study points, but I don't feel that this is really tied to evaluation. [NOR1e].*



### *Student satisfaction*

With regard to the next concept of student satisfaction, respondents across the subunits spoke of the increased focus upon student voice in the evaluation of programmes. This concept overlaps a great deal with those previously outlined. The QA frameworks all highlighted the importance of gaining student feedback and ascertaining demands. In NOR1 it was clearly felt from within the system that reflecting student voice was important, but was also experienced within the programme group.

*What I think is that the student voice should come clearly out, and we are focused upon getting feedback from students in relation to development and improvement and that it is put on the agenda, discussed and used [for these purposes] [NOR1d].*

In NOR2 there was a noticeable perception too that this was a major focus in the organisation:

*Our goal is to make sure that student needs are met; that's the point for us. And are they, then we have achieved what we set out to. [NOR2f].*

Another respondent considered this to be the preeminent focus of evaluation in the organisation. However, at the same time, this was a different reflection, recognising that it was not easy to assess how such a satisfying of needs should be interpreted and at what level:

*What you can say is that in the organisational evaluation system, so it is all about whether the student is satisfied, the happier they are the better it is. The happier they are with the teacher, the more approval they give. I've never experienced that the evaluation system has ever been used for anything else that to check if people are satisfied, and perhaps do something if they are not satisfied with a programme. Perhaps with the administrative side, and we have had a lot problems with that... but they've be dealt with well, based on some of the evaluative comments. But the inner life of the programme is not touched by the evaluations at all. [NOR2h].*

Evaluating at such a level while initially might appear quantifiable over time it was much more problematic. As one respondent in ENG2 noted,

*Yes, we have to demonstrate that what we're doing meets University requirements and expectations. We've got to be able to demonstrate that it meets student needs and expectations and requirements, and that one I think, the first one in a sense is relatively straightforward, there are systems and structures in place that I was talking about earlier. The second one is the problem, and ultimately the only way in which you do that in the kind of situation which we're*

*in is if people keep on coming back, and if you continue to get your numbers. And the numbers on the programmes have been quite consistent. [ENG2m].*

Another respondent agreed with this reflection that satisfaction was under focus, where students were expected to come back:

*Do they do one course and then they come back and do another and another, or do they stick the course, I mean recruitment and retention. To be honest those are the bottom lines for us about evaluation in the end aren't they. [ENG2n].*

While there were differences between the programmes across the subunits with regard to how places were funded, whether participants were fee-paying or not or whether commissioning bodies or the state funded grants were used, it was interesting to see that these issues raised above formed the broad sweep of Institutional demands. And as was outlined above, it was recognised that these demands all appeared to coincide as well as being on different levels:

*There are demands that we just have to address, those set by NOKUT, from the organisation centrally, as well as our Faculty...[NOR1].*

### **7.3 Perceived within-unit demands**

In the sections above there is recognisable within institution variation as well as different external demands, but there are many comparative themes. What is perhaps most striking are the patterns of demands affecting the subunits which have pointed to a complex context for the subunits to negotiate. As we will see in subsequent chapters these affect decisions made about adoption and implementation of models. But there were also important reflections made by respondents concerning their perception of within-unit demands. In chapter 10 I will explore how these framed the decision processes as the sub-unit attempted to explore the perceived demands, but here I outline some of the main concepts drawn from the data with regard to within-unit demands.

Demands and pressures	Underlying issues	
Perceived within-unit demands	Improvement focus	
	Academic anchoring	Projecting competence Protecting professionalism

Perceptions of within-unit demands focused mainly on two interlinked concepts with relation to evaluation and assessment. The first was recognition that the subunits considered themselves to be improvement focused, the second that the programme's academic basis and values were central in driving this focus. I do

not deal here with the perceptions of the subunit decision process and the different roles performed by members, but rather the overall demands experienced within the groups.

In NOR1 there were many who referred to the within group demands that they experienced, with a strong improvement focus and drive to evaluate and develop. These initial demands were felt to be preeminent. One respondent reflected that:

*But it's more about the demands that we place on ourselves, and these are much more demanding. That we continually want to give good supervision, and to lecture such that students get the best they can etc. They are tough demands... but the [formal] system is quite simplistic. [NOR1c].*

And the respondent further noted that the subunit considered themselves quality focused at a deeper level, which was about being part of an academic group that worked closely together:

*I think it is part of our culture here that we are focused upon quality... and that is the most important for me. Because I am part of a professional community, I'm focused on quality. There is strong desire to learn amongst the staff, to learn more for oneself and develop our programmes. I think that is why people want to come here, and others want to cooperate with our organisation and our group. It's a good advert, a stamp of quality. [NOR1c].*

Another respondent reflected that the attention given to evaluation within the subunit was not considered to be externally driven but rather followed from an internal desire to develop and improve the programme:

*So it isn't just because a regulation comes down from above asking us to evaluate our programmes and we do it out of duty just so someone can tick it off and put it in a drawer. It's much more about dynamic processes and continual improvement of our practice. And we talk a lot about this, if you consider evaluation more broadly like I am now. I think it is wise to see it as a big field. [NOR1h].*

This sense of talk and reflection over evaluation is dealt with further in chapter 10, but the same respondent also recognised that the wider demands from the organisation for systematisation, evidence of throughput and degree of student satisfaction was unproblematic as the data could be collected from their already existing database as these were areas of focus for the group rather than demands to be met. The respondent referred further to the degree of within subunit pressure to develop the programme, which had involved focusing on such areas. But, what the demands had done was create a need for greater systematisation of the evaluation processes, and these were a natural progression for the group:

*We are trying to systematise this and [the programme leader] is working with systematising all the contributions we've made in order to inform all those making requests. But it's not anything that worries us. We don't experience it as tiring or worrying; at least I don't see it that way... [NOR1h].*

The implementation of these wider demands had appeared to benefit the academic group with regard to requiring an extra level of reflection as the information needed this systematisation, despite the lack of experience in receiving any feedback. There was a sense that any increased understanding that would lead to improvements in programme delivery was welcome and helpful. There was still after all a great deal of freedom for the academic group. These processes are taken up further in Chapter 10 in relation to decision making as I consider a special case of 'collegial construction'.

In NOR2 there was also agreement that improvement was in focus across the unit, and this again was based on the improvement of own practice, but far from being tied to any wider framework it was left to the individual to decipher for themselves:

*In my logic this is something that I am always doing something about; I always do something about the weakest point each year. So I try to find the weaknesses, where we are worst and do something about it. And next year we ask the same question, what is our weakest point now... [NOR2g].*

And within ENG1, despite outlining a relative cynicism to the accountability focus of evaluation in the organisation, one respondent recognised the positive influence of the improvement attitude within the course team which was directed at programme development and based upon the professionalism and professional judgement of those comprising it:

*The real area where quality will kick in will be at that course team level, where you have a group of professionals discussing matters and wanting to ensure that what they are doing is of a high quality and meeting the needs of schools and teachers. [ENG1g].*

And there was also reflection within ENG2 that over and above the institutional demands a critical reflection regarding programme achievements took place:

*But then inside, for us internally, it will be about, it's much more complex, it's whether we feel we're providing a worthwhile learning experience and that can depend upon people's perspectives. [ENG2n].*

However, as we will see in subsequent chapters, not all the units respond to these demands in the same way and the decision processes described were quite different. There was though a noticeable reflection in each of the subunits that the combination of different demands caused a general tension between selling

programmes, educating individuals, producing study points and defending an academic position. Responses were reflected in a statement made by a member of the NOR2 subunit:

*It's clear that within the whole [organisation name] system it seems to be that you sell a product and that product should be both academically defensible and at the same time relevant for the customer. And it's clear that there is a field of tension that is even more greatly noticeable in such a setting as an institutionally bought management programme that is to fit within a master's programme. You have nearly everything, the customer who pays out millions, and it's not just a leadership training programme it's a masters. And it's actually a problem, there's so much, you need to have a compendium of 2000 pages and focus as much as possible on processes. How are you supposed to focus upon the curriculum? So I feel like in my programme all the demands come together [NOR2h].*

#### **7.4 Summary**

Following from the responses given this area of demands was divided into three major areas, focusing on the task environment which included the externally oriented demands from policy, participants and commissioners, in addition to institutional demands and within-unit demands. These groups were also generally recognised from the literature review in chapters 2 and 4, however the focus on commissioners was an interesting finding in the Norwegian subunits as were reflections over the NCSL in the English setting. The main foci are summarised below.

In terms of the more “external demands” from the task environment of programmes, it was generally recognised that moves for greater accountability were felt throughout the system. Despite the different bodies involved in England and Norway, and different perceptions of sub-group members, there was an increasing focus on demonstration, or at least discussion, surrounding practical impact of programmes. Sub-groups discussed defending the elements of academic reflection. The mechanism to achieve this greater accountability was formative self-evaluation evidenced within the University quality assurance systems, although it was evident across the different sub groups that the introduction of such systems was by no means uniform. This also accounts for the variation that will be seen across the subunits in subsequent parts of this study. In England there had been increasing pressures to demonstrate impact of programmes, and these had been highlighted since the forming of the NCSL in line with national demands based on improving school effectiveness and contributing to evidence based policy. In the Norwegian setting, the involvement and influence of commissioner groups, seemed much more immediate in terms of the implications of programme evaluation, yet at the same time I have outlined and will outline further certain challenges regarding perception of their competence.

Student demands appeared to be increasing with the ascent of the role of student as consumer. Their responses also were felt to buffer their reflections of the wider organisation as well as perceptions of what the programme was able to offer. All groups recognised that the evaluations that students took part in were limited, due to the longitudinal nature of development focused studies, and there were difficulties in developing good measures to respond to demands about impact upon students and their progress. Additionally came the challenges of comparing personal preferences of students with programme frameworks and goals.

With regard to Institutional demands, there was an increasing reflection of the impact of the organisational system upon the evaluation process, where the demands were generally perceived to be bureaucratic and little response was forthcoming top-down. Subunit members perceived this as mainly based on satisfying quality assurance systems that are downloaded from national policy and adapted from the wider accreditation and evaluation bodies. These seemed to be given greater credence due to the wider acceptance of the use of evaluation to demonstrate fitness of purpose. This meant that alongside funding demands there was a greater demand for throughput and good scores of satisfaction. At a wider institutional level there were noted to be development of units to take responsibility for quality assurance, particularly of student surveys. Although these processes are not explored in detail, the perception of respondents with regard to the impact upon the subunit is dealt with.

With regard to within-group demands, this was an area of interest to the respondents and worth exploring even further in future studies. Although this area will be dealt with in the next chapters, it is important to note here that there was suggestion across the different groups of different needs, due both to academic experience, structure of programmes, focus of leadership style and size of staff etc. However, an important area of interest is the sense in which groups felt they needed to defend decisions to one another academically, described as an occupational community.

It might be therefore considered that this is a framework of responses to constraints and contingencies. There was an increased perception of competition in the English subunits, where nationally sanctioned training as compared with development programmes is being offered. Those selling distance learning to domestic school leaders has decreased, more of the distance programmes were being sold abroad. Different funding demands were to be met when designing programmes, in addition to the requirement of bids to deliver programmes and increasing demands placed by different funding agencies on evidence of impact. In Norway where programme are sold (either in part or whole) to a local authority buyer the growing issue of negotiation was raised, where subunits had to account for what the buyer expects to get out of the programme, allied to the

fact that it is a HEI programme subsequently answerable to the organisational framework.

Attempting to combine the demands had appeared in each subunit to raise a question of balance as well as perceived overlap. I reiterate that the purpose of this study is not to focus upon this variation per se, but to outline the processes of response to demands placed. Subunit members discussed the question of what were the necessities but not necessities of the evaluation task. There was a sense in which all were now required to perform evaluations in a more structured way but in each case there was a lack of definition of what they should be looking for. This was complicated by the demands across the levels. Such complexity was an overriding theme in the responses given. Weick draws on Huber and Daft's<sup>173</sup> (1987) "environmental determinants" of "perceived environmental uncertainty" as an "occasion for sensemaking" (1995: 87). One of these determinants is "complexity". This environmental complexity increases uncertainty "because a greater number (numerosity) of diverse elements (diversity) interact in a greater variety of ways (interdependence)... [affecting] what people notice and ignore" (Weick, *ibid.*). This environmental complexity is thought to trigger decision situations in organisations, in this case it is thought applicable to the subunits under study. Maitlis and Lawrence (2007: 77) noted this importance of this concept of "environmental complexity" and applied it further in discussion of "triggers and enablers" of sensegiving, an attempt to influence others sensemaking. These concepts and allied approaches were first outlined in chapter 5 and will be taken up further in chapter 10. However, it is important at this juncture to note that Maitlis and Lawrence's research findings in this area was mainly isolated to leaders in organisations. In this study, conversely, it appears that the subunit members as stakeholders also attempted to influence the sensemaking of those in the wider environment as well as internally as a result of perception of environmental complexity due to their competence area and boundary spanning tasks (Thompson, 2003).

It is recognised that some other sources of possible demand or influence appear will no doubt be present. Based on respondents' perceptions the major areas of demand have however been outlined. The findings from this section are interesting, as there is a shift in expectations as well as their source; for example, the external environmental demands have become narrower, or at least are perceived to be so, and internal requirements both reflect this and also react against it. These reflections will now be dealt with in the ensuing chapters where I consider the perceived responses to these demands. This leads into the next chapter which focuses more closely upon the evaluation designs that were implemented and later in chapter 9 focus is placed upon how decisions were taken to form them.

---

<sup>173</sup> Huber, G.P. & Daft, R.L. (1987). The information environments of organizations. In F.M. Jablin, L.L. Putnam, K.H. Roberts & L.W. Porter (Eds.), *Handbook of organizational communication* (pp 130 – 164). Newbury Park, CA: Sage.

## 8. Evaluation designs

In the previous chapter it was recognised across the subunits that members outlined a complex web of demands, although there was seen to be variation according to the contexts under study. At one level demands were noted to be related to their particular subject area, educational leadership, and associated national policies and at another level in relation to being HEIs and associated frameworks for quality assurance. This chapter deals with the problematic area of evaluation definitions and designs, as discussed in section 5.9, and outlines and discusses subunit members' perceptions of the evaluation designs implemented on their postgraduate programmes. The intention is not to outline the evaluation frameworks, these are described in the documentation each HEI has developed and as was recounted in chapter 4 and 6 these were seen to be generally similar. The purpose is rather to investigate how and why the frameworks develop as they do, accounting for any divergence from the written presentation. But in order to discuss these processes in a more detailed way, the basic frameworks of evaluation are referred to.

Chapter 3 outlined evaluation definitions and how different models and frameworks can subsequently develop dependent upon one's conceptual understanding and approach. It was recognised that evaluation is thought to be a value based judgement with a utilisation focus, and is considered to be decision based (Owen, 2004; Scriven, 2003; Stufflebeam et al., 1971). The utilisation focus creates, however, a dichotomy where evaluators often consider that the findings they report are prepared to be archived rather than for resultant action. The theoretical underpinning of evaluation was also discussed, where it was proposed that different evaluation models that are implemented are little influenced by theory, amongst both experienced and less experienced evaluators (Christie, 2003). The main foci of interest in the evaluation field were considered by Christie to be stakeholder involvement and method proclivity. This chapter also attempts to consider the influence of the organisational setting on evaluators. Following Stufflebeam's (1983) research into evaluation models and designs, focus with respondents was also placed upon discussion about choices related to traditions, expectations and experiences. There is an overlap between this and the next chapter; responses that mainly concern designs are placed in this chapter and those primarily related to the decision process in the next.

In discussion, respondents were asked to consider the purpose and focus of the models implemented on their programmes as well as their own attitudes to evaluation. In doing so, they were invited to discuss theoretical and practical factors thought to influence the design process as well as any limitations they saw in the models. Respondents also commented upon the issue of ascertaining effects and impact of the programmes and how these were reflected in the designs.



## 8.1 Underlying frameworks

The initial area of focus for discussion about evaluation designs and models was the underlying frameworks for understanding evaluation as perceived by subunit members. Within these themes 3 related intertwined issues were noted in the data collected. These are outlined in the table below.

Evaluation Model and design	Related issues (interlinked)
Underlying frameworks	Attitudes to evaluation
	Reflection over influences
	Current evaluation designs

### 8.1.1 Attitudes to evaluation

During the interviews respondents were asked to outline their views about evaluation, and to consider what, if anything, had influenced these views from any espoused theory as well as their own experiences of practice. Respondents spoke about their attitudes to evaluation and structures within their organisations as well as wider in society, exemplified across the subunits with discussion concerning the difficulties in balancing the different demands of evaluation within the current structures imposed on HEIs. A clear example of this was commented on in interviews at NOR1. This involved taking into consideration the goals of the programme, alongside that of various stakeholders, whilst attempting to include the relative importance related to aspects of the programme by the participants, in addition to demands from different mandators. When asked to consider generally the purpose of the evaluation process and the complexity of reflection, the respondent reflected over the formative importance of evaluation from a theoretical standpoint. This appeared somewhat different from the current models of evaluation being downloaded from policy frameworks. The respondent focused more heavily on the importance of ‘voice’, discussion and improvement rather than upon control and legitimisation:

*I don't see evaluation as something that just happens at the end... Something that I have experienced strongly ... is that both the authorities and institutions [have created] easier systems and we are very preoccupied by systems and structures. And the summative evaluation is part of the legitimating aspect for what we've done and what we should have done and no more. Development comes if the evaluation is to be used to develop something new, so just as much comes through the processes, ideas and experiences along the way... that can come from the students, or my colleagues, they can come externally, where you try it out to see if it works. When you evaluate you do it along the way and of course afterwards. It is important to get many voices heard, but not necessarily those who've experienced it have the strongest voice though. [NOR1e].*

These comments were not uncommon across the subunits where respondents outlined the importance of stakeholder feedback for programme improvement

combined with professional assessment. Whilst few respondents related their reflections to specific theory a pattern emerged in the responses across the subunits where evaluation was seen to be a professionally developed mind-set that formal systems should not negate or detract from. These reflections were also mirrored by another member of staff in NOR1, who recognised that the formal system had its place but it was within the process based, professional assessment that much of the reflection took place and this was seen to be about developing an “evaluative way of thinking”:

*Evaluation isn't just about the formal form of evaluation but about an evaluative way to think in practice, such that we become very conscious of it. And when we are organising we constantly have an evaluative glance at what we are doing, asking what have we experienced now, do we observe when learning took place or good processes developed. So I recognise that we ought to and must have an evaluation system, but it mustn't take away from us the daily focus upon observing and searching for and reflecting upon whether it is a good situation that allows for good processes or not.... So I don't believe in those processes where you sit down now and again and perform some kind of formal assessment, then put the results in a drawer. I think that it has to be a daily part of the work, where you have, in addition, an evaluative way of thinking. [NOR1c].*

This was also described by a respondent in ENG2, who considered the importance of developing a more “mature” form of evaluation that was focused on everyday activity, balancing the best of formal structures that could support the professional reflections as well as assessments gained formatively. Respondents across the subunits saw this as developing one’s own *evaluation vocabulary*. In NOR2 discussion reiterated these points where respondents also spoke of the challenges from the reductive nature of central QA systems to such a vocabulary. In ENG2 one respondent reflected upon how the evaluation theory taught on the programme matched their subunit members’ own particular view, highlighting participation and based on self-reflection. This suggested that drawing out the best from practical approaches, rather than having any theoretical or methodological favourites. In considering this issue another academic member of staff in NOR1 more fully outlined the benefits and merit of considering evaluation from different paradigmatically oriented positions, while noting interestingly that often, problematically, the processes became confused:

*I think that evaluation has two types of reasons, the one tied to seeing how, in a way, things went and the other is tied to seeing how it is going along the way. My perspective is that evaluation can have both of these purposes, and probably should have them, but what I think is problematic is that you should try to do both at once. So what I think would be good in the future is that someone develops, and I think they will, different methods for the two purposes, just because it is a kind of mix-up when you perform an evaluation where you often want to look at the processes and these are disturbed by the focus on results. So*

*these aspects need different theory bases if they are going to be mixed together. [NOR1j].*

These reflections highlight a perception across the subunits that evaluation was becoming synonymous with a focus upon results, and was increasingly based upon evaluative theories that supported this approach. Finding a process orientation within the structure was not considered impossible, but it was not the main goal of the evaluation exercise with regard to demands coming from above. Here it was identified as being a legitimating activity, rather than a developmental process. At the same time the different subunits reflected that the within-group attitudes and processes represented more of a developmental focus than was evident in the formal frameworks of their institutions and wider afield. These within-group attitudes and processes are dealt with in more depth in the next chapter.

### **8.1.2 Reflections over influence from subject field / profession**

This subsection concerns the influence of the subject field /profession upon respondents' perceptions of evaluation. It follows on from the reflections in the previous subsection concerning the relationship of programme goals and content to respondents own perceptions of evaluation theory. Interestingly the "field" that respondents described straddled the idea of being a professional academic or administrator within higher education and that of the subject specific arena related to school leadership and management with the accompanying impressions of how activity in schools should be assessed and evaluated.

Following on, one respondent from NOR1 reflected that this agency often challenged the basis of programme intentions. This was exemplified on the one hand by the fact that evaluation was one of the programme topics. How evaluation was handled on the programme raised a point of interest:

*My experience is that we as an academic group [at this institution] work to quality assure [our programme] because we want it ourselves, not just because someone asks us to do it. It reflects that we work a good deal with evaluation; it's a theme in our field. We teach about quality assurance and evaluation for school leaders and teachers, so we are no doubt observed to see how we do it ourselves, and so perhaps we approach it with a little more consciously than some of our colleagues working with subject areas like the sciences [NOR1h].*

But this was also seen to be important when considering in what manner specific content and subject matter of the programme was delivered, exemplified with regard to how the theme of leadership of educational institutions was explored in lectures and tasks related to the way the academic group operated. The respondent recognised that their approach could on occasion be construed as normative:

*We need to look at how we speak about leadership, how ideally leadership should be practised. Sometimes we try to be descriptive but end up being normative. So we need to hold up the mirror and ask, what do we think about leadership, what does follow up and assessment mean for us, how do we work as a group, what are the interrelationships like? In a way we have to explore the themes that we teach and examine ourselves. [NOR1h].*

These reflections reflected also in other comments coming from NOR1, were slightly different from the other groups. There was a general reflection across all groups that the content of the study programmes related to evaluation influenced the way the groups evaluated their own work and their attitudes to evaluation. Group members varied in experience, but all to a greater or lesser degree had experience of evaluating. What was different in NOR1 was a more regular group based reflection about evaluation, as opposed to more general group discussions on study progress. This, as I have already mentioned, is taken up more fully in the next chapter.

### **8.1.3 Current evaluation designs**

As was noted in the previous chapter subunit members had recognised an increased demand within their institutions to systematise and formalise the evaluation and quality assurance activities on their programmes. Across the organisations, respondents outlined how the frameworks and procedures had developed while they considered their perceptions of structures, purpose and focus, and the strengths and weaknesses of the operative models.

In all of the organisations respondents related the discussion concerning the models of evaluation to the quality assurance systems within their wider Institutions. At the same time there was reflection concerning how evaluation was carried out both formally and informally, and how the subunit activity fitted within the wider frameworks. In this subsection I briefly outline the structure of the evaluation systems as perceived at the local level. The purpose is to gain an understanding of how subunit members observe the framework to function rather than to outline a detailed explanation of all constituent parts.

Despite recognition by respondents that different organisations would likely have developed their models differently or be at different stages of development, there was a sense in which all the subunit members perceived an increasing convergence across HEIs in terms of the purpose, focus and structures of evaluation and quality assurance frameworks. As one respondent from ENG1 put it, “I think that you’ll find that [the framework] is quite a common pattern for British universities”<sup>174</sup> [ENG1g].

---

<sup>174</sup> The same point was also made, along with time restraints, as part of the reason why one of the subunits that was approached declined to be part of this study.

When asked to outline the framework of evaluation within the subunit, respondents described the structures as a combination of formal and informal methods and approaches. In NOR1 one respondent provided a good overview of the basic structure of the evaluative framework in operation across the subunits investigated:

*Evaluation? It's partly formal assessment and informal assessment. It's also the feedback gained along the way, at every gathering, where we systematically collate what we have learned and how it functioned etc. [NOR1b].*

As will be outlined briefly below, the systems were now considered to reflect the more formal demands to provide data concerning the quality of implementation, examination and throughput of students, based upon internal analysis, external assessment, grounded in and supported by student feedback. Further discussion concerning assessment of programme impact will be outlined throughout subsequent sections, but with particular focus in section 9.5.

An initial point that is important as a backdrop to consideration of evaluation models and designs is that all of the programmes under investigation had been validated by the equivalent programme committees within their institutions. As was recognised by a respondent in ENG1, this means that programmes are framed within an institutional structure and thus considered to comply with specific demands, which as seen elsewhere increasingly include formal plans and designs for evaluation:

*We are responsible for designing the programmes, but as with all Universities we have to have them approved through a kind of validation panel and there are certain rules and regulations, obviously certain things that we have to agree with, and concur with. [ENG1g].*

Respondents widely recognised this as providing a remit for the course, although admitting that this was not always an easy process – having to strongly present the case for the need of a new programme. In addition, as exemplified by a respondent from NOR1, there was consideration that this provides part of the quality assurance of the programme, as well as ensuring an academic base to work from. As can be seen from this response in the case of NOR1 the Institution has accreditation enabling it to develop its own programmes at master level, as well as institutional autonomy over programme content. But it is interesting to consider that these two processes are equated:

*And it is such that [HEI-name] decides; we don't have to have our programme approved by the Ministry. We have established our own routines for approving a HEI programme, and there is a built in quality assurance process because for us there is an academic assessment tied to it. The [HEI-name] is free to design programmes but you have to be able to argue for them as there are many different interests, and there can be conflicts of interest. Because of that you need good*

*argumentation for why you might think it's a good idea to establish Educational Leadership as a field of study. [NOR1k].*

This consideration of the academic anchoring of the programme creates an interesting backdrop to the discussion which will be outlined in the next chapter with regard to how decisions are made about evaluation and on what basis, as well as how the units respond to external demands.

## **8.2 Structures and approaches**

Despite an observed variation in programmes, including form of delivery, there was a great deal of similarity with regard to the main components of evaluation within their respective Institutional frameworks. Respondents outlined frameworks consisting of a configuration of processes including module evaluations, both formal and informal; student feedback frameworks, use of participant representatives and reference groups; monitoring of programme delivery and progression; regular course team meetings to assimilate feedback and assess implementation progress; involvement of external examiners; annual course reviews; summative programme evaluation; and periodical cross programme evaluations. These clearly reflect the ENQA standards and guidelines referred to in chapter 4.

But while the frameworks generally appear similar, discussion concerning the control over the parts, focus and implementation of them differed. An additional factor of interest that arose through discussion with respondents across the different subunits related to their recognition of the interplay between formal and informal evaluation processes within the structure. The following section contains reflections concerning the evaluation structures and approaches in place across the subunits.

Structure of evaluation models	Related issues (interlinked)
Structures and approaches	Basic structure
	Formality and informality

In ENG2 respondents described how the subunit operates within the wider Institutional framework for evaluation. In addition to evaluations run by the programme team, one of the faculties within the HEI was given responsibility for carrying out the formal survey of students, with some possibilities for the addition of module specific questions from the subunit. This was described by one respondent as following “a standard format”. There was again a common pattern of these results being registered within the system, as well as being discussed within the subunit by those responsible for the different modules as well as those lecturing on them. In addition there was a strong implementation focus to the process evaluations, with the subunit monitoring staff and student output, as well as consideration of how their programme materials functioned.

The annual processes were further assessed by an external examiner, and an annual report was delivered to the Faculty. This was supplemented by periodic programme review. The evaluation model was considered by one respondent to be comprehensive, declaring that:

*There are those who would claim (laughs a little) that because of the amount of quality assurance that we do and because of the quality of our materials people on the whole get a better deal out of us, but then we would say that wouldn't we. [ENG2n].*

The respondents also outlined how evaluation activity took place at course team level, where there was reflection over lecturer feedback and monitoring of student work and feedback. One subunit member described this process as essentially being response driven, and outlined the way the informal processes developed within the course team:

*I suppose we also evaluate as a team, in terms of the response we see, in terms of what [course lecturers] are saying and what students are saying in their written work, any letters or feedback we get directly from students. [ENG2k].*

Within this approach, it was emphasis on monitoring the programme that was perceived to be the strongest part, focused mainly upon ascertaining the standards of the programme content and the teaching given within a comparative approach:

*it's quite a substantial process, yes... We have two obligations, first of all to make sure that the materials are of a standard that is suitable for the people who are working on the programme or the course, and secondly we have to make certain that the quality of the tuition they receive is suitable and commensurable with that which they'd get at other universities. [ENG2k].*

Similar reflections were made at ENG1, but in addition respondents discussed the impact of student feedback within the system; the two main forms being voluntary summative module feedback and student representatives who channelled formative feedback to course team meetings in the subunit. As one respondent noted:

*We've also got a student representative who gathers evaluations from the students and brings it along to termly meetings that we have... and we also have an evaluation discussion group on the web where students can respond to specific questions on things that come up between our formal evaluations. And the other thing that I suppose that we do is, when they put in their draft we actually have a box that asks them if there is any specific comments they'd like to make. [ENG1a].*

Programme participants were encouraged to evaluate the whole course within a year of completion, in order to provide some essence of impact data. Such actions were seen as part of the move towards greater formality and accountability, which had placed the subunit under greater difficulties in terms of reporting. As one respondent recognised, this was not unproblematic for the evaluation structure that the HEI had developed over time, creating complexity with regard to the increasing external demands for codification and impact assessment:

*But in a way we've relied very much on kind of gut feeling and we know that what we offer is valuable and worthwhile because people tell us so. It's that kind of level of intuition and gut feeling rather than, you know, here's some hard data. But government agencies want to see hard data, but it's quite difficult to get that data, when you start to unpack the complexity of the link between a course or an intensive master programme over 2 or 3 years and the quality of leadership. [ENG1g].*

Reference groups were also used in Norway to meet the demands of increasing student involvement in the evaluation of the programmes. Respondents intimated that the underlying purpose of these groups as seen within the subunit was to promote dialogue at programme level, exemplified in this comment:

*As I understand and experience it, dialogue is important, that's why we have established these reference groups, which are organised around the base groups for students such that they have to space to talk together before the appointed representative comes to talk to the reference group which will contribute to the [programme] evaluation. By doing this and making time for it the evaluation becomes more interactive, creating space for dialogue.... [NOR1d].*

In slight contrast to the subunits in England, the subunit in NOR1 devised and implemented surveys that were based on the institutional framework, during and at the end of each module, where the data gathered was presented for reference groups and subsequently the wider course team in order to evaluate implementation and programme content before the data was reported upwards through the system. One respondent reflected over an example of this process, which helped create further reflection within the subunit prior to formally reporting the results:

*We presented what we saw as the tendencies, what was functioning well and in what areas we need to work on, based on this material. We concluded and were agreed on what needed to be prioritised of the areas seen as requiring development. So we ended up with some "headlines" of the areas to work with [during the programme] for improvement. And at the end of the semester we will have a final evaluation which is taken forward to the next round. And that's about everything: content, form and curriculum. [NOR1b].*



In NOR2 there was once again a similar format to the evaluation design across the other subunits. When it came to the formal student survey, similar to the subunits investigated in England this was designed and distributed by a separate unit within the organisation and later discussed within the reference groups and panels. But in the other subunits, reference groups were also used to gain formative feedback from the programme participants, and this had recently been developed to form a panel for the whole programme:

*But we also have evaluations which aren't just based on the [electronic survey] but are based upon discussion with panels, made up of representatives randomly selected from the classes, which bring out more of a group discussion. [NOR2e].*

These programmes, in addition to be a response to demands for student voice, had created greater opportunity for formative discussion between provider and participant to balance overemphasis on summative responses otherwise used in the system. Another respondent considered that these reference groups had been particularly effective with regard to the commissioned programmes run by the subunit. These groups had given opportunity for wider dialogue between the different stakeholders and were seen as important for developing the evaluation system and development of the programme:

*[They are made up of] us, the students and the commissioners, all together. And these are excellent fora; we use them a great deal. They are a learning arena for us and provide good discussions and are thus very important. We use these on all of our programmes, which give us a very solid evaluation system. [NOR2g].*

However, as will be seen in the next chapter, a potential source of conflict was perceived to emerge with regard to the commissioned programmes, as it had been noted that the commissioners, as employers, responded to feedback from the programme participants, their employees, and took this criticism up with the subunit members. Despite potential difficulties, these groups were still considered to enable a better process of discussion.

#### *Formality and informality*

Following on from the previous subsection, another important issue related to structures arising from discussions with the different subunit members concerns the relationship between the wider formal system and the more informal activities. In this section some examples are given from the different subunits. Three example areas were outlined. One area concerned the efficacy of the formal systems in place to provide the required information. Another theme concerned the academics and what information they felt they needed about their programmes. Another example focused upon how academics approached evaluation in their programme delivery, particularly their own assessments of progress and development, what could be described as the formative assessment performed as part of the professional activity of subunit members at the

programme delivery level. Although each focuses slightly on a different situation where the dilemma is raised, the respondents were answering questions concerning their approach to evaluation within the formal system. The responses within this data can also be seen in relation to the section on perceived limitations, outlined in section 8.7 below, and will be further discussed in relation to decision responses in the next chapter.

In response to a question regarding how the information considered necessary within the wider organisational system was gained, respondents noted that the evaluation systems allowed for more informality than might first have been perceived. An example from ENG1 showed how the subunit had adopted a more informal approach that allowed them to collect information beyond the more formal surveys going out to students. A respondent noted that a cross-faculty internal meeting had highlighted the difficulties in gaining responses from participants, and this had been noted particularly, but not exclusively, on distance learning and hybrid courses. The respondent noted that the message coming from the meeting was that:

*it doesn't always have to be formal, the evaluation process isn't always formal, you can have it informally so that students might say "Oh that we was really great". So it's up to us, we need to collate that informal feedback as well and put it into some kind of form. [ENG1h].*

The respondent was later asked if the more formal systems provided the information required within the system and noted that:

*Well no, that's why we've started to try and implement this other way of doing it. Actually it was very useful because we had an internal meeting about a month ago and we were able to have quite a lot of comments from students then, I mean it's not formalised in any way but they know that that is what is happening to the comments. I mean, they're anonymous but they... know that these comments are going forward to a meeting. So in a way it prompts them as well to do it. [ENG1h].*

Respondents also discussed the evaluative needs that they had with regard to the programmes that they run. In ENG2 an example was shared concerning how some subunit members recognised a need to gain some evaluative feedback for themselves regarding a new part of the programme, and had approached it in a more informal way, as it would not necessarily be covered within the structure and approach of the formal system. A meeting was to be set up independently to consider how to answer the wider demands, although the respondent recognised that planning for informality was not always easy in a busy schedule and with a fairly rigid QA system in place:

*this course that I mentioned, [NN – course title], which is this new compulsory course, ... we want to do an evaluation, to get some feedback I should say, we've*

*got a meeting coming up in a couple of weeks' time to talk about it, and we haven't decided yet whether it will be just from the tutors who supervise it or whether we will actually go to the students, but we want to get some feedback, to, because this is to inform next year's [NN – course delivery], [as] we could make minor changes. So we are doing that off our initiative basically, but we're also working with [NN – named person], who ... has got the overall responsibility. We haven't actually planned this, we've just said we've got this meeting coming up, to work out a way of getting some longer term feedback on students' reaction to having this compulsory course. [ENG2p].*

In one sense it seemed therefore that the more formalised system could on occasion help increase the amount of reflection around the programme and problem-solving concerning evaluation. But this was still with the recognition that the growing demands related to impact were often perceived as untenable.

In addition to specific reflections about the structure of evaluation and its efficacy within the formal system, respondents also reflected more generally over their own role as a professional, as was seen above. As one respondent from NOR1 noted:

*So when you talk about evaluation, I've thought that one thing is these reports and the quantitative surveys that we do and the formative evaluations along the way when you implement a programme, but another thing, in a way, is all the assessments you do yourself as a professional actor, with regard to improvement and development. [NOR1f].*

A member of the subunit at NOR2 also referred to this theme, considering the informal evaluation performed as part their professional judgement to be the basis for programme development and assessment of quality rather than any feedback generated from the formal system:

*In relation to the development of the programme so I think it is more that it's part of the process. When you work so closely to the students and where you have them in so many situations where they can develop, you see it. I think it is more in the daily conversations with students. [And there are two of us who supervise] all of those writing Master theses... So I think that all of the adjustments we've made are a result of that type of dialogue rather than as a result of these [organisational] evaluations. I do a lot of evaluation in the form of everyday communication... [NOR2h].*

As was stated at the beginning of this section, briefly outlining the evaluation designs and frameworks provides a backdrop to understanding the responses given by the different subunit members. Relationships between the structures and approaches in current use have been briefly outlined. The issue of formality and informality was also raised, as the subunit members reflected over their own attitudes to evaluation and their practices compared to and within the formal

evaluative frameworks. The examples outlined above inevitably overlap with discussion concerning the decision process that will be outlined in the next chapter. I turn now to look at themes raised by the respondents with regard to design. These themes centre upon how they described the purposes and focus of the implemented designs, considering also designs for programme effects, as well perceived tensions and limitations.

### 8.3 Purpose

Responses from the subunits suggest that the purposes related to evaluation design overlap, while at the same time results are utilised differently at different levels of the organisation. While this is not unrealistic when relating to evaluation theory in this area, perception of purpose and future utilisation appears to be most important to those at the micro level. As was seen in the previous chapter the perceptions of institutional demands from subunit members showed that the main focus appeared to be about systematisation, operative control, throughput and ascertaining student satisfaction. Within-unit demands were considered to be more focused upon improvement of the programme basis and participant learning.

Across the different subunits there was some degree of tension between the formal requirements of the evaluation systems and the processes at programme level. For example respondents spoke of the increased demand for, and formal systematisation of, student feedback and evaluation of the course, while at the subunit level respondents often described their purpose as improvement focus and the possibility for student voice. There was increasing lack of assuredness of what the purpose across the institutional system was and there was considered to be little feedback ensuing from higher levels. Respondents were unsure whether evaluative activity merely provided the data at higher levels to satisfy legitimated or control purposes, in order to meet external demands and criteria, or if there was an interest in improvement, or if this was intentionally left to the micro-level. Discussion frequently centred on whether the purpose of evaluation was focus upon accountability or improvement.

Purpose of Evaluation Model	Related issues (interlinked)
Design Purpose	Accountability focus
	Improvement focus

#### *Accountability and improvement*

As was taken up in Chapter 4 balancing accountability and improvement in evaluation has been seen as an issue of increasing complexity. It has been very evident within debates surrounding quality assurance and accreditation of HEIs, although the focus as noted in Chapter 4 has predominantly been on the external accreditation bodies (Danø & Stensaker, 2007; Vroeijenstijn, 1995; Westerheijden, Stensaker, et al., 2007). As Harvey (2004-8) was shown to have

noted, increasing the formal demands, that is for greater accountability, upon academics can demotivate and create a culture of compliance rather than promoting greater focus on continuous improvement. Thus factors of accountability were considered to be greater control and compliance<sup>175</sup> whereas factors of improvement are more learning focused based on micro-improvement of the programme.

In the subunit at NOR1 respondents spoke about a within group focus upon programme improvement, contributed to by their evaluation models. At the same time there was recognition amongst members that evaluation also has an accountability stream to it. Respondents saw the main purposes of evaluation as ascertaining student learning and developing the programme to facilitate this. Respondents across the subunit considered that the formative evaluations that were implemented internally to be the primary focus of the staff, exemplified by the comment below. In the next chapter it is noted that the subunit members maintained that these were already in place before external demands came, the problem is not necessarily the fact that they have to provide information, but rather the focus of it as well as the format it is to be provided in:

*We have an internal evaluation, and it's that which is important; the continual evaluation that we do with regard to the students. And we have pretty good routines that we have developed for this... It is this that is important for us, that we continually assess the programme, its content, working methods and forms of evaluation related to the goals we want to accomplish. [NOR1k].*

In concert with these goals another respondent summed up the main functions of evaluation of the programme to be about learning:

*I am most focused upon "the learning assessment, that's the keyword for me. And in that is very much that we have a conscious relationship to ourselves, that is, what is actually happening within the study programme. And so we invite the students into metacognitive processes, such that they are conscious of their own learning and how they can utilise it. We capture a lot of this in many different ways. [NOR1b].*

Another respondent went on to reflect how the programme structure was built around this concept and used as part of the learning tools. The purpose of the various design approaches used was to ascertain and respond to the students' perspective at the micro level, but also to gain greater participation in the process, which was also a wider goal of the programme:

*The main purpose is improvement, and this should be based upon experience, where of course the students' experiences are decisive. If we can't manage to present or organise a study programme suited to the group we are teaching then*

---

<sup>175</sup> Harvey sees these as different aspects

*we will never succeed. So I am primarily focused upon increasingly drawing the students' experiences into the evaluation of the study programme. [NOR1i].*

There was a perceived complexity within the organisational system with regard to evaluation, noting that there were many rungs in the hierarchical ladder of the organisation. This implied multiple purposes for and uses of evaluation, where the higher levels appeared more interested in the control side and the improvement focus was left to lower levels:

*So in a way, from the side of the leadership [evaluation] is a tool of control, to ensure a certain level of quality is reached. So it's more the quantitative form they are interested in, whilst the focus on development of the programme and study as whole is the responsibility of the academic group... [NOR1i].*

As another respondent at NOR1 outlined, it was important to link evaluation with learning processes at the micro level, and this the subunit tried to achieve, ideas which were linked up to organisational learning:

*... evaluation is closely linked to organisational learning, which is a goal for the programme. The students shall experience how they can systematically develop their own teaching programmes and by experiencing that the programmes that they are involved in here are evaluated continuously we think gives them a good example how things can be done in practice... [NOR1k].*

This again was linked by another respondent to the concept of “voice” and creating “dialogue”, how this process was allowed to develop:

*So for me it's about encouraging “voice”, both ours and the students', and from that background you can go deeper into the materials... and developing a dialogue and discussion, which is the purpose of evaluation. And this all of time to develop the best conditions for learning on the programme, within the framework we work in. [NOR1d].*

Interestingly in alluding to boundaries and frameworks, the respondent also begins to take up an issue that will be approached in more detail in the next chapter, when it will be considered how demands and interpretations are dealt with, and how subunits describe how they respond to the different pressures that they face.

There was not reported to be the same level of discussion within NOR2 and the subunits in England. As was outlined in the previous section on structure of design, the members of the NOR1 subunit generally expressed having more control over and proximity to the evaluation design process than did the other subunit members.

However, there were still responses concerning the purpose of evaluation. Respondents at NOR2 outlined how the formal evaluations appeared to be very much focused upon the implementation of the programme, rather than on the content. As one respondent noted:

*Evaluation? Yes, it's a bit varied here. That is, in relation to the students experience shows that to a lesser degree there is a discussion concerning themes and content, it's more about implementation: who functions well, what's the administration like, the organisation of the programme and suchlike. [NOR2g].*

At the same time the respondent considered there to be some elements of the central evaluation that aimed to ascertain what learning had taken place upon the programme. However, the additional factors that were part of the centrally implemented student survey were considered to be of little interest and relevance for understanding quality at programme level:

*[the focus of the surveys] is on whether [the programme] has contributed to their learning; well that's how I understand the forms. Whether they have gained any new insight... new understanding, increased their level of reflection, contributed to learning; that's what I'm looking for. I don't look at the rest; I don't really care about it, because if I have those [points], I'm safe. And if they take up these things about organisation and how much they liked the lecturers and all that, well, it's not important... I use evaluation in such a way that it's a measuring point to find out the main tendencies. [NOR2g].*

Part of the reason for this limited value and which is similar to responses from NOR1 and the English subunits is the regard for surveys being part of a general process to satisfy the wider quality assurance system and focused therefore too broadly. In this case the respondent suggested that the methods used followed a standard operating procedure in the organisation, becoming an institutional expectation:

*It's usually the [electronic learning platform], with focus on have the prepared, is the curriculum okay, have they learned something, was the teacher good, Well, there isn't so much about the teacher, a little perhaps, and a little about outcomes. They are not especially, well, they are pretty much just the usual "grind" [the organisation] puts into action. [NOR2h].*

As was outlined earlier, at micro level however, there were also the reference groups in which focus could be placed more specifically on the programme content and organisation, both within the different modules as well as holistically. One respondent recognised that evaluation was directed:

*...not just on each individual module and the "happening" but also how the programme has by and large met their needs; to look even more at usefulness, and discuss alternatives. Are there things in the modules that should be covered*

*on a programme like this...is there unnecessary repetition. Because they discuss the portfolio in great detail... and now we have a joint panel for all our school programmes, such that they can give each other comments for adjustment across the different groups. [NOR2e].*

The final point is interesting, as the level of evaluation is lifted from within group to across groups, with an underlying assumption of some similar and shared experiences that can be evaluated based on programme organisation and implementation. The focus on accountability and improvement here was thus lifted to a comparative level, but only appears to have any impact at micro level.

There were similar responses to NOR2 across both subunits in England with regard to the purpose of the evaluations. In ENG1 one respondent reflected over the purpose of evaluations within the wider organisational structure. While considering the limitations of generalising, the respondent considered how it was not always easy to see the purposes of evaluation beyond the fact that it was done, once again, because it appeared to be an institutionalised expectation:

*From this experience that I have, which is limited to these [postgraduate programmes] around the same subject area, there is an emphasis on evaluation but sometimes, I don't know, it seems to be more because there is a prerequisite to do it. People just haven't got time [and] I think it needs more time. And also thinking about how to tie in the evaluation in with what the purpose of it is and how it's actually going to improve the course for students and people who work on it. I don't know, sometimes you just do evaluation because you're supposed to do it without thinking, well, why? [ENG1h].*

This response had been based on discussions concerning the evaluations directed from above, within the quality assurance framework. This work was tied to particular organisational units set up to deal with QA issues. Another respondent, as was noted earlier when considering operative control, considered that the information required was to provide evidence of activity rather than any focus on programme improvement, and as such was of limited value:

*Well, we have within every faculty we have a kind of quality assurance group and through the University as a whole we have a Quality Assurance / Quality Enhancement Committee, which kind of, you know it's quite a bureaucratic process where they will require of us evidence and some of that evidence is kind of form filling, going through the motions, have you got minutes of meetings and so on, but also sort of student feedback, module evaluations, course evaluations, all of that. But it's at that kind of level; it's not very sophisticated I think. [ENG1g].*

These frameworks created a dilemma compared to subunit members' professionalism. The same respondent admitted that more sophisticated models



were used when evaluating processes in other circumstances outside of the organisation:

*I don't think that the levels of evaluation that we employ are very sophisticated, I think they are very basic. We are better at doing evaluations of others' programmes than we are of doing our own programmes. So I don't think we do, as I said earlier, very much in a sophisticated way, but we are getting better at it; I think it's going to take a bit longer. [ENG1g].*

One respondent from ENG2 further outlined the structure of the evaluation systems across the organisation and how the programme was evaluated, recognising that there was a declared desire to balance accountability and improvement. However, at the same time the respondent declared some uncertainty as to whether the improvement purpose was fulfilled or not, and at what level:

*Well, it's looked at various levels within the organisation, particularly the global student evaluation studies that are done by the [NN – Institute title]. ...it will be looked at, at a course team level by particular course teams, it will be looked at, at Masters Programme level, it will be looked at within the [programme committee] within the faculty to see trends, patterns, what's going well, what's not going so well, whether there's a case for continuing courses, revamping them and so on. And it will also be looked at, at university level, so it's looked at used in different ways, and hopefully used responsively and not just put on the shelf. I was mentioning earlier the different purposes of evaluation, evaluation for accountability and evaluation for improvement, and I mean one hopes that in an HEI like this that is committed to on-going improvement that it is used creatively and constructively and not just used as an accountability tool to say its ok, or it's not ok. [ENG2k].*

The issue of perception of utilisation will be returned to in the next chapter. The next section is linked closely to the responses about accountability and improvement, considering the focus of evaluation activity.

#### **8.4 Focus**

In the previous chapter examples were taken from responses concerning the perceived demands placed upon them. The purpose was to begin to explore how members of subunits saw the framework of demands and pressures for evaluation upon them and how these appeared to interrelate. An important area of discussion in relation to demands concerned at what level evaluations were focused. In Chapter 2 it was seen how one of the growing themes within the field of school leadership training and development has been the greater interest in evaluating programme impact as well as increased discussion concerning the complexities of such approaches. An example was taken from England, where Leithwood and Levin (2004) had been commissioned by the then Department

for Education and Science to account for the possibilities and challenges of evaluating programmes with such ends in mind. As was recognised earlier, the authors suggested that the majority of models applied within this field measured a degree of participant perceptions of satisfaction and self-reflection over learning, rather than ascertaining the effects on the work place. In addition to this many other evaluations of programmes appear to have been based upon deliverer reflections concerning the programme basis and how well it had been implemented.

In relation to these themes two major interlinked issues arose in discussion with respondents, the first regarding the unit of observation and levels of analysis of the evaluation models applied on their programmes and the second considering more widely the concept of evaluating for impact.

Focus of Evaluation Model	Related issues (interlinked)
Design Focus	Participant perception
	Deliverer reflection
	Commissioner framework

#### 8.4.1 Participant perception and deliverer reflection

Members of the subunits were asked to consider the level at which their evaluations were directed. Once again their comments are not isolated from other areas, and discussion took place within a holistic reflection of the evaluation processes. Interestingly, responses across the subunits in this area were linked to members' recognition of the different structures, processes and approaches used. When considering the unit of observation and levels of analysis with the evaluation models on the programmes there was acknowledgment across the subunits that the general focus was based upon students' own reflections concerning the programme content and implementation, as well as an attempt to ascertain what they had learned, and it was at this level that "measurement" was perceived to take place. This is exemplified by a response from one of the subunit members at NOR1:

*You measure in a way, well you can say that which you do at the end is a measurement of students' perception of what they have learned; that's really what you measure... [NOR1f].*

Within the same subunit there was also recognition that evaluations were mostly focused upon ascertaining the degree of user satisfaction. However, while formative evaluations might appear more reminiscent of user surveys, as was noted briefly earlier, subunit members perceived that they did give opportunity for the course team at the micro level to interpret the results for progressive improvement of the course. One respondent related this:

*Otherwise, there are the continual evaluations that take place throughout the year; these are more investigations of student satisfaction, user surveys in a way, but are of course also a starting point for continual revision and change in the studies. [NOR1j].*

As was mentioned earlier the formative approach was highlighted, where participation was encouraged in order to enhance the student surveys:

*So we use the feedback from the students we gain along the way to develop the programme. It's more the qualitative assessments that aren't necessarily captured by a form. [NOR1h].*

While comments like these were not uncommon across the subunits, responses from subunit members in NOR1 suggested additionally that the focus of the evaluation processes was shifting from self-perception of participants concerning their own learning and towards greater focus upon the programme itself. One respondent noted that they had attempted to produce a model that might match these two positions:

*We can in any way see a development, if we look over a wider period of time, from a reflective process, that is that students shall reflect over their own learning processes towards the next step which is that they reflect or can say something about what we do and assess the choices on offer here or the learning arena we create. So the first part is in a way a self-evaluation and the second is an evaluation of what we offer. I think that we are on our way, or it can look as if we are on the way to think about what we offer, what [participants'] ambitions are when joining the programme, such that we connect students' responsibility and our responsibility. I think that's where we'll end up at some point this year, perhaps? [NOR1j].*

In NOR2 focus on the evaluations was seen to be a mixture of participant perception gained through student surveys and panel discussions, and deliverer reflection. As was seen in the previous section, one respondent outlined how the only really important factor of interest from the evaluation was concerned with student perceptions of learning. Another respondent also reflected over using the responses in the student survey as a guide to judge how well the programme functioned. In doing so great weighting was placed on the data based on participant perception on a survey common to the organisation's Master programmes. The respondent considered that this data was adequate enough, and purposively looked for evidence in the students' responses for the degree to which they considered that they were satisfied with the implementation of the module in question. The respondent summed this process up by declaring:

*I look at whether I am meeting their needs, do I meet them with the correct questions, and I can see this in the evaluations. [NOR2f].*

In addition, as was also recognised in the previous section another respondent from the same subunit remarked how their use of on-going formative assessment of programme progression was thought to play the most important part of the evaluation. This was based on everyday observation of programme implementation and discussion with students concerning their learning, rather than the more generalist, formal models run in the organisation.

#### **8.4.2 Commissioner framework**

As will be dealt with more fully in the next chapter the externally commissioned programmes implemented in Norway had created an extra level of evaluation. In addition to using designs similar to those on the internal programme, both NOR1 and NOR2 had formed “steering groups”, which also included representatives from the particular commissioning body for the programme at hand. These groups met to discuss programme development and formed part of the evaluative process. This added an extra level of focus, introducing the perceptions of the commissioners themselves. The commissioner reflections and responses showed great variation and within the evaluation process were perceived often to be mediations of participant responses rather than any specific analysis of their own. However, at the same time this offered an extra level of reaction for subunit members to relate to. I will return to this point in the next chapter with regard to the impact of such responses on the decision process concerning design of evaluation.

This process was described as on occasions leading to a much greater “hands on approach” and formative perspective throughout the evaluation process, especially with regard to programme content and implementation, attempting to ascertain and respond to the participants’ perceptions of programme quality. However, there was also evidence that some commissioners placed less focus on the programme development than might otherwise have been expected. A respondent from NOR2 outlined how the subunit had experienced that some commissioners, due mainly to perceived capacity problems, retreating to accepting measures of participant satisfaction as a basis for measuring programme quality and in doing so relaxed their controlling role:

*So, one can say that [the commissioners] should, based on what they are buying in, attempt to have even greater control. But the owners, or buyers, are busy, they have a lot to do and they haven’t got the capacity to go into this and control in detail. So they measure to a great extent students’ level of satisfaction.*  
[NOR2g].

Underlying these issues is the perception of how effective commissioners’ evaluation designs are. In reflection over this process, another respondent from NOR2 considered how the data from these designs could easily be used instrumentally in an attempt to change the focus of the programme towards their own demands:

*But in addition [the commissioners] have implemented for nearly every group an ex-post evaluation where they go into more detail, what they have learned, what they liked, what they didn't like and they use, perhaps overuse in my opinion, these comments to inform what they want changed in the [overall themes of the programme]. So, it has perhaps become a little too instrumental. [NOR2h].*

However, as has already been noted, the level of competence of commissioners was considered to vary greatly and this moderated the evaluative activity and the level to which assessment was directed. While some of the commissioning bodies were more interested in the general running of the programme, others had expectations concerning the on-going impact on practice. These examples from NOR2 were similar to those related from NOR1. This point will be returned to in the next chapter, when considering how the evaluation design process develops and what decisions are involved in it and how commissioner competence influences this.

## **8.5 Current design limitations and debates**

Respondents across the organisations also raised issue with the difficulties and complexities with the systems under which they operated. Key overriding issues for the respondents here were commonly basic structure, level of participation, analysis and interpretation of data collected, amount of information collected, and reporting format. These issues are presented in the table below and are exemplified by comments raised across the different subunits. The intention is to outline some of the main areas recognised rather than suggest any generalisation across the subunits. Once again, the categories are not considered to be mutually exclusive. They do however explain some of the spread of tensions with the current designs faced by the different groups. It is acknowledged that this data reflects the individuals' experiences of their own evaluation system rather than a combined perspective of a particular subunit, although there was noted to be a general commensurability. The purpose once again of presenting this data is to consider the complexities and challenges associated with the task of making decisions about evaluation models.

Current Evaluation Model	Problematic issues (interlinked)
Design Limitations	Structural problems
	Student characteristics
	Interpretation problems

### **8.5.1 Structural problems**

Respondents outlined different categories of problems with their evaluation designs relating to the structure of evaluations. The structural problematic related from examples within subunit members' responses concern the organisation of evaluations, their general level of sophistication, their frequency

and timing, data generation and the use of resources necessary for implementation.

One member of the subunit at NOR2 considered that the formal evaluations implemented throughout the wider organisation as well as upon the programme, contributed little more than give an overview of general areas that might be improved within the programme. The respondent recognised that this had generally been accepted and highlighted as a weakness, and that the feedback system and dialogue with programme participants needed to be improved:

*It's a weak point, one of the weak points. Evaluation only gives us feedback that we should have seen something better and I suppose that's enough for me. I get the barometer I need, my starting point to do something. But it can't tell me [how to change], well unless of course we go into a dialogue, so that they can help me change it for the better. But then we have to go into a dialogue and we don't have such good traditions for this; we haven't had such good traditions for talking with students. [NOR2g].*

Another issue related to the organisation was related to the comprehensiveness of the evaluation model in relation to the task. As was noted in section 8.5, one respondent, from ENG1, described the evaluations implemented as unsophisticated in relation to the programmes being run and despite the impression of improvement, this issue appeared to signify internal systemic weaknesses, as, at the same time, the respondent perceived there to be competence within the team to have been able to implement something superior to that currently in use. Similar comments had been expressed by members of the subunit in ENG2.

In NOR1 one respondent reflected further that there was felt to be a basic weakness with the evaluation structure, especially concerning the end of programme evaluation. While recognising a difference of opinion amongst subunit members, the respondent considered that the current structure did not support the underlying intention of continual improvement within the team. The challenge raised by the student surveys was the type of data that they constructed, which was perceived to be purposeless, as well as the engagement of those filling them out and the problematic this created of how to use the data with future groups in mind. The respondent recognises that this type of evaluative activity becomes somewhat symbolic, lacking an utilisation purpose, despite giving the perception of scale of response:

*And it's my opinion that, the type of evaluations implemented at the end of a study programme... don't have a great deal of value. That's my opinion. Because, my task as the leader of the programme is, in away all the time, to capture what's happening. But it is also such that I know something is going to happen, I know that some are going to find things difficult, and I know that when we are going to have IT-based supervision it isn't going to go well because they*

*can't use the tools. So when you are talking to me now I can see less value in these larger evaluations than many of my colleagues here. I think they are more like "window dressing"... And perhaps it's based on my experiences that show that they don't have so much to say, it doesn't mean anything, and doesn't lead to anything very much. [NOR1e].*

The structure, frequency and timing of evaluation were noted to be especially difficult within this type of postgraduate programme, where the participants are part time. This raised questions as to what was being evaluated and to what degree. At the same time the efficacy of the evaluations comes into question. A respondent from NOR1 reflected that it often felt like a new evaluation appeared to be implemented before there was any opportunity to reflect over the results from the previous one. This challenged the possibilities for formative improvement focus and creating a situation, not unlike the idea of performance paradox (van Thiel & Leeuw, 2002), where the frequency of evaluation and impression of quality assurance worked against the purpose:

*So it is valuable, but I think it can quickly become too much; you don't manage to have any lectures before you have to evaluate. I think that if you are going to evaluate in that way, you have to have something to evaluate. Our students are only here once a month, so we don't have so many gatherings before the electronic evaluation is sent out, and so we start with the next thing pretty quickly, before we even see the results. So I think that there can quickly become too much evaluation. [NOR1c].*

This point also leads into the next sub point concerning participant characteristics, which are a challenge to evaluative activity.

### **8.5.2 Participant characteristics**

Respondents discussed the problematic issues with relation to evaluation design with regard to student characteristics. These characteristics were partly recognised to be specifically related to the profession that programme participants are drawn from, as well as more generally related to issues of gaining feedback. Another interesting issue that was raised concerned the institutional preferences of students. There was some question of whether parts of the current evaluation design were appropriate for the student body enrolled on the programmes.

An example of the general difficulty with gaining feedback was raised by a subunit member from ENG1, who considered that issues of workload might be contributing to the problem, but which was also later recognised to be a general one across the institution as a whole:

*We've been getting a very small rate of response. So this year we thought we would try and actually ask them to do it slightly earlier along with their*

*assignments. I think it's a course wide problem, this thing of getting them, so it's just another thing for them to do really. [ENG1h].*

The respondent also went on to discuss how many of the forthcoming issues related to the students perceptions of how well the programme was implemented with a great deal of focus on practical issues, many of them recurrent problems. Within NOR1 there was also recognition that participants, especially in the initial stages of their study, generally raised issues concerning the structure of the programme rather than the content and basic premise of it:

*But now I think that we've only come up with small things ... because they are so pleased with so much, so it is more like they want different rooms, better breaks, more structural things. It's just like when I was a teacher, I was so fed up of parents who were only concerned with whether the pupils had chewing gum or not, but never asked me about the content of my teaching. Please come and ask me about my teaching. It's also that they are quite fresh, first year students at Master level, so maybe they can't be so critical about the content. But we do get a little of it now and again... [NOR1h].*

An interesting interjection came during an interview with a subunit member in ENG2 which straddled the three major thematic categories in this section, student characteristics, structural and interpretation problems. The respondent recognised that the nature of the way some participants progressed through the course and the way it was organised could lead to some degree of fragmentation. This had created a situation where participants did not really have any opportunity for overall programme assessment within the current evaluation model. The subunit had commissioned extra research to investigate the participant attitudes to the programme but the respondent recognised this was complicated by the limitations of understanding the students' relationship to and identification with the wider organisation. The complexity within the task environment was problematic in both collecting the data but also interpreting it when collected. This contributed to a recursive difficulty in gaining a holistic overview of the programme:

*I mean in terms of that they get the opportunity to evaluate courses, they don't really get the opportunity to evaluate the whole programme as such. And that's what we were trying to find out through some of this ... research we did, but of course it always falls into, you've got people who like what they know, and in fact, in the main, people, if you're well into studying for an award, you like what you know... you put up with what you know, because otherwise... you'd have got out earlier. From the organisation's / our perspective, I do think it makes a difference. We don't know what students think about the full programme. [ENG2p].*

This also then highlights the problematic issue of interpretation of evaluation findings which is considered in the next sub-section.



### 8.5.3 Interpretation problems

As was seen in chapter 3, issues of interpretation related to evaluation models are considered by many theorists to be generally complex. Particular areas of complexity referred to by respondents in this study included uncertainty within the model as to the purpose of the evaluation, problems regarding representativeness of feedback within both formative and summative models, applying both qualitative and quantitative designs. Both structural and student characteristics are once again noted to overlap in these responses.

A response from a subunit member in NOR1 exemplified the perceived difficulty in interpreting evaluation data due to the lack of understanding as to the purpose of evaluation. As was also noted by others, the respondent discussed how the workload situation might explain a lack of feedback, but also recognised that the inconsistencies in understanding the purpose created difficulties in interpreting the answers given:

*Well it doesn't take very long to tick the boxes on forms like these, and [there were also] additional comments, but not always. There are spaces available to leave comments. But what I miss is focus upon the purpose of the feedback; is it, in a way, there to enable them to express their dissatisfaction or satisfaction or is it to contribute to improvement of the programme? We could have done a better job with the form, but otherwise I think the feedback was serious. But they are busy people, they're school leaders and teachers with hectic lives, working full-time and studying part-time. So, sure it's hectic and it's not certain that they prioritise this. And what does it really mean from time to time to have a low response rate? It might mean that they are satisfied? We couldn't measure that. And those that answered were perhaps those who were most dissatisfied? I don't know... [NOR1g].*

The use of reference groups could be equally problematic. While their introduction had been seen by many as way to increase the amount of feedback, in both qualitative and quantitative terms, a respondent from the subunit in NOR1 related how the course team had wondered to what degree the comments made reflected collective issues for programme participants. The respondent considered the issue of how the subunit members tried to interpret the representativeness of the reactions that had been raised by the student representatives:

*We get the representatives from the groups to come and give us feedback on different issues, but we cannot always tell to what degree this is their own personal opinion or whether it is the result of discussion from within the groups. Therefore we just have to consider how much weight we are going to place on the comments and what we need to respond to. [NOR1c].*

And these meeting points were further noted by other respondents to still be plagued by problems of low response rates and students not turning up to meetings.

Another issue raised concerns about information overload, which also created problems for subsequent interpretation. With the increase in volume of evaluation activity related to the programme along with requirements for both quantitative and qualitative data, the production of information made the process of interpretation more complex, raising questions about what kinds of design would be applicable and practically relevant to implement:

*Another problem... is the amount of information collected in... because traditionally an educational programme is built up with plans, a selection of literature, and some criteria, perhaps from the authorities, for it should all be about, and that's it. But what happens when you begin to evaluate, and not just once but many times, as well as evaluating the surrounding systems that affect the programme, is that you begin to experience that the volume of information is so great that it becomes contingent, and perhaps even random what you focus on, that doesn't have a controlling effect. Therefore one needs to create a balance between the amount of information you collect and how it can be used constructively. [NOR1i].*

Interestingly all of the responses in this subsection so far were generated from the subunit in NOR1. As will be seen in the next section, this group of respondents had more control over the development of evaluation designs used upon their programmes, even though they were, like the others, still required to work within institutional frameworks. It is interesting to note though how they reflected over design weaknesses, as this will be revisited in the next chapter.

Before leaving this section a final comment is given to a respondent from ENG1 who when discussing the surveys, that were designed and circulated from a central unit, agreed with the perception that it was hard to interpret the questions and understand what the formulation might mean. In commenting on the content of one of the evaluation forms the respondent sympathised with the general problem of filling out such surveys as well as the particular points that were hard to define on the current model:

*I think I'd prefer to give comments or feedback on things because sometimes it seems a bit meaningless and you're not quite sure; there's no space to ask questions is there. Like this one, administrative support here could mean anything; there are so many different aspects so you think, well, what am I actually judging here? [ENG1h].*

These perceived challenges and weaknesses had not been made any less complex by the increasing demands for formalisation and systematisation, as outlined in the previous chapter. Growing alongside these demands that were

primarily related to more general HEI reform, were those linked more closely to the subject field of school leadership and the increased interest in and requirement for demonstration of programme impact and the effects of study.

## **8.6 Designs for programme effects**

In this section focus is placed upon how respondents perceived the concepts of effects and impact in relation to their evaluation designs. In the previous sections of this chapter more general attention has been given to the purpose of designs used and the level at which they focused. The issue of ascertaining programme impact was raised in the previous chapter as a perceived policy demand. In England respondents across both subunits had noted how funding of programmes was commonly becoming more dependent upon demonstrating in some way impact upon practice, attempting to link programme output with outcomes within programme participants' workplaces. This was considered to be a measurement of value for money. In Norway, despite there being little to no central pressure of this kind, the commissioning bodies "buying in" the services of programme providers varied in their demands and the expectations for the participants 'sent' for training and development, some moving towards impact focused demands.

Although, once again, the data in this section is considered to be related to previous and subsequent sections, there were four major, overlapping themes raised by respondents with regard to consideration of evaluating for programme effects. The first regarded discussion of the way that effects of study were ascertained. The second regarded discussion about issues raised concerning designs and models when considering evaluating for effects. Part of this theme has already been dealt with in the previous section that deals with perceived design limitations more generally. The two remaining themes, focusing on the within unit discussion related to the issue of effects and impact, and their responses to perceived demands, will be dealt with in the next chapter which focuses upon the decision processes associated with the internal models chosen.

### **8.6.1 Ascertaining the effects of study**

This issue relates very strongly to the previous section concerned with understanding respondents' perceptions of the focus of the evaluations associated with their programmes. The concept of effects was defined in very general terms, as respondents were asked to reflect over how their current evaluation models sought to ascertain any effects of the study. These discussions were framed by the wider debate taking place in the field of educational leadership and management about impact and improved pupil outcomes, outlined briefly in Chapter 2, which each respondent acknowledged.

This section is divided up into subthemes drawn from the data, outlined in the table below:

Evaluative Purpose	Mechanism employed
Ascertaining Programme Effects	Programme activity
	Post-programme reflection
	Investigating improved pupil results

*Ascertaining effects through programme activity*

Across the subunits respondents outlined how micro-level impact measures were incorporated into the tasks and activities that were part of the programme reflecting upon within programme impact. This reflective activity was often intended to help participants to consider their own learning as well as any impact they perceived it might have on their practice and the practice of others in their organisations. Therefore, in this setting and associated with this mechanism, impact was commonly defined as self-perception of learning and change in practice. An example is drawn from ENG2, where one respondent outlined how this reflection as part of one of the compulsory examined tasks was used to assess participants' understanding of their development as well as providing the programme deliverers with data useful for gauging what impact the programme was generally having. Interestingly, then, this evaluative activity was built into the course work:

*we have a section in the... report, that they're asked to do for [part of their examined work], which explicitly asks them about the impact on their thinking and practice of doing the course and whether it's had an impact on their organisation and colleagues. I think it's useful to help students to reflect on how they are able to apply course ideas and it's also useful for us to be able to see if it's having an impact, and if so in what ways. [ENG2k].*

Another respondent from the same subunit recognised that these activities were more specifically focused upon change in practice as an outcome of their work, rather than just explanation of what research they had undertaken. At the same time the respondent recognised that generally there is little within the evaluation procedures that approach the topic of impact, although there was some focus within the central surveys. The respondent also reiterated how such data will only be forthcoming from the students themselves:

*First of all it's very difficult and I don't know that in terms of our evaluation of the course we do anything very much to try and ask about impact. That said, the [NN – Institute] surveys do ask for information from the students, and it's ultimately only the students who will give feedback and information on that; they are the only ones who can make that judgement. But, both [NN – course title] and [NN – course title] include within the projects an expectation that there will be an evaluation of the difference the project has made. In [NN – course title] it's very explicit, it's not a research report, it's a report on a management activity that they've undertaken and part of what they are asked to do is to say what difference it appears to have made to practice. [ENG2m].*

A respondent from ENG1 also showed how impact data was interpreted through programme activity, investigating perception of changed thinking generally as well as related to personal practice:

*But, what you can see is the impact in terms of their thinking about issues and in terms of their thinking about their own professional practice. And I guess that is the kind of impact that I am looking for when I was talking about the dissertation etc. [ENG1a].*

A similar response was made by a member of the subunit at NOR1, who also emphasised conceptual and practical difficulties with measuring any kind of effects of the programme. As is observed here, the respondent refers to one assessment of impact as the ability to connect learned theory with practice in one's own workplace:

*And what we think is that we can't really measure [impact], but we can see it through the programme activities. There is a document where they have to apply this competence, where we can see if they are able to choose practice situations and analyse them by applying theory and debate the issues. So we can actually gain a great deal of insight, but then we would really have to do documentary analysis and we haven't really done that systematically. But it is something we meet in our supervision, through the papers we receive, and we can follow it up in the lectures. Most of it is about the way they talk about things and the way they apply theory [to their own practice]. [NOR1c].*

Attention was also further drawn to these processes in the evaluation and how the practical tasks were used to ground the theory and participants were asked to reflect upon the ultimate practical relevance in their own workplace.

*So it's clear that we're focused upon how students apply their knowledge to their own practice, because our students have a job and our educational activities will always indirectly be related to what is useful knowledge and what has practical relevance. And this is done through the compulsory demands, where the students themselves reflect the relevance of what they [are learning]. So, on our evaluation survey there are questions that are focused, amongst other things, on usefulness for a leader in the workplace. So, in a way it's about results, well not results as such, but relevance of what we do in relation to the needs of the schooling system. This is important. [NOR1i].*

Interestingly this also applied to the externally commissioned programmes. One respondent reflected upon how all of the programme tasks were specifically connected to "obligatory development work at each school" that the participants had been drawn from. In this particular commissioned programme all of the participants were from one educational authority, and as the respondent

recognises the tasks were additionally supervised by representatives from the commissioner who also observed the activity at the workplace:

*So there was a compulsory task to write a report, or outline cultural and structural traits in the organisation, present it for the personnel and gain feedback, such that there were a lot of processes among the staff. [We supervised them about] the content of the documents, while they had a supervisory team from the regional authority who were at their schools and supervised the processes, giving them feedback related to how it was experienced amongst the staff and challenging them a bit. It was an expressed goal that the development work would lead to changed practice at each represented school. [NOR1f].*

This respondent went on to emphasise that there was however neither demand nor purpose nor possibility to consider impact on pupils in this project:

*But to able to say, or rather, to begin to test out [the effects] in relation to the pupils... of course that is where one would like to see a difference, in better results, but it wouldn't be possible in this type of project, and neither was it our goal to do it. No... [NOR1f].*

The desired focus of the evaluations more generally within the subunit at NOR1 was greater understanding of participant learning. While quantitative surveys were developed these did not reveal the types of 'effects' respondents were most interested in. The respondent also spoke of the importance of the "individual reflection reports" produced by the programme participants with regard to their perceptions of their own learning, considering that these provided a better basis for evaluation than the quantitative surveys which just measured a response to particular competence goals. There was, then, a clear emphasis on developing evaluative mechanisms in relation to programme activity. And this was reflected to some extent across the different subunits, where evaluation was more tied to what came out of the students' work and submissions.

#### *Ascertaining effects through post programme self- reflection of participants*

Building on from the use of programme activity another example concerning the essence of impact was attempting to ascertain effects through post programme reflection. It was a common idea across the subunits, but as will be seen in the next chapter some groups had more fully implemented this idea, while others were still talking about and developing their response. One example was given from the subunit at NOR2:

*And we have evaluations that are implemented when everything is completed, about six months later, where the students and maybe the commissioner if it's a commissioned activity. These discuss, what happened really, what was the*

*outcome, not just the output. We don't just discuss the delivery, or output, but also the outcomes; did this produce any lasting change. [NOR2e].*

Another respondent from NOR2 had experienced a similar evaluation related to the programme, but one that was undertaken by one of the commissioning bodies upon their own employees. The commissioner had arranged a presentation of those participants of whom had completed a Master degree at least six months previously, attempting to focus on effects of the programme within the local authority:

*And it was an attempt to what they had learned and whether it had had an effect. And it was quite impressive actually, because I don't think there was one negative word. Everyone said that they had learned something; everyone said that they used it, and everyone could point to specific results within their own daily practice. So it is visible. [The commissioner] has arranged such days before, with follow up to see what has been the effect of the leaders. [NOR2h].*

This was also referred to by another respondent who outlined how the commissioner had attempted to investigate further the impact on the workplace by asking participants to reflect over how colleagues had experienced any change and gained any benefit.

*[The commissioner] asks them if they have become a better leader, as well as if their colleagues think they have become a better leader; whether they have got any benefit from the leader completing the course. On this point they are a little more modest, they are, after all, Norwegians, so they don't necessarily believe that everything is going ok. But they do give interesting answers; they think that there are visible results, that is, effects on the organisation they work in. [NOR2e].*

As we have already seen, the commissioner in this instance had a clear motive in what impact they expected from the programme, the employability of candidates and the improvement of school outcomes. Despite recognising some wider cultural barriers to the asking of such questions, the respondent considered that such investigation added legitimacy to the subunit's approach to the programme by offering them an evaluation at a level that was not normally undertaken.

In the subunit at ENG1 there was recognition that on some of the programmes run by the course team they had attempted to introduce a more longitudinal reflection to their programme evaluations, and as will be seen in the next chapter it had also become a discussion point for the other programmes:

*We're one of the few programmes across the University where we've had this 6 or 9 months evaluation after they've finished. But it's a kind of voluntary activity; I don't think we're obliged to do that. [ENG1g].*

### *Investigating improved pupil results*

This area of investigation of improved pupil results did not receive much attention with regard to the models currently implemented on the programmes studied, however as will be seen in the next section and also in the next chapter there was a great deal of discussion concerning the topic. However, one point of interest came from one of the commissioning bodies in Norway which as was outlined above was attempting to gain some understanding of what effects, if any, their investment in the programme was yielding as far improved results. One respondent noted that the commissioner intended to combine a number of data sets:

*So the latest thing that they are trying is to take their measure for reading, maths and the national tests and all that and trying to see if these schools where the head and possibly others have taken a Masters have produced better results. So they want to measure, I think they want to try it from next spring... two to two and half years since [they completed the programme]. [NOR2h].*

#### **8.6.2 Problems with designs for the study of effects and impact**

A widespread issue for the subunit members concerned the problems of ascertaining programme impact through evaluation. As has already been seen the major focus was in assessment of programme activity, where it was attempted to build impact related processes based on participant self-perception. As was noted above when discussing designs for programme impact respondents spoke more commonly of the problems in evaluating to discover effects. The issues arising in the data are outlined below, and as is seen from the answers these were generally perceived to be overlapping and interlinked in respondents' comments. Examples are therefore taken for the different points, which at the same time interrelate. These have not been split up in order to avoid spoiling the flow, decontextualizing or misrepresenting respondents' reflections.

Evaluative Purpose	Interlinked issues
Problems with effects design	Cause and effect
	Level of observation
	Complexity of programme purpose
	Time Frame
	Organisational constraint (endemic)

There was a widespread reflection over the difficulty of ascertaining the relationship between a particular programme and resultant impact upon the organisation a participant might be working in. It is recognised that this discussion builds upon the on-going debate that was outlined in Chapter 2 with regard to impact of school leadership upon pupil outcomes. While there was more disagreement both within and across the subunits with reference to that



debate, there was a clearer agreement with regard to the challenges limiting the evaluation of programme impact.

#### *Cause and effect*

Within the discussion two major arguments arose concerning the problem of ascertaining cause and effect in such evaluation designs. The first was with regard to demarcating the independent and dependent variables in such a study. This was exemplified in one of the comments from the respondents in NOR1 subunit, where part of the basis of the argument is the perspective on leadership that the programme is grounded within:

*[It's a problem] to think of it as a linear cause-effect relationship; I mean, what is the cause and what is the effect? If there is one thing that has problematized within the research field on school leadership then it is that attempting such cause-effect studies is extremely problematic, and often only leads to new questions, just because what the effect is and what the cause is so difficult to ascertain in most cases. And someone can be an excellent leader at one school and have no effect at another, because school leadership is relational. It's about what kinds of relationships you manage to create, both upwards and downwards in the system as well as relationships with the local community. It becomes so complicated that a simple cause and effect model based on a school leader programme; well I think that's particularly naive to establish [such an idea]. [NOR1k].*

Not only was the identification of the independent and dependent variables considered difficult, but also the problem of intervening variables as well as the ability to delimit other relationships was thought problematic. This was exemplified in a response from NOR2:

*I think there is a pretty long way from the independent to dependent variable. There are two different things. One is that when you talk about effects of executive programmes and leadership programmes, then you are really talking about effects of isolated, much shorter programmes, which I still believe are incredibly difficult to measure the initial effects of at organisational level. But here there is talk about effects of a [Master] programme, which might have some effects, as they themselves say, they feel, but they don't know... I think its very difficult to be sure that you are not measuring spurious connections. [NOR2h].*

#### *Level of observation*

Linked to the problems of ascertaining cause and effect was that of level of observation. It was highlighted across the subunits that the majority of models implemented on the programmes were based on student self-perception and reflection, and while as the examples below show, there were opportunities to

interpret these responses the same limitations applied to the extent of the data gathered. One respondent from ENG2 recognised that it was difficult to ascertain the context participants were working in and make claims about change, but the reflections they made suggested some evidence of changed practice in the workplace:

*We have a student working in an organisation; what would be really helpful would be to get some kind of semblance of what other people working with that student see them having got out of it. I think that the reflection that we ask for is very important, but the nature of the task and the nature of any kind of study is such that you don't know whether or not those reflections upon practice actually reflect the practice that was undertaken, that they had gone through, or indeed represent their true reflections upon it. I think one of the things which makes me fairly confident that perhaps they do, is that you do get in the reflection of the reflective bits that you read, you do get quite a lot of people talking about, quite a lot of writing where people are talking about the things that didn't quite work, the things that they would have liked to have done better, they things that they would do differently next time. And that's very reassuring, because actually it would suggest that you are getting people to think quite carefully about their practice. [ENG2m].*

And these comments were reiterated by a respondent from NOR1 who considered that the focus was only related to the individual's experience of usefulness and that the models implemented reflected this:

*It is the subjective, individual's experience of the effect of the programme, that is what is being measured; if, that is, you can measure what [people] think the outcome is. The questions are formulated, as far as I remember and can see... "how was it for you?", that is "what's in it for you" – so that's the effect you really want to measure. And then we know that it's about how they have experienced the programme, [whether] it was relevant, were there relevant themes, a good way of working, were there too many lectures, too many submissions, and that sort of thing. [NOR1g].*

#### *Time frame*

Linked to the issue of level of observation was that of time frame. Respondents spoke about the effects of the time frame they were placed under when expected to report on programme impact. One respondent from ENG1 considered how they faced a dual problem with regard to models available and the framework for expectations of results:

*Well, you know, there are two levels of constraint really. The first is that ... the technology for actually evaluating impact are not desperately well developed on the one hand, and secondly, well if you are trying to look at the impact of a university course you don't want to wait for 5 years until you find out what your*

*students are actually doing. I think there's a real problem with getting authentic forms of evaluation on these types of programmes. [ENG1b].*

And this point was reiterated at NOR1 concerning the type of effect that one might be interested in measuring, namely pupil results and how long it might take before changes were visible:

*It's difficult for a study programme where the effects can't be measured before some years have passed. And whether results can be ascribed to a programme or related to practice is very complex. Ultimately it's just how the students perform in the exams, how they write their thesis; it's these types of results that are decisive for us - unless of course you follow students out into the practice field, but that is a different kind of study than is being asked for. [NOR1i].*

### *Complexity*

The issue of time frame was also combined further with the general complexity of such a programme as compared to other types of training with clearer, measurable goals. A respondent at ENG1 succinctly described the problems related to where and when the evaluation should take place and how varied this would be from participant to participant:

*when you are talking about programmes, leadership development programmes, the complexity of it is at a different level, and it's not only the complexity but also the time frame as well. You go on a training course to learn how to use Word and the outcomes will be obvious after the end of the training course, but with a master programme, one of the interesting things about those post course evaluations is, quite often people say it's only six months down the line that I now appreciate, you know, something or another [ENG1g].*

The question then arises as to what parts of a programme should be evaluated, and how demands to provide certain kinds of evidence will impact on the rest of the programme and its basic premise and values, as exemplified in this response from NOR1:

*It's very difficult to know what the education leads to, and how to measure this, because how do you measure the development of reflection? How can you measure development of fields of knowledge; that you develop competence in networking? There are so many "side effects" from school leader development programmes that are difficult to measure, at least quantitatively. But, it could however be interesting to see whether those who have completed our programme become better leaders, but then we would have to decide what that meant; what is better leadership in relation to what? [NOR1h].*

*Organisational constraint: endemic to HEIs?*

There was also reflection concerning HEIs and evaluation, inclusive of the problematic of gaining impact data. In discussion with one member from the subunit at ENG1 the topic arose concerning a current research project into school leadership and the impact on pupil outcomes and whether any results might change the modes of evaluation on postgraduate programmes within HEIs. The respondent considered that while results of the research would change the knowledge base that programme providers draw upon, HEI evaluations would be unlikely to change. It was considered to be more of an issue of organisational constraints, endemic to HEIs, rather than any lack of knowledge:

*...universities have a particular rhythm to the work that they do... So yes, the research will have impact because it will be demonstrated through a major research project that there is a connection between leadership and student learning, but all it does is to serve to confirm common sense and knowledge in the first place. Now, I think the project will have an impact because of that and will take knowledge a little bit further, but really it only confirms what has already been known. Now the impact of that on national programmes will, I think, be that the knowledge base of Master programmes will be changed, but whether it will impact on the evaluation of master programmes I doubt very much. I still think that there has to be a much wider recognition across universities as a whole that perhaps the forms of evaluation they take are inadequate to forms of learning that they want to produce. It is a university wide issue rather than one related to educational leadership programmes. [ENG1b].*

## **8.7 Summary**

In this chapter I have outlined reflections of subunit members over the evaluation designs that were implemented upon their programmes. This was in response to Stufflebeam's (1983) research into evaluation models and designs, where focus was also placed upon choice related to evaluation practices and traditions, expectations and experiences. As part of this discussion respondents were encouraged to consider their attitudes to evaluation. There was a notable overlap with the responses discussed in chapter 7 with regard to balancing internal and external demands within evaluation models. There was a notable tension between the models proposed and the data demanded by groups within the task environment and the concepts and attitudes to evaluation within the subunits, where the role of professional judgement was felt to be under challenge. The subunits favoured a methodologically varied approach to designing and implementing evaluative models, which they felt to be under attack and the varied processes under consideration were becoming confused. While an internal drive to evaluate was evident across all the subunits, it was expressed most clearly in the responses within NOR1. This drive for internal development was connected to themes thought relevant within the field of practice of educational leadership. This influence was an interesting connection for the subunit members.

Respondents also discussed how the models of evaluation were related to the quality assurance systems within their wider Institutions. There was a tendency to see these formal processes as convergent across HEIs, and members expressed more interest in their own more informal evaluations and professional judgements. This form of “academic anchoring” was seen to provide the groups with an internal legitimacy. However, they also expressed how external demands and frameworks intertwining with institutional processes were beginning to overtake them. The groups were generally open to many of the new demands, including that of greater student voice and focus upon teaching and learning quality, but there was a general scepticism to the measurement models and understanding of causality that had been introduced, perceived to be linked to accountability demands rather than being improvement focused. The impact on decision making about evaluation is outlined more thoroughly in the next chapter. This reflection over the purposes of evaluation appeared to have sharpened the focus of the group members, especially as the programme participants were experiencing similar demands upon their own workplaces. The groups were also sceptical to approaches that had been developed with regard to ascertaining effects of leadership development and training on pupil improvement. One major problem was considered to be the endemic difficulty of linking HEI programmes to the practice field, and therefore accounting for programme impact.

It was once again interesting to have the Norwegian cases and their relationship to commissioning bodies in the form of regional and local authorities. The subunits in Norway experienced in the same way their involvement to be accountability focused, but recognised also that this was directed towards more local processes. As a result it appears that they were more able to defend their own approach to evaluation, often having greater competence than those commissioning their programmes. In addition, it seems probable that mandators would more readily identify with the programmes of the institutions they had commissioned.

In the next chapter I outline in more depth at the decision processes within the subunits. However, in the table below I outline a simplified form of the perceived tensions between the focus on evaluation designs within the subunits and the external demands and frameworks they must address.

**Table 12: Tensions related to evaluation designs**

<b>External demands and frameworks</b>	<b>Tensions</b>	<b>Internal focus</b>
Formal evaluation designs		Informal evaluation processes
Hard data		Academic anchoring
Participant satisfaction		Programme development
Accountability		Improvement
Summative feedback		Formative feedback
Quantitative data		Qualitative responses
Institutional Systematisation		Subunit systematisation
Matching QA demands		Satisfying programme goals

## 9. Evaluation decision making

There are demands the subunit as programme providers have addressed, the designs are in place, but a question still remains: “how then have the decisions been made?” Such a question is focused upon understanding who have been responsible and what form the decisions have taken. These questions are the theme of this chapter, related to the categories decision making and decision makers, which as described in chapter 5 are derived and developed from Stufflebeam et al.’s (1971) evaluation problems. The data dealing with involvement in decision making will be further analysed in the next chapter with regard to decision making models that might be used to interpret these processes at hand.

Without attempting to oversimplify, it has already been seen that there are relatively similar demands on the subunits in this study by the very nature of being a part of an HEI during the implementation of quality assurance policy allied to the Bologna agreement. As was seen in chapter 4, these demands stem widely from the impact of the Bologna declaration and subsequent activity in the development of quality assurance systems. These demands have led to similar frameworks for evaluation and the information required from them, but as was observed in the previous chapter there are notable differences in relation to the way that the subunits looked at evaluation at the micro-level. Additionally, within the field under study there has been a notable increase in demand for impact evaluation, on a national level in England whilst more varied within Norway due to the different organisation of school leadership training and development and the role of commissioning bodies. The perceptions of these processes are outlined within this chapter. However, the chapter does not claim to account for all the differences and variations within and between groups, but rather attempt to explore, illuminate and understand the processes taking place within each organisational subunit as they decide how to respond to the demands placed upon them and how to arrange and implement their evaluations and related activity.

Additionally, it is important this chapter should not be thought to occur independently as a linear response to varying demands. As was outlined in Chapter 5, the particular phases, or rather “elements”, of the evaluation process are considered to be intertwined and recursive (Dahler-Larsen, 2004a; Dornbusch & Scott, 1975). Therefore it is noticed that the previous chapters consist of elements of the decision processes both on the individual level and also more widely across the sub-units under study. This section aims, however to deal more specifically with the decision making processes building upon perceptions of within sub-unit behaviour and responses to demands and frameworks for designs. The focus here is on what subunit members perceive to take place. The main areas considered from the data concern the roles, processes and experiences of evaluation. As I outlined above, understanding these

processes was part of the focus of Stufflebeam et al. (1971) and their description of problematic areas to be understood. In this chapter the main focus is on the two elements, that I have reworked, focused upon decision makers and decision making. While the other areas of definition, demands and designs have been outlined in the previous two chapters, the themes will also be revisited here as programme providers discuss their experiences.

This chapter also focuses more deeply upon roles. This exploration of involvement in decision making at the subunit level builds on the work of Hardy et al (1983) and Hardy (1990b) outlined in chapter 5. Additionally the background for the study appears to have developed in a similar way to that of Tourmen (2009), who studied decision making regarding programme evaluation from a situated perspective, exploring “pragmatic knowledge” of practitioners through analysis of their practice. Tourmen observed a difference between experienced practitioners and beginners with regard to decisions about design. The experienced practitioners were observed to be more politically aware, appearing to focus pragmatically on “feasibility” and “legitimacy” with particular regard for three main dimensions; the evaluative object; the means to conduct; and the strategies to use (Tourmen, 2009: 28). This was based on their “conceptualizations” based on theoretical understating of evaluation and their own experience. As a result they often adopted descriptive theory approaches. Beginners, however, focused more on the specific frameworks commissioners set out and how to respond to them by applying more prescriptive evaluation theory. In this study the respondents could all be considered similar to Tourmen’s category of experienced practitioners.

In line with the areas drawn from the literature review summarised above, the interview guide was developed to include questions and prompts concerning these processes. The themes covered in this part were among those considered more abstract by the respondents, which was expected to be the case as they are more latently related to perceptions of decision making action. However, as the ways in which the members had implemented their evaluations was discussed, the respondents began to reflect more deeply over the processes that they had experienced. During data collection three major areas were outlined in the discussion, the roles of the involved, the processes of decision making and responses taken. I begin to highlight the major reflections related to roles within these processes.

### **9.1 Role issues**

In this section the different roles that influence decision making are outlined and the issues that arise that impact decision making. Data is drawn from across the subunits as programme providers were discussed both directly and indirectly different actors thought to be involved in decisions about evaluation models. Wider demands upon the decision process were considered in chapter 7. The most influential roles are outlined in the table below:



Decision making	Related issues (interlinked)
Role issues	Academic staff
	Administrative staff
	Programme participants
	Wider organisational roles
	Commissioners

### 9.1.1 The collective role of academic staff

Across the subunits members reflected over their own and others' roles within the academic team related to decision making concerning evaluation. These reflections were related to the specific roles and responsibilities had for decision making, but also to their perception of the working relationships between staff members and how this influenced processes.

Role issues	Role	Related issues (interlinked)
Role issues	Academic staff	Level of cooperation
		Degree of socialisation
		Evaluative competence

In NOR1 group members discussed how the structure of the programme team from its inception had contributed to a clear working relationship between staff members as well as formal responsibilities for the participants. There were many examples of from members in terms of development focused evaluation and collegial discussion. There was a pronounced description of discussion concerning how evaluation models meet internal requirements but a clear focus on how these would contribute to a process of improvement at programme level. One respondent noted that the system they had developed was systematic, allowing formative decisions to be made and allowing for participant feedback that could influence change along the way. The respondent implied that these responses were taken back to the academic group for discussion:

*We've had a system from the very beginning, where those who are coordinators for [the different levels] meet the students to get feedback [related] to what has functioned well, and it's organized in such a way that the midway assessment offers us the possibility to make some changes, whilst the final evaluation only has meaning for next year's students. So I would say that internally our evaluative processes are very systematic. [NOR1k].*

Another member of the subunit ascribed this to the academic interest of the team members in the development of the programme, which included the evaluative activity of the programme:

*So if I should say what I think influences our [decision making], what kind of evaluation influences our study, then it's really what I was just talking about, that it is just as much about the interests of our academic staff who contribute to developing our study programme... as well as [what develops] in their meeting with the students. Definitely. [NOR1e].*

When discussing discussions about the development of evaluation models, another participant referred to a socialisation process of becoming part of the academic group and the role played by more senior members of staff in the development of evaluation models:

*...we have had a pretty steep learning curve with regard to the number of students and [subsequent] growth of the academic group, but there are some of the group who have founded and established, and have it them through and through, whilst others of us are newer to the game and need to be socialized into it. And I think that this is taken care of in an incredibly wonderful way. The academic community. [NOR1d].*

This latter point will be linked to the description of processes outlined in the next section.

Similarly in NOR2 the academic team had developed a process whereby the internally felt needs were taken care of amongst themselves, and the focus on external frameworks was minimal:

*But NOKUT don't receive detailed reports about this programme, so I don't really work very systematically or use much time on this. You must not get that impression. We work more organically and improvised with this, but we do have a great deal of information at our disposal. [NOR2e].*

In ENG1 there were also similar reflections, where some joint processes led to reflection over the way the programmes should be evaluated:

*Because all of us thought, you know, well it's great to have all of these ideas but we'd just spend all of our time evaluating, so it's ways that we can actually... it's formalising a bit more so that we keep things together. [it] was discussed in the context of not just evaluation, but how what is fed back in evaluation can be implemented in terms of improving the course... [ENG1h].*

However, another respondent considered there to be a tension between what one wants to do and one's own value system and the bureaucratic structures that exist around programmes. This framework was also influenced by the perceived demands from the task environment, about what is possible to know about the programmes. The interviewee speaks clearly about trying to balance these factors so that the students will gain the best teaching and learning experience possible.

### *Competence Development*

The subunit members were asked to consider what competence they had in relation to evaluation. Related to the issue of competence were the perceived opportunities within the organisation for training and development in evaluation. Across the subunits the responses were similar in relation to the preparation and development for the evaluation task. Respondents reflected that competence in evaluative activity was generally drawn from academic ‘training’ in situ and subsequent experience. There was an assumption that the professionalism of academics was preparation enough for the task at hand. There was a general reflection that academics considered themselves competent enough to evaluate their own study programmes and felt that the institutional frameworks were limited in form. There was very little in the way of formal training for evaluation; what there was limited to some courses related to the introduction of quality assurance frameworks. These varied between courses run centrally for representatives of the programme staff and topics provided for academic groups to work through. Against this backdrop, two interesting episodes were discussed. The first from NOR1 was related to competence development within and across the subunit. It was noted that competence in the institution was generally based upon developing one’s own experience over time. The respondent felt that the work undertaken on behalf of the academic group provided many opportunities for development, but not necessarily in relation to the kind of evaluation that was increasingly being demanded:

*Competence is built up by evaluating, trying out surveys, considering them within the academic group. We’ve also considered them together with other academic groups, who’ve commented upon the forms, a kind of collegially based assessment. Apart from this I have access to evaluation forms from 2 other institutes, but no training other than the competence I develop through attending seminars, doctoral courses, and international workshops.*

Further discussing a project undertaken with a fellow subunit member, who included application of quantitative and qualitative methods and issues related to evaluation, the respondent followed up this reflection by noting:

*There is a competence development in everything we do, but there isn’t any special training to implement user surveys within higher education [NOR1i].*

Another respondent confirmed these reflections noting that those members who were also attached to the doctoral programme had received training within statistical and evaluative techniques which had given them a theoretical underpinning that they could relate to programme evaluation, whilst it was otherwise generally considered to be the responsibility of the individual to update themselves:

*The doctoral education is a goal-focused initiative, but lecturers just need to take any opportunity that presents itself, whether it is a lecture from a visiting professor or participation in open lectures. In practice training is the responsibility of the individual. [NOR1e].*

However, on the initiative of the programme group focus had been placed internally upon the area by arranging a 2 day seminar to discuss the subject. One respondent recognised that this seminar had become the internal reference point for the group with regard to quality assurance, assessment and evaluation:

*With regard to training, or what we call competence development related to evaluation, supervision, feedback to students, we've had a 2 day seminar within the academic group where the theme was taken up. When we discuss evaluation and supervision, that seminar is referred to. [NOR1g].*

This idea was also exemplified in responses across the other subunits. In ENG2 one respondent with a great deal of experience could not recall any provision being made by the institution, again highlighting the general expectation that academics would already have such competence, or develop them in situ:

*Well, to my knowledge I don't think that they've been offered. It's kind of assumed that you pick them up by osmosis really. I mean there are sort of research methods and evaluation methods kind of courses that we offer to students, and we ourselves would have done that as part of our postgraduate degrees and so on. And most academics are involved in research and so would have developed research and evaluation skills. But that's largely about evaluating programmes out there, as much as evaluating your own programmes. [ENG2g].*

In ENG1 there was a recognition that some courses were made available for staff associated with programmes, concerned with implementing frameworks and measuring effects, where the most notable point had been how widespread a struggle there was with addressing such issues. The reflections across the groups highlighted the problematic issue of having an evaluative focus that they professionally had difficulty accepting in the form it was presented, and which they did not feel they were adequately prepared for. While all of those interviewed agreed that they had attempted to meet the demands set, they also discussed the decisions made along the way concerning the extent the models were valid. As will be seen, however, there was greater variation with regard to the extent to which these issues were discussed collegially.

### 9.1.2 Role of administrative staff

Role issues	Role	Related issues (interlinked)	
	Administrative staff		Level of cooperation
			Degree of socialisation
			Evaluative competence

In each of the subunits there was at least one administrative member of staff assigned to the programmes. While the tasks varied from institution to institution related to organisation frameworks, there were some key tasks that influenced the decision making processes regarding evaluation. While I refer in more detail to the case of NOR1 below, some commonly described elements for the administration across the subunits were to ensure the institutional frameworks were met satisfactorily, to liaise with and help keep the subunit members updated with regard to the formal content require and deadlines to be met, to provide the documentation and statistics required and help prepare the reports. These demands were experienced as coming from within and across the institution and externally from QA bodies and the various Ministries. There were a number of comments across the subunits relating to two issues, the importance of the administrators in quality assuring that the data required was provided in the required format but additionally that these comments would temper the decision process as the subunit members attempted to consider how they would approach the task at hand. This administrative control, or rather adjudication of the process is interesting, related to the idea that the academic members of staff who would design and perform the evaluations would do so based on their experience and professional judgments, but the administrators would help adjudicate the process. However, in each of the subunits the administrative members of staff reported to having and were reported to have a close relationship to the academic members of staff they worked with and alongside. They were therefore drawn more closely into the decision making process. The implication with regard to processes will be outlined further below.

### 9.1.3 Interaction with programme participants

Role issues	Role	Related issues (interlinked)	
	Programme participants		Level of cooperation
			Degree of motivation
			Evaluative competence

In the previous chapters it was recognized that across the subunits respondents considered how the demands for increased student involvement had changed the evaluation frameworks. However the demands for increased student voice did not automatically lead to greater involvement of students in the decision processes concerning the programmes. There were many issues related to this observation, from lack of motivation through to more endemic issues within the

organization regarding evaluation use. It was considered important to ensure the meetings with students were more than a mandated formality. As one informant in NOR1 described:

*So, evaluation is valuable but it has to be used and utilized, and preferably together with these informants. But we ensure this through these monthly meetings that we have with the reference groups where there is one from each study group and we have a programme where we go through the study demands, the lectures and projects. So I feel that we have a great deal of information from the students... [NOR1c].*

Part of the motivation issue was related to the perceived level of satisfaction of participants, which as was seen in chapter 7 was considered to be relatively high. They were perhaps less likely to involve themselves in a critical evaluation:

*Well, generally our students are very satisfied... and we've gained some feedback that it's going really well. [NOR1f].*

While similar processes have already been outlined with regard to NOR2 one respondent outlined how this involvement could be improved, and this involved having a more holistic approach to the evaluative processes on the programme which also included the quality of evaluation of students' work. The respondent intimidated that an improvement of these processes would raise the quality of evaluation generally:

*Where would I wish to see a better system? Paradoxically enough I think within the learning processes, where we could be even better at developing our students' way of working, the learning processes, acquisition of knowledge etc. Feedback in relation to the evaluation that we undertake in relation to students, shows, that, well I think that will be one of the next big issues we tackle, across the programmes... I think that we are pretty much a classic, academic institution when it comes to this. I've changed the evaluation forms... such that they become more of a learning situation for students and this has been very popular; the quality of work has improved. But this is an area where we have a challenge. [NOR2g].*

This issue was raised with regard to the background, experience and competence of students and was noted across the subunits. While there were students willing to join reference groups and respond to requests for evaluation, there was regular recognition that the participants were generally busy people. This was highlighted even further within the web-based programmes. Additionally in NOR1 the subunit members were aware of the breadth of backgrounds students were drawn from, and how this affected their engagement with the programme – a key task was adaptation and involvement based on competence and experience:

*This is, of course, an experience-based programme so we have different groups, and it changes a little dependent upon who is recruited to the study. So we need to make sure that we adapt to the students we recruit... we need to be aware that when [we say] that its experience based then we allow students to use and apply their experience [NOR1b].*

Nevertheless there were those that contributed to the decision process. As will be seen, students were often more active within the commissioned programmes, where they also met on occasions in reference groups with their employer and related the expectations of them there.

#### **9.1.4 Relationship to the organisation**

<b>Role issues</b>	Role	Related issues (interlinked)
	Relationship to organisation	Level of leadership
		Degree of feedback
		Organisational constraints

Important issues were also raised with regard to the role of actors within the wider institution in decision making processes. Respondents outlined 4 related issues: the role of leadership, degree of subunit autonomy, the degree of feedback and organisational constraints.

Within NOR1 there was the general perception that issues for programme development were left to the academic group, but were discussed, challenged and ratified within the Institute programme management body. Therefore information gathered for use higher up in the system was seen as for control purposes:

*If one thinks of evaluation at the Faculty level, and I also consider Institute leadership as well, so they would be more interested in evaluation as a tool of control; I suppose I would have been more interested in that if I was sitting in that position. Of course you want to know if the quality is good enough [NOR1f].*

Respondents did not see this as problematic as on the whole, feeling it to be within the duties of the institution to check on the quality of the programmes delivered in its name. However, across the groups there was a perception of a lack of feedback and members were given little insight into how the information was used. As such the control mechanism was limited.

In NOR1 a respondent recognised that the subunit had a great deal of freedom to decide about the content of the evaluation, and what would be the basis for

decision making at the higher level, but it was also recognised that the rationale behind the task and the purpose of it was to meet the demands set externally as part of the Quality Reform within Norwegian Higher Education, for which NOKUT was established to oversee. As a result the quadrennial programme evaluation was recognised to meet these external demands and designed to focus upon the main indicators as set out by NOKUT to be central within the institutional quality assurance system.

*It's a part of the Quality Reform, so we have to look at student feedback, what changes have been made as result, their results and grades, statistics and such like [NOR1j].*

As a result although there was a freedom to shape the content of the evaluations, there was a real sense that the general directives drew the evaluation focus in a particular direction, and thus the focus of the staff was directed in a particular way. The perception was that the programme had gone generally well and that students were satisfied, and therefore the respondents assumed that they would have heard something from leadership if this had not been the case. But built in this was also a degree of a lack of expectation of more feedback, with recognition that the academic group was responsible for the development of programmes:

*But I feel that is our academic group who is responsible for this. There is no-one that comes to control us or say that we need to do it like this, this and this. In a way it's the academic group who controls it, but with some input from the Programme board. [NOR1f].*

However, this was also reflected to be a response to the fact that the subunit was generally perceived to have evaluation under control. There was recognition that there were many other subunits in the organisation that had not progressed as far as theirs with regard to having satisfactory systems in place:

*I think that as an institution, in relation to meetings with NOKUT, so it's all about having the right information in the right place. Now we did have it so it wasn't a problem, but I understand that there are others who don't and then you suddenly have to gather such information when you meet them. But that is just an impression I have [NOR1f].*

This, the respondent considered, was a challenging position for the academic group with regard to interpretation and communication, especially as they had little experience of feedback further up in the system:

*There is also programme commission for our study, who receive what we are serving them now. But how this is used afterwards can be questioned. [NOR1fh].*



The issue of lack of feedback from higher levels within the organisation was also related by respondents in NOR2. One respondent reflected that it seemed that because the programme was considered successful, that it satisfied the system requirements and student and commissioner demands, leadership had not given any extra help to improving what was currently on offer. An example was given in regard to developing literature for the course, where the respondent considered that little had been done to follow up the evaluations made by programme staff and there was little to no feedback:

*But they don't want to develop [the programme] any further. School leadership development in Norway is a priority area, where there is funding made available year on year. But they seem happy with a little success, whereas it could be an even greater success, and it could have given us a leading position in Norway, especially if we were given the support to develop more research-based literature<sup>176</sup> ... No we never get any feedback from above, never. The capacity of leadership to give feedback is equal to null, but we don't really care about that either. [NOR2f].*

While this might appear peripheral in comparison with the questions about evaluation and decisions here, it was clear that it had affected the respondent's faith in the system, and pointed particularly to the concept of quality assurance, which seemed, as outlined above, to be focused more on control for success as perceived by participants than in terms of assuring the quality of the programme itself. The respondent appeared therefore to be differentiating between quality assurance as control and assessment of the programme itself. As the respondent declared:

*There isn't any quality assurance in a way, not at all. [NOR2f].*

This point was also followed up other respondents in NOR2. There was again a clearly felt feeling that information was passed on and as long it showed signs of satisfaction no further follow up was necessary: When asked what happened with the evaluation information one respondent replied:

*I don't know. It probably goes into the system somewhere, and I would imagine the system has the same kind of attitude as I have; if it's bad they come, as long as it's going well they don't come at all.. [NOR2g].*

---

<sup>176</sup> This was also to some extent reflected in the data gathered from the evaluation of the particular module, where programme participants indicated confusion with the literature available, both in terms of what should be read and why there was a lack of literature compared with other parts of the programme.

This point was again reiterated in ENG1. One respondent had considered how the evaluation system would be followed at higher levels, and what space there was for feedback:

*it is interesting because I was wondering what would happen. Well, so far we haven't had a really bad evaluation so if you did then that might be something you'd want to look at. [ENG1h].*

In ENG1 another respondent outlined how under normal procedures there was very little feedback. It would require a wide-scale evaluation to receive feedback from higher up in the organisation:

*You don't get feedback from the Institution normally unless you're part of some mega-review, but we tend to put bullet point evaluations for the students, and we get some feedback from there, but no, not a huge amount [of feedback]. [ENG1a].*

Again similar comments followed in subunit ENG2. Additionally one respondent connected the lack of feedback, as was seen in NOR2, to the fact many of the comments being passed on were resource related issues which the higher levels tended to leave with the subunit to resolve. This was evident even when the issues were raised by external examiners, which the respondent had felt might place greater pressure on the organisation to respond to:

*A great deal of emphasis is placed upon the comment of the external examiners, although it's interesting to find that if those comments have implications for resources they tend to be batted back to the team to deal with rather than resources being found to deal with the particular concerns. Yes, yes I think that's a fair statement because one of the concerns that was raised by the last external examiner, last year's external examiner report for [NN – course title], was that because we'd had quite a lot of illness and the course teams were fairly small, then it was important that there should be sufficient staff to be available to be able to cover the demand of actually running the course. And this was more of a problem now than it used to be in the past because they have stripped back to the bone within the Faculty [ENG2m].*

These comments clearly overlap with processes and responses from the subunit decision making, but are outlined here to show a more realistic picture they experienced of roles within the institution. A very clear concept throughout the responses across the different subunits related to the internal policy of the Institution. As has been noted as a result of developments from the Bologna Declaration all the Institutions have developed central plans for their QA systems. However, the interpretations vary as do the internal policies to meet the external requirements.

*Institutional structures and constraints*

Another issue that arose concerned the impact of the institutional structures. For example, there was again a perceived problem across NOR1 with the use of information and feedback to the subunit. One respondent saw the problem as structural, and although affirming that the wider QA system was a relatively recent development noted that there were problems with its construction and intention, as well as underlying leadership of the processes:

*When you are building something then you need to create something, you must construe it first, so I don't think there are many higher education institutions that have come much further than to build a structure and create some expectations about measurement and systems. And maybe they are about to discover how we are going to use them. And this requires more than just bureaucracy, it actually requires leadership. [NOR1h].*

In ENG1 there were similar reflections, but where it was further recognised that the systems across the organisation had been limited and there had been little follow up previously. One respondent declared that this contributed to greater learning within the system, where decisions could be based on shared good practice rather than constantly initiating new ideas and processes. The respondent noted one programme group had used their own initiative to begin a cross-group process:

*[we've] started with our [NN – type of course] to have meetings with the other groups in [NN – org name] so that we can share this kind of information, or otherwise you're just starting from scratch each time. [ENG1h].*

### 9.1.5 Role of commissioner and other external bodies

Role issues	Role	Related issues (interlinked)
	Commissioners	Degree of demand
		Level of cooperation
		Evaluative competence

With regard to the responses gained from the subunits in this area, little focus in the English cases was placed upon commissioning bodies for their programmes. As has been seen throughout this thesis, this was not considered a key area with regard to postgraduate programmes being delivered. There was reference made to demands placed by funding bodies and respondents also referred briefly to programmes being run on behalf of the NCSL at other parts of their institutions but these were not specifically seen to be related to the postgraduate programmes. But while there was no direct perceived connection to programme commissioners, there were considered to be indirect relationships with funding bodies.

Due to the way national policy and strategies had developed in Norway, commissioning bodies had a much closer role to and influence upon programme providers. As has already been stated, in the Norwegian cases the commissioners were local and regional authorities. In section 7.1.2 it was recognised that these bodies varied in their demands, ranging from the concrete to the diffuse, as well as their on-going interaction with the programmes. This was also coupled with diversity concerning their interest for and competence within assessing and evaluating the programmes implemented for them. As was mentioned earlier, this was described as an issue of ‘commissioner competency’.

*A question of “commissioner competency”*

In discussion with subunit members in Norway a key term used regularly with regard to the process of responding to a tender and negotiating a contract to provide school leadership development was “commissioner competency”<sup>177</sup>. But it was also noticed that this was only one part of the process, which was described more widely as discussion and negotiation about purposes, implementation and intended outcomes. Meeting the varied needs of commissioners required the subunits to assess the tenders and consider what was appropriate to deliver. This often required and was based on an ‘informal’ assessment of the competence of the commissioners, both before, during and after the initial negotiation process. In conversation with subunit members from both NOR1 and NOR2, reflections concerning this concept were discussed, being linked to their perceptions of impact on decision making with regard to the evaluation process.

A general reflection across these subunits was that the commissioners were not prepared for the role that they had received in this area. One respondent in NOR2 exemplified these comments, noting there to be a widespread lack of competence with regard to the tendering process:

*And it’s clear that some of the problems here are that the owners aren’t very competent; not very good at commissioning. They could have been a lot better... than they are. [NOR2g].*

And as a result of this, the same respondent considered that many commissioners often merely accepted the premises the programme was founded on rather than coming with their own demands. This acceptance of the “whole package” on offer also included the evaluation focus and designs usually employed to ascertain programme quality and goals attained:

*I think that we lay more premises for this than the buyers, because they are not there all the time, they don’t have enough insight, they don’t have the capacity*

---

<sup>177</sup> In Norwegian, “bestillerkompetanse”

*to lay down any premise and that makes this difficult. But they also have great understanding and respect for what we are able to do, and that it has to happen in a certain way, and that evaluation needs to be done in a particular way, and the programme content needs to be such and such. They have a lot of respect for that really; well that's how I experience it. So there is never any conflict... there can be discussions, but never any conflict. [NOR2g].*

A similar pattern had been noticed by members of the subunit in NOR1. One respondent recognised that this was often a problem of scale, where the smaller authorities lacked specialisation within their staff, in particular with regard to the educational sphere. As was noted in Chapter 2, from 2004 the local and regional authorities had taken over the role as employer and their competence in overtaking this role varied (Bæck & Ringholm, 2004). This created some tensions in terms of the control these authorities had over funding and the relationship to HEIs (Wales & Welle-Strand, 2005, 2008). Respondents in this current study confirmed these impressions, recognising that often this meant the commissioning body might only be concerned that a programme was implemented. A respondent from NOR1 considered that the subunit was concerned to improve commissioners' competence themselves:

*Well, we have to "bring them up" a bit. It's a bit about having what we call "commissioning competence", and we see great differences between authorities. For example, we have negotiated a contract and had a programme for one county municipality, which is a large and an important actor, professional with high commissioning competence and competence related to what they want results from and what they want to use money on in schooling their teachers and leaders. This is different from the smaller authorities. Now, it isn't that the smaller authority in this instance isn't as good, but they haven't had a leader who has been employed for the position. So a consultant at the education office is doing a job without the necessary competence. So, it's more that we see they don't know what to evaluate or commission, they just want some leadership development. [NOR1h].*

This also meant that there was a degree of variation in understanding the purpose and role of evaluation, and once again another respondent considered it up to the programme providers to lead the way:

*Very often the mandator doesn't see the purpose of assessment and what should be evaluated, before all the evaluation is completed. It is often up to the institution delivering the programme to make this clear. And to the degree you discuss this with the mandator it is more a question of negotiation. [NOR1i].*

There were, though, examples of commissioners perceived to be highly competent, and these were recognised to already have their own agenda and plan for evaluating the programme. In NOR2 a respondent reflected over the variation between active and passive involvement of commissioners and the fact

that although the norm was disengagement with the programme after commissioning, one local authority had been particularly involved in the on-going processes as well as evaluating post ante, which had also increased the focus of the subunit members on evaluating and had a positive impact:

*I think that when you have an active commissioner there are many advantages. Firstly, you know they are watching you, so you need to keep on your toes, you need to deliver, you need to explain, you need to tell why, and you have a commissioner that wants to measure results afterwards... So there are many advantages, and you could say that usually commissioners don't really have any demands, they just put up and shut up; that's the usual. [NOR2h].*

However, it was suggested that despite being heavily involved in the process, the level of influence or cooperation was considered to be more with help in fine tuning of the programme than in direct influence over and demands for change in the programme content, and evaluation based on these processes. Despite the commissioner performing their own evaluation, the subunit was still perceived to retain control over the basic evaluation process:

*They have been relatively involved in the fine tuning of the programme, whilst the other authorities that have paid for their students have not been so. So it's clear that they have a much greater influence on the programme than the others, but on the other side, it isn't that much influence. It's a little more about evaluating which themes to focus on [more of one thing than another]. But it is at the headline level, rather than concerning the content. But I've never experienced, or pretty much never experienced that an external customer has been so heavily involved in the development of a programme than [NN – mandator] has been. Most of them have contractual frameworks and haven't done much more than say what they want and thank us for what they've received. [NOR2h].*

While the respondent reflected over the involvement of this particular group and their willingness to evaluate the programmes for themselves, there was a slight degree of scepticism over the interpretation of the results of the analysis. The respondent recognised that the additional evaluations that the commissioner had chosen to run appeared more instrumental, focusing upon the perceived satisfaction of the participants:

*Additionally, they have implemented, I think for nearly every group, an ex-post evaluation, where they go into more detail about what they have learned, what they liked, what they didn't like, and they use these, perhaps over exaggeratedly in my opinion, in their comments concerning what they want to change at the headline level. So it becomes a little instrumental, perhaps. [NOR2h].*

Reflecting over this another respondent recognized how the commissioner's own evaluation was presented to them:

*In the bigger evaluation that the commissioner implements they ask if the themes are relevant, if they want more or less of something. To some degree we take notice of this, but not fully. We get this report analysed in advance [by them], and take it for what it is [NOR2e].*

Interestingly, the respondent recognises that the commissioners have focused their evaluation designs, but indicates that the academic group still interprets the findings within their own values and purposes. This idea will be revisited below as an example is given from a commissioned programme at NOR1.

#### *Commissioners and evaluation competence*

Commissioners were often a lot closer to the design and implementations of the programme as a result of the Knowledge Promotion reform, and as we have seen in many respects played a more active role in the evaluation process. But as was seen in the previous chapters there were different degrees of involvement depending on competence factors.

As was seen above in NOR1 it was recognised that commissioners often had a limited understanding of evaluation and the purposes of such a process. This placed greater responsibility on them as provider to lead the development, but at the same time recognising the need to discuss and bargain concerning the focus of the evaluation and deliberation over what information might mean. It was also recognised within the subunit that those responsible for the programme delivery would need to develop methods to help commissioners and qualify them for their role and this might often lead to a bargaining situation but with the recognition that the process was a partnership:

*But we then find a form that perhaps will help them find out a little more about what they want and why they want it. So you can say that we are an actor working together with the school owner to qualify them for their role. But we also learn a lot as well about what the field wants and what we should together evaluate. [NOR1h].*

But the respondent also recognised that the process might be about difference of opinion over the basic course purpose and structure, especially when an actor with clear demands was part of the process:

*You can negotiate about meaning as well, we did that with one regional authority where there was an important actor who knew what they wanted and wanted what they paid for. They wanted a great deal of leadership development in a short space of time, which it is possible to problematize, and we tried; we asked if it is a good idea to take 30 study points in 8 months, problematizing the learning effects. We asked if it was better to take it over 2 semesters and rather have processes that could develop a bit. But they didn't want that, they want to*

*get it over with quickly... but we negotiated with them, to say that the learning effects wouldn't be so good, but it's they who are commissioning and paying. So that's how it is. It's difficult, and we can problematize it as academics, asking how can they measure effects if they haven't taken account of the processes needing time to develop? We can ask such questions... [NOR1h].*

*Commissioners: An extra level of tension or release?*

In chapter 7 it was reported how one respondent from NOR1 had experienced the clear, “concrete” demands of one commissioning body as they focused upon school change. As the purpose was to reinforce change the mandator had wanted and taken a very proactive role in the process of the project. As the programme had progressed, it became clear that such a process added an extra level of complexity with regard to assessment of the programme. The subunit members now had to relate to the programme participants and the commissioners, as their employers. The addition of the commissioner as an extra link in the stakeholder group had led to a negotiation process in terms of what impact the programme should have and how it should be evaluated. The members of the subunit responsible for this programme recognised that they would have to develop a way of formatively evaluating progress that connected these two groups and to avoid the tensions of trying to direct a programme as well as overseeing the implementation of the processes at school level. To accommodate this, this meant devising an additional programme including the training and involvement of their staff in supervisory roles for programme participants. This meant the programme provider group had “two groups” to relate to:

*And it was very interesting because we had two groups to related to; firstly the commissioner who was the county authority... and so all these students. In order to combine things we developed a model where were trained supervisors who were employed by the authority and involved themselves at school level whilst we carried on with other things. We provided the teaching whilst they focused on the processes, following up at school etc. [NOR1e].*

When reflecting over the impact of this upon the evaluation processes, the respondent recognised that the result of this decision was a much more dynamic assessment practice. The programme team had then in cooperation with the mandator developed a series of indicators for change and development, particularly that what they had learned would be trialled in their own schools. The most interesting facet appeared to be the way that this compared to the formative evaluation processes and the decision making processes on the full programme. The respondent outlined how the proximity of the commissioners, programme participants and programme providers had led to much greater interaction with the task environment leading to more fluid and responsive interaction and programme development. As the respondent noted, “we focused heavily upon a process evaluation and the feedback we got; they responded continuously so the programme was being adapted all the time”. The application



of programme theory and constant interaction led to dialogue between the groups and discussion concerning fundamental programme issues. This was contrasted with the perceived experiences from evaluations of the full HEI based programme, which the respondent felt to be less effective, with delay on all decisions seemingly made retrospectively and appearing to follow a plan for evaluation rather than producing a more dynamic process like that experienced on the external programme:

*We don't do that here. You could say that even though we have a structure for evaluation where we ask them once a year and they need to answer different things and sit in groups talking about this and that, so we [already] have a plan. We have lecturers coming in there and then. So the adjustments we make are much smaller, in addition to having a tradition where we bring in these lecturers with their own materials. But we only a few outsiders, so after a while we made the programme for them all the time, we developed the theory in a way that they would understand it. [NOR1e].*

The respondent recognised that this appeared to conflict with the subunits approach to decision making and evaluation, creating a paradox with regard to the design that had been implemented:

*So if you ask about evaluation, so then it's just about when you have a programme that is implemented over time such that you have opportunity to run a process evaluation such that you can use the feedback in a way that you can change the programme. And I feel that's a bit paradoxical here. When we finish each level we ask them what they think, and we are a little disagreed what this evaluation means. But, we try to make some adjustments, but its new students coming in and maybe it doesn't suit them. [NOR1e].*

Another respondent agreed that the external programmes offered an important contrast as the extra level of commissioner with their demands and expectations offered another interpretation of the programme and its goals. On another occasion the commissioner had recognised a need amongst their leaders for increased competence in analysis with regard to interpreting league tables. The Master programme as whole was adapted to accommodate this, with methods teaching moved from the final year to all levels. The respondent recognised that they would have been unlikely to do this without the feedback from the educational field coming directly through programme negotiations:

*So I think it's a good example about how a process of negotiations with school owners contributed to a realisation that we needed to do something with our master programme. [NOR1k].*

## 9.2 Decision processes

In this section I consider what processes were perceived to take place within the subunits when discussing evaluation models. As has been seen in the previous sections of this chapter, these processes are set against the backdrop of different roles. By way of introduction it is important to highlight that the respondents across all of the subunits characterised in some way the organisational frameworks as bureaucratic and as was outlined above they perceived very little feedback.

A respondent at ENG1 exemplified these comments when outlining how the institutional system had become more bureaucratic over time noting how it was often unclear whether evaluation was implemented to improve quality or to satisfy external demands. This theme was originally outlined under demands but here the discussion relates further to how processes develop and how respondents interpreted them in their decision making. The respondent wondered whether the focus on quality had in fact improved the quality on offer:

*I think that's it a bureaucratic process in a way ... I've become quite cynical about quality initiatives, quality management and in a way I think there is an inverse relation between the degree to which organisations say that they are quality organisations and the quality of what they offer in the classroom. I think there's an inverse relationship because people spend an inordinate amount of time going through bureaucratic quality processes, form filling, box ticking etc, and it impacts on the amount of time and effort devoted to the classroom. I'll give you an example here, for the last five years in my reports, every year I explain about the quality of accommodation including things as banal as squeaky doors of the lecture rooms. The doors still squeak and bang over there. Every year for 5 years I have asked if something can be done about that. So that's just an example of why quality has become fairly meaningless I think. [ENG1g].*

Bureaucracy was tied to the lack of feedback noted earlier. Another respondent at NOR1 had noted a legitimative tone to evaluation processes in relation to the wider system. There was a necessity to follow the formal system, providing the data that was required, but with some freedom to influence the choice of methods applied. However it was demonstration of control that appeared in focus:

*The systems, in a way you just have to follow them, but you can influence to some extent – you can influence how you implement a midway evaluation and a final evaluation. I feel perhaps though it's more like we have to show that we are doing something, it's that they are interested in, what we are doing. I suppose that's fine, we should report what we are doing. [NOR1c].*

It was experiences of use, or lack of them, that appeared to influence the attitudes and decision making of the subunit members. Another example was drawn from the subunit of NOR1 with regard to a forthcoming programme review. When this review was compared to annual cyclical process within the system the respondent recognised that that the cycle of decision making differed at the lower levels. Although there was no feedback, the results were interpreted locally and used to develop further designs. There was a sense that the formal processes had little effect higher up whilst they encouraged more informal responses at the local level:

*I have absolutely no idea what they use the results for within the university system. So if I should summarise my understanding of the quality assurance system then I would say that it functions well at programme level where we have evaluations that we can develop to be used for a purpose, but it's not visible how they continue up through the system. We recently had a visit from NOKUT and this was one of things that I raised with them, that the QA system is very much channelled in one direction in the system, while at the same time the Quality Reform has made us much more aware of the importance of the formative evaluations that we implement ourselves. [NOR1j].*

This was an interesting comment whereby the respondent considered the formalised systems to be irrelevant to the programme but recognised how it had highlighted the importance of the work they were already doing.

### **9.2.1 Perceived internal proximity within the subunit**

This section and the next consider how the subunit perceives the group processes surrounding discussion and design of evaluations. It is recognised from the data collected that the issues overlap with other processes taking place within the group. In analysing the data from the interview transcriptions four major areas of focus in relation to the phenomenon are outlined. When interviewing members of the subunit in NOR1 focus was drawn towards collective group action, which was portrayed as a democratic, collegial process. When relating it specifically to evaluation design, I chose to refer to this as a “collegial construction” of evaluation models. Such an approach, as will be seen in the next section, was not as recognisable in the responses of the other subunits. I do not however suggest that it wasn't present in the other groups. But if it was present, it was not mentioned in the same way as in NOR1. At the same time the groups demonstrated responses that I coded under another decision making response, ‘dismissive submission’, which is outlined in the next section. I have therefore chosen to relate these descriptions as a vignette of processes recognised by this one group.

First I briefly outline perceptions of its basic characteristics of “collegial construction” and then extend these by drawing on descriptions of relationships and processes, before recognising some challenges outlined by group members.

These reflections are then followed by a presentation of the data collected from the other subunits with a description of ‘dismissive submission’.

### 9.2.1.1 Reflections from subunit NOR1 - *collegial construction*

Within NOR1 there was a notable discussion concerning the within group roles and a sense of strong collegial proximity and positive group dynamics related to evaluation and related decision making.

#### *Characteristics*

Theme	Descriptive element
Collegial Construction	Collective action
	Local initiation and power sharing
	Participative and critical approach
	Development focused
	Sensemaking/giving approach

The various respondents described the basic characteristics of the way the decision processes unfolded within the subunit. Attention was drawn to a collective interest in developing the programme, with one respondent describing it as a “voluntary community action”<sup>178</sup>, implying that a greater collective effort than normal was required. Within this ‘project’, it was considered that the processes were driven by an open and comprehensive approach to communication across the team using different channels but particularly through the frequent subunit meetings. The format of these meetings was described as encouraging participation from all members, based on discussion around themes concerned with the implementation of the programme and focused upon development. Focus upon evaluation was considered to be particularly prevalent within these meetings:

*We have a meeting for the academic group once a month, where we amongst other things had in the spring a great discussion about what survey form we should use and what questions are best when implementing a written evaluation at the end of each module. [NOR1a].*

Members of the subunit were under the impression that they had greater control over the processes of developing evaluation designs. As was seen in the previous sections, within the wider Institutional framework, as well as other demands for quality assurance, the information required at higher levels of the organisation appeared to be at a much more general level and so could be accommodated within the approach the subunit had developed. Despite, then,

---

<sup>178</sup> Translated from the Norwegian word, “dugnad”, which implies a communal based, often voluntary action, to achieve a collectively considered worthwhile task or project. This is an important and widely understood cultural reference point.

the existence of demands there was a strong perception amongst subunit members that considered that evaluation design processes reflected their own focus upon programme improvement. And this was also exemplified by the perception among some members of the freedom to direct the course of the evaluation focus, rather than as a result of a demand:

*This time, what we are going to evaluate, well it's based on a desire that we have had, and a felt need for improvement, so there isn't any pressure, not at all. [NOR1g].*

This perception of the group and its decision processes also spilled over to other areas of the programme development. One respondent remarked over the openness and reflective nature of the group in which a critical approach was adopted:

*[we] have an academic group where we discuss and challenge one another; and I have a questioning nature to many things, so it is easy to come in and be allowed to be a contributor in this group. So for me it's a very creative, challenging and stimulating milieu; the academic group is a pretty important part of the study. It's there we discuss plans, how we will frame the work. We can help each other with supervision, as we spend a lot of time doing that; and also support the students' learning processes, how we can supervise in a way that contributes to this... as I am new to the field, or new here, I've gained a great deal from the academic group. [NOR1b].*

The control the academic group had was felt by one respondent to be based upon the legitimacy of being members of an academic, professional community. This was evident by the perception that it was the responsibility of the group to defend their right to adjudge the quality of work produced by students:

*It's much more up to the group to discuss the results for the different levels over time, taking a professional decision for what is good or bad. So the academic group has great legitimacy with regard to making decisions, when it involves this kind of thing. [NOR1i].*

This is also linked to what, as was related earlier, respondents in ENG1 considered as anchoring for evaluation, academic experience and socialisation within an occupational community, even NOR1 group members referring to the term 'community of practice' about themselves.

Despite the fact that the model had been developed by some of those who had been there longer, there was still perception that the group was being collectively developed and that new members of staff were contributing to the evaluative assessment and progression and bringing their own expertise:

*My experience is that this community of practice that we are developing now is the continuation along lines already laid down, but that we are developing a great team now. I came into the group during a period when there was need for coordination, especially since the first two levels work with similar themes. [NOR1b].*

Another respondent was asked as to whether the programme basis and particularly the evaluative models that had been adopted at earlier stages of programme development had driven the current approaches and therefore only gave a sense of collegial development. The respondent agreed that earlier models had obviously framed much of the discussion, but at the same time there was openness for change and development:

*Yes, I'll agree with that, but there has constantly been a development and I think it's important to point that out, as new people have come into the academic group (they) have questioned things... [NOR1j].*

A collegial approach was considered to be of particular importance due to the rapid growth of the subunit and there was recognition that this could potentially lead to fragmentation, with a split between, as one respondent put it, those who had established the programme and the “newcomers”, but this had not been the case:

*There hasn't been anything else than that we have talked about creating the room to raise questions in a more holistic perspective, that we create this, because we have had a steep rise in the number of students and development of the academic group. So some members of the group are entrepreneurs who established this and have it in them through and through, whilst others of us are newer to the game and need to be socialised into it. And I feel this is dealt with really well (in the academic community). [NOR1d].*

However, as will be seen later in this section, despite appearing to contribute to the development of collegial processes, the growth of the programme and subsequently the numbers of staff, the expansion also led to a perceived increase in complexity for decision making.

#### *Processes involved*

In this subsection examples are given of some of the processes respondents in NOR1 associated with decision making concerning evaluation. The process by which decisions were made and the degree of involvement were interesting within this subunit. In describing activities and decisions made respondents gave an insight into the underlying mechanisms influencing these processes.

It was interesting to hear how the evaluation models were perceived to have initially developed. In particular one respondent explained how the current

model had been built upon a design for an externally funded project that preceded the Master programme, where there had been an expectation of a summative evaluation focused upon participant satisfaction. This model had later been developed to also meet the felt need by the group to self-evaluate the Master programme as it progressed, albeit within the institutional frameworks that were being laid down:

*[the evaluation system] was developed with an idea in mind that the programme would be assessed for external purposes. But then of course we also have of course a need to evaluate; and it's about how you develop a system along the way, and this part of the internal decision making processes. [NOR1i].*

Despite this internal development of evaluation models, there was still recognition that the decision processes were framed by the institutional structure. Another respondent recognised that this meant the subunit was not decoupled from organisational frameworks, even though there was a sense in which there was openness in the system for manoeuvring:

*With regard to decision making process, well they are formalised in a way that the academic group is a discussion group. We can take decisions within the group but then there is a programme committee for [our programme] which also includes external representatives, and it's there we present our decisions. And so there are new decisions made there; where the Committee for Academic Affairs<sup>179</sup> are the ones who decide if changes are to be made. So this is within a structure tied to our Faculty. You need to use the structure that is there when making decisions. [NOR1k].*

The specific processes that were outlined in the previous chapter were developed to include formative interaction with the students, which was part of the Quality Reform in Higher Education. This creates an interesting area of reflection concerning the interaction of the subunit members with the programme participants. In this regard it was reiterated how important the dialogical processes were. An example was given from the reference groups, where this collegial construction was recognised to require further rounds of discussion and decision making, before returning to discussion with participants. One respondent reflected over this need in responding to the students required discussion amongst the academic staff, both at module level and within the whole subunit before following up:

*We can of course comment upon, there and then, some of it, but other parts we need to talk about afterwards – is this something we need to do something about, or develop further, should we go back to them for more information? So we need a meeting after to give feedback to the students about what we've considered, and maybe we get new feedback (from that). So it's not just*

---

<sup>179</sup> Institutional level, <http://termbase.uhr.no/?lang=&q=S>

*decisions made from above, where we decide what to work with; it has to be done interactively, where it's important that they get to say what is on their minds [NOR1c].*

The collective image of evaluation processes and the way that this underpinned the whole approach to the programme was also reflected upon. This description of and reflection over a collegial process was more evident in this subunit than in the others under investigation. This will be to some extent explained by the structure of the programme group, but there was also a common conceptual connotation of evaluation and its purposes, markedly different from the other groups. One respondent considered that the subunit interpreted evaluation within a broad perspective, seeing it as a part of their on-going improvement focus, which was a continual activity. It was this approach that gave the members confidence in their programme and method of delivery:

*...I think that we get it pretty much right here; we're on the way. We try and we fail. We're working on developing our study, and we will develop it and we want to generate research on the field. And so evaluation of our own practice and programmes (is key), where we evaluate many different parts of our activity; everything from the curriculum, to the way we organise things, to how we choose lecturers as well as using ourselves, the form of our lectures, use of powerpoint or not, working methods, choice of room - that is, everything is questioned, and we try to think in a new way, every year. [NOR1h].*

There was recognition from the same respondent that although the organisational frameworks must be followed, the response to these demands was based upon a collegial reflection:

*So it's not just because some decree comes from above such that the organisation tells us to evaluate our programme and we do it conscientiously, delivering it just to be ticked off and put in a drawer. It's much more about dynamic processes where we continuously develop our practice. [NOR1h].*

And this was exemplified in this case that while the different module component evaluations were administrated by the various coordinators, the process of formation, collation and summarising took place within the collegial group. Another respondent described how during the process of reporting findings from the different modules the group had:

*... collectively discussed how we can (report findings) and found a common understanding of what it could be like, that is, a system of our own*

And the same respondent went on to describe how the process was as important as the results of it, as it would commit the subunit to utilisation of the findings:



*That type of reflection that evaluation creates room for is extremely fruitful really for an academic community, but it demands that we make time for it and that it isn't just an instrumental activity that we do just because we say that evaluation should have a systematic form. But my experiences are that we actually use it. As such it becomes more meaningful than to be part of organisations that we have evaluated, but that was it, it was just placed in a drawer. So in my opinion the processes surrounding (evaluation) are very, very important. [NOR1d].*

The commitment to collegial and local utilisation was seen therefore to be vital, but as is also recognised in the subsection further below dealing with challenges, that one of the issues raised regarding collegial approaches concerns the time that they take to implement. But in NOR1 respondents spoke about the importance of these processes and how they had been built into the programme from its inception, with a particular regard for internal utilisation. Interestingly the subunit members now saw that many of these demands were coming externally. Respondents were subsequently encouraged to consider how these processes had developed and who led them. Their responses focused heavily on what one member referred to as “integrated patterns of thought” around learning assessments, even though it was the academic group leader who “kept the processes in order”. Interestingly the group were observed to follow institutional rules because the organisation “can come to look at our cards, and that has importance with regard to that we take it seriously with regard to documentation and such things. I would imagine that plays an important role (in our decision making). But the one who keeps control of the processes is the leader of (our) academic group, who coordinates the activities” [NOR1b].

### *Relationships*

Another part of this overlapping concept concerns the internal relationships within the subunit that were perceived to drive these collegial processes. In describing their “hands-on” system for developing evaluations, one respondent reflected further how the programme group as a decision unit played a central role in acquiring and utilising the information. As has already been outlined, the programme group at NOR 1 was described as being “development oriented”, emphasising the cooperative spirit that existed, developing models in advance of demands and ones that had been adopted elsewhere in the organisation. It was interesting to hear how members constantly referred to the ‘group’ as the point of reflection:

*I have never experienced a place where there is such a high degree of interest and willingness to change and develop systems. So we have a very development oriented group here, and as such we lie pretty much ahead (laughs a little). It's actually true! [NOR1i].*

A number of the respondents referred further to the “culture” within the group, focused on change and development, where evaluation was seen as a key part:

*It has been a great deal of fun and very interesting, especially with a culture for change... and development. [NOR1h].*

Another respondent also reflected that as a new member of staff had been surprised over the focus upon development and improving the quality of the programmes, perceiving the interest around their activity as indicative of their approach and that it was worthwhile to get further involved. At the same time the group had been willing to hear the new member’s thoughts and reflections from the very beginning, where the processes were deemed to be open such that it was easy to make a contribution:

*There aren’t many deadlocked thoughts that you need to be like this or that, but you can quickly become involved in the development of the programme... I think it’s part of the culture here that we are interested in quality... and that’s the most important part for me, to be part of academic community; where there is a considerable pressure to learn. [NOR1c].*

The members of the subunit highlighted the importance of critical discussion as part of the collegial approach, seeing leadership as a relational process. This affected the way that they designed and approached their evaluations. The interaction between subunit members and between the subunit members and the programme participants were considered crucial in this respect, recognising that the different forums were channels to support these processes, which were followed up in the academic group:

*(Feedback) can come out in the reference groups, where we believe that good arguments can win through... and we challenge students that if they want to (change things) then we are very willing to hear their argumentation. [NOR1k].*

#### *Relationship to administration*

The role of administration on the programme at NOR1 was also interesting in relation to the concept of collegial construction. It is highlighted here as one often sees a separation between academic and administrative staff. One of the roles of the administrative staff in NOR1, as outlined earlier, was observed to be to keep the academic members of the subunit aware of demands related to evaluation and to control that the activity is done, particularly concerning the internal quality assurance system and the preparation of an annual plan. An administrative member of staff outlined the distribution of tasks like this:

*It’s the academic members of staff responsible for each of the modules in the Master programme who carry out the evaluations and summarise the results, but I am also a member of the academic, and I need to remind them and control*

*that things get done, things like the formative evaluations... this is to ensure that what we do is in agreement with what is demanded of us. Because there are demands and we have to act according to them, for example from NOKUT and from the organisation centrally as well as the Faculty... It's the administration's task to make sure things are done correctly and sufficiently in relation to the demands. [NOR1a].*

Both the academic staff and the administration spoke warmly of the relationship between them being one of close involvement. More than that, administration was described as being a part of the decision making process within the subunit. One of the reasons given for the close involvement of administration in NOR1 was the academic background of the administrative support. A number of informants referred to the fact that having a pedagogical based qualification meant the administrative representative understood and supported the underlying values of the programme and the approach that was being taken, identifying with the occupational community. The respondent in an administrative position commented upon this influence within the decision process, and how the role had developed:

*I am part of the discussion on an equal footing with the academic staff, and you could say that I was given this role because I am qualified within the subject area. And so I've got a kind of a double role, because I am part of the academic discussions. [NOR1a].*

*Problems with success: time for collegiality?*

Due to the policy reform for competence promotion and the resulting availability for funding for school leadership development, combined with perceived success of the programme since its inception, the subunit had rapidly grown in size. One respondent recognised that this challenged the collegial approach to decision making that had been taken, which had been more democratic and open but was time consuming. After restructuring, the leadership roles within the subunit had been divided up, with one member appointed with specific leadership responsibility as programme coordinator. Within the general collegial approach and the wider freedom group members felt, this respondent reflected over the ways the process was directed by the different leaders. One led the processes more democratically, allowing for much discussion and input but also with a strong theoretical persuasion, while another who had led these processes was considered more goal oriented, focusing upon the task in hand. In addition the new leader was considered to have adapted experience as a school leader into the role, taking more control over processes, acting as when a school leader, "working in a more political...very structured way". The direction was still then perceived to be the same, but the process by which decisions were being addressed was different. The respondent was then asked to reflect over whether this had had an impact with regard to decisions concerning evaluation models:

*Definitively, because evaluation then becomes used as a tool for steering. Then, one adapts the evaluation to what you want to achieve, and I think we will see more of this in our group. But this is dependent upon being narrower in what you want to discuss, you can't just talk about everything. Such is it when a group grows, you can only take up some things [NOR1e].*

*Balancing accountability and improvement: legitimacy for difference in common frameworks*

Despite the reflection that evaluation decision processes were initiated within the subunit, the progressive tightening of the institutional frameworks challenged their freedom to act. While still considering that there was a good degree of control, there were noticeable effects particularly in relation to the Quality Reform in Higher Education that had filtered through the HEI system. One respondent at NOR1 recognised how the discussion at subunit level was still extensive, but that the frameworks being drawn up at higher levels were becoming more accountability based and more restrictive. But challenges related to the linkage between the internal subunit system and the formal demands within organisation were also seen to have produced positive repercussions at the subunit level. One respondent exemplified upon how preparation for the quadrennial programme evaluation interplayed with the internal evaluation processes. The processes at the subunit level had become quite methodical based on the QA frameworks set up system wide, where each level responsible for preparing a report that was discussed further and more widely at the next level before a final report was prepared from a plenary discussion. The respondent saw there to be developing greater interplay, where their processes had become more systematic and likewise had increased discussion surrounding internal practices that were not part of the final report. The respondent recognised that they could not report everything they had done and so there would be some gaps. For example, when suggestions for improvement came, these had often already been addressed and were in place and were not reported. However, these reports had also been seen as positive as they provided written documentation for the subunit about what they had achieved and gave them a good record when they reviewed programme development:

*...so it shows how we, in a way, see the external and internal in interplay... but that type of discussion has been the basis for our production of these annual reports. Personally I think it is fine to have to produce such reports, because they contribute to systematisation and you get them in written form so that you can keep record over time. [NOR1k].*

Even within these collegial constructive processes one sees how the external demands direct the work of the subunit. The increased demands in the system was shifting the framework for the group, and they saw the dichotomy between greater freedom to influence a developing system whilst being directed towards

certain ways of evaluating that if not different to that which they had put in place, were formed for different aims. Internally, then, they had maintained a collegial approach but were being more instrumentally directed towards certain themes. They consider then collectively how to respond to the demands placed upon and the results that the evaluative activities bring. The greater systematisation the group gain is counterbalanced by the requirement to highlight and report on areas that they are not always in agreement over. But as has been seen already the group then consider how they will respond and whether or not they can suggest and offer changes to the system. On occasions this has proved to be a successful tactic, where the competence of the group has been recognised by those at levels above them. However, these constitutive processes also over time do cause a substantive shift in focus for the subunit. Such responses are discussed further in section 9.3.

#### **9.2.1.2 Reflections from the other subunits: *dismissive submission***

The presentation of the data from the other subunits should not be considered as presented as a direct comparison or contrast with NOR1 for the reasons that have already been outlined, particularly that the groups were organised differently within their wider organisations, that portfolios of programmes both in terms of number and method of delivery showed some variation. Although one cannot be certain that elements similar to those described as collegial construction did not fully exist elsewhere, the data gathered suggests that it was not as widespread or approached in quite the same way. Therefore I only reflect over that members of the NOR1 subunit discuss processes surrounding evaluation decision making that offer an interesting deviation from the other units. However, the data collection process began with the NOR1 subunit, which meant that I had already considered their responses before visiting the other units.

It is also recognised that these categories should also be seen as fluid, with programme portfolios developing, internal structures changing and external demands within and beyond the task environment shifting. However, in this section it is recognisable that the subunit members in NOR1 were specific in their description of and reflection over the collective nature of their approach in a way that the members of the other subunits were not. Taking these degrees of freedom into the reckoning, the next section outlines descriptions of the processes as outlined by subunit members in the units NOR2, ENG1 and ENG2.

#### *Discussion processes within the group: Looser within unit coupling?*

Analysis of the transcriptions revealed that there was not the same degree of collective discussion across the subunits regarding evaluation. As I have already stated, many factors may influence this, including not least involvement in and workload pressures concerned with other tasks apart from the postgraduate programmes under focus here and the arrangement of the programme portfolios and different organisational structure. However, in the three subunits there was a

sense of less interest in discussing the subject of evaluation, but additionally less internal cohesion in the group between the different members. In ENG1 respondents referred to the organisation of multiple programmes within the subunit as a hindrance to cross-unit discussion, whilst in ENG2 and NOR2 the modular format and organisational structure also led to more internal fragmentation.

Therefore, overall areas of interest from the responses given include the sense in which the different subgroups are perceived to coordinate with their central evaluation/quality assurance system and the level to which they appear to be internally coupled. As a part of this, discussions were focused upon perception of pressures faced and levels of agreement with evaluation frameworks (which also overlaps with evaluation values). One resulting area of consideration, as will be seen below, appears to be that the reduced control and influence over the central surveys also appeared to reduce the amount of discussion, but this ought to be investigated further. As I have already stated respondents from NOR2 did not refer to the same kinds of within sub-unit discussion as NOR1 with regard to the formation of evaluation. Responsibility for design and data collection and collation was taken centrally, although the administration for the programme attempted to coordinate with those working in the wider system. There appeared to be much greater decoupling from the central frameworks in NOR2. I deal therefore with these decision processes first before turning to the English subunits.

In NOR2 the decoupling was evidenced by the lack of participation in the central evaluation process. Respondents appeared to give little regard to the evaluation findings, and this may partly be explained by the limited role they had in the process. One respondent who was responsible for a particular module in the programme reflected over the fact that these evaluations were “nothing to do with” him:

*I don't even hand out the evaluation forms, but just get a copy of the evaluations when they are completed... [NOR2f].*

The subunit members did not play any part in their formation and felt that the interpretations placed on the findings at central level were made without any reference to the academic staff involved. In addition the evaluation process was considered by another respondent to be more of a ritual, wondering whether the organisation paid attention to the findings:

*It's implemented centrally without any connection to [what we are doing]... I think the evaluation is more like a ritual action than a serious evaluation, but that's a little how [the organisation] functions*

The respondent went on to compare the internal quality assurance system and central evaluations with experience from providing programmes for other

mandators. The respondent reflected over how the academic group had felt the benefit that one of the commissioning bodies in particular had their own evaluation process. This view, however, was tempered slightly by their experience of the way the data was analysed and used:

*So for us these more comprehensive evaluations that [the commissioning bodies] have done are very useful. I'm sure of that, but I think also that they have been a little too instrumental. I mean, if someone didn't like the study trip, well then they don't want a study trip. I think that's just a bit too simplistic, a bit of a cowboy attitude to their utilisation...[NOR2h].*

The group had not really taken the content of either of these forms of evaluation into full consideration, but rather had briefly considered how the results had been interpreted at higher levels. The processes of evaluation and decision making concerning the programme itself were considered to be taken within the subunit, but not in a collegial sense as was exhibited at NOR1 but rather based on the individual members own professional judgements of what the programme should involve, how they should be implemented, what the outcomes were and how they should evolve. As the same respondent followed up:

*Any change of direction for us, the changes, these occur primarily from academic developments that are anchored professionally, in an academic environment and academic discussions. These are the main sources for change of the programmes.*

This idea of academic anchoring will be further discussed below in the section on responses. However, according to respondents, despite this understanding of the process there was perceived to be only limited discussion across the subunit with regard to the evaluation processes, both in terms of design and follow up. The programmes were considered to be independent and left mainly up to the module leaders to organise. One respondent was asked whether evaluations and their findings were discussed across the subunit:

*Yes, but not to any great extent; it is pretty much happens between the individuals [responsible for modules] and their teachers, that is those teachers that are brought in to the different modules [NOR2g].*

So while there was some discussion this was not to any great degree raised for further discussion at the subunit level. This was also recognised by another respondent, who considered that this way of working was a result of how the overall programme had changed in structure over time and its relationship to the programme structure in the wider organisation. This respondent considered there to be:

*...little [discussion], but not so very much really; the programmes are, I suppose, relatively independent, because that's the whole logic of [the master*

*programmes in the organization]. The way that we have kind of forced them together in a package is completely, if not unheard of well, it's unusual, even though it wasn't originally like this. Originally there was a common concluding programme and a little structure throughout. [NOR2h].*

Another member agreed, considering that the main focus of discussion was about the implementation and follow up of the programme administratively, rather than on the programme content, and hence there was little discussion about evaluative processes. This meant that the central surveys were given relatively little following up as well, but this was partly perceived to be attributable to the success of the programme:

*It's only administrative things that are discussed, that is if the administrative things have not functioned well enough then they are discussed. But there hasn't been much reason to discuss the programme content because it has gone so well, and everyone has been more than happy. [NOR2f].*

Additionally, however there was a lack of feedback from higher levels. As the respondents discussed the way they reported information through the system, a picture emerged of a group that felt leadership had little if no interest in the areas that they raised concerned with suggestions for programme improvement and development, especially when these might be related to requests for increased resources. This reflection was noted across all of the subunits under study. Another respondent from NOR2 considered that the organisation did not know how to accept and respond to feedback, describing the system as “terrible”. It was here, perhaps the decoupling from the system developed:

*So it's all about doing it yourself, finding solutions yourself and there is very little support in terms of staff in relation to this. But I do have support from [members of the Faculty] and [the programme administrator] and this is a fantastic support and form for dialogue and it's this that is decisive. But that is on the horizontal level. But if you think about the vertically in the system then it's completely dead; I'm clear about that. [NOR2g].*

The respondent was asked to reflect further about these statements, particularly as they raised the issue of the place of the subunit within the wider system, in particular in terms of decision making structures. This was interesting because it was a theme taken up time and time again on the programmes for school leaders in terms of how they should organize their working systems. The respondent was asked in particular about this tension in the area of feedback and reporting within the system and how it could be solved. The problem was perceived to be one where leadership was not “genuinely concerned with academic quality, quality of teaching, development of pedagogy – it's completely dead to say it plainly”. There was, however, recognition from elsewhere that some feedback



did come in terms of the central evaluation to the subunit, but this was little discussed:

*There is a little reflection that takes place, the programme leader, institute leader and programme administrator all get a copy of the evaluation, and there is of course a little discussion around this, but I don't see it as any deep discussion into what goes on here. Perhaps a little in terms of the 'headlines', in the most general sense, if there have been poor or unpopular lecturers. But there is little of this and we pretty much do the lecturing ourselves; if others come in then it is only temporarily. [NOR2h].*

There was also seen then to be a discrepancy between the formal evaluations and the follow up within the subunit. The information was generally thought to be superfluous to the academic reflection that was tied to the modules making up the programme. Another respondent also picked up on an idea mentioned earlier in our discussion in terms of loosely coupled systems, recognizing that the programme had developed in what seemed to be quite a laissez faire way:

*But, it could be that you can call this a loosely coupled process; it's not a process where we have developed an annual plan and decided upon a particular date. We must of course be aware of the calendar, and the startup of programmes, but it has been pretty improvised here. The process has been improvised and steered by situations outside of our control. [NOR2e].*

This of course should not suggest that there was an absence of any consideration of evaluation and what the designs should be. There were some examples of changes made to please the commissioning bodies. But as has been intimated already, these decisions were perceived to be taken as a professional judgment with regard to the modules that the individual academic was responsible for. This was the reference point for assessment of any responses drawn from the evaluations findings. A respondent recalled the way a programme was redefined by the subunit to meet the understanding of the commissioner:

*We get feedback from the [electronic evaluation forms], the panels and from discussion with the commissioning bodies... and we have an internal discussion about these. And this has amongst other things led to [a programme] that has developed a great deal and we have changed it to become closer to the school system. [NOR2e].*

Despite all of this the members of the subunit felt that the evaluation structures in place were suited to the loose programme structure. Individuals made decisions about the evaluation of their part of the programme based on their own self-reflection. While it might be seen as a hindrance not to have feedback in the system this was mostly because they felt there to be a lack of resources available. Their day to day contact with programme participants and the

flexibility in terms of control of the programme meant that they could adjust formatively as things progressed:

*In many ways it's completely satisfactory form of evaluation, and we know when the students are angry, and we are capable of making changes; the programme has been significantly changed by student feedback. SO in many ways we have sufficient control of our individual programmes. [NOR2e].*

And the nature of their loose coupling did not stop the subunit members developing ad hoc systems to address pressing issues. There was, though, a sense in which the academic members left the decision making processes to the academic leader and administrator. But this was seen to be suitable for the way the subunit worked.

#### *Process of decisions about model construction*

In ENG1 the decision making structure had not previously been formalised at any of the levels. Issues for evaluation were taken up at a central meeting for the different programmes on offer. These dealt with evaluation in a more ad hoc way, appearing to draw out responses as issues cropped up:

*[There is] an internal group for, in the [NN – org name] we have a regular meeting of the MAs, there's [x] MAs which are round about the same academic area so we have that there. And basically the Course Director will ask for feedback on any of the issues that people can help with and try and implement, but there's no kind of formal procedure, that happens more through the annual review in the external department where issues are taken up and addressed on the decision making level through regulations and things like that...so everything has been kind of developed as we go along. [ENG1h].*

However, following on from external demands to the organisation, the amount and focus of centrally devised evaluations was beginning to leave little time for evaluative discussion and decision making at the subunit level. Despite these extra pressures, the subunit members did seek within their programmes to try to evaluate their evaluations annually and setting aside some time would be for discussion:

*Well, always every year we look at our evaluation, in fact we changed it, as I think I intimated earlier, we've changed it this year and we look at how we do it and whether it's effective and whether we can do it differently, so it's not static. [ENG1a].*

Combined with the internal structure of the subunit increased evaluation demands had brought additional challenges to the decision making process. Despite the routine and bureaucratic nature of the central evaluations there were

attempts to make programme specific changes to the subunit frameworks. In ENG1 one respondent discussed how there was an attempt to widen the focus:

*So what we are trying to do, what we've been trying to do in the last, I don't know, 8 or 9 months is try to have a wider approach to evaluation than we've had before and not just rely on X or Y, but rely on a wider remit really.* [ENG1a].

In ENG2 there were similar responses regarding the amount of time available for cross-unit discussion. Respondents noted that there were many different methods employed in the evaluation of the programme, from trying to interpret the central evaluation and student feedback to evaluating the programmes and supervision. As was seen in the previous chapter this variation showed a widespread understanding of evaluation and its purposes. While all of these processes were in place the respondents spoke little of decisions made across the subunit. There appeared at first to be some disagreement over this issue within the team. One respondent referred to the team being able to add programme specific questions to the student evaluations. The respondent had been able to add questions related to ascertaining student perceptions of programme impact on their own organisation. These questions had been useful to identifying how the programme was coming and evaluating its development with regard to a subject that was becoming of increasing importance. There was a sense in which members of the team were taking on board such topics of interest into the evaluation processes. At the same time the subunit members were very interested in the general subject of evaluation and were experienced as external evaluators in many different contexts. Another respondent considered the importance to subunit members of reflection over the concept of evaluation:

*But the question of how to evaluate and the basis upon which you evaluate has always been one that's been of interest to us as a group, and we do see that as being a problem, it's a problem for the people who are engaged in the process, you have to decide on what basis you are making a judgement on the quality or otherwise of the subject of your evaluation and the reasons for that and the rationale for it* [ENG2m].

But despite this common interest the subunit members had in many ways ceased to further discuss this across the group:

*To be honest there are very few discussions about evaluation – although we are talking about having a session to which we invite people / practitioners outside the organisation to brainstorm new Masters Courses for educational leadership. I appreciate this is not really evaluation.* [ENG2p].

In addition, due to the financial pressures on the organisation as were outlined above, focus had to be placed on evaluating the initial implementation of

programmes, and this meant that over time little emphasis could be placed on more formative evaluation:

*so by the time a course comes to an end it could be [many] years since those materials were first pulled together and there might have been a little bit of updating but nothing major. [ENG2n].*

The same respondent went on to reflect how this lack of in-depth evaluation could also lead to a misinterpretation of data that had been gathered. The respondent felt that such development of the evaluation process was necessary to avoid a decision making process based on the superficial data that had been collected, particularly with regard to the interpretation of what strong levels of participant satisfaction might indicate, particularly in terms of the quality of programme on offer.

*So there are also comments about the lack of current materials within the courses. I mean there is also lots of great stuff about how fantastic the tutor was and how comprehensive the materials were and how they stimulated people to think about their practice and so on, but there are critical comments which I think need to be picked up. And so I don't totally subscribe to the view that I have heard here that "people are happy on the courses", I think at some levels there is a mismatch between what we are offering and what some of our school practitioners are engaging with in their schools, and that's something that we haven't totally taken on board. [ENG2n].*

Within the English subunits the processes surrounding evaluation decision making also appeared similar to those expressed in NOR2. Respondents did not report to discuss processes in the way that members in NOR1 had discussed collegial processes. Respondents focused more on the parts of the evaluation systems, discussing more the frameworks put into place by the wider organisation and the tasks performed at subunit level. Responses from both subunits covered reflections over the routine-based nature of the evaluation process, evaluation model construction and discussions taking place about models.

#### *The impact of organisational structure on decisions*

Although this study is not essentially focused on the utilisation of evaluation within the organisation, as was seen in chapter 3, utilisation and perception of it is suggested to influence decision activity. In the English cases it was interesting to see how various members of the different subunits referred to how responses from levels above them made an impact on their attitudes to evaluation and their related decision making. The subunit members at both ENG1 and ENG2 discussed how they faced challenges in terms of their evaluation structures in relation to the frameworks in the wider organisation. Again within the context of

a loosely coupled organisation the groups also appeared loosely coupled internally with regard to evaluation. However, in ENG1 there was a sense in which there was more discussion across the subunit in this area than there was in NOR2 and ENG2. It was rather a problem of time available and overload that the subunit members expressed themselves to have struggled with. This is not to suggest the same pressures did not contribute to lack of discussion in the other units, but this was an area highlighted in ENG1.

In section 8.3 an example was drawn from the subunit at ENG1, where the respondent referred to the professionalism of the course team as the important driver for improvement of the programme, through within unit evaluation, referring to some discussion over programme implementation. The respondent considered that additional to these processes were the mandatory tasks related to accountability frameworks that were perceived to add nothing but yet had to be done. These processes were mandatory and considered bureaucratic. However, processes at the course team level offered opportunity for internal discussion about programme implementation and development. The balancing of these demands was not straightforward however and a cause of tension as the teams attempted to decide what to do. Due to the structure within the subunit each member of staff had responsibilities for different programmes but yet taught on each other's programmes. This meant that discussions took place surrounding the evaluation of the programmes at course team meetings. It was here the pressure of time and of demands caused problems for the team in discussing further about evaluation. One respondent related how the course team had good debates about processes but often did not get past the demands from the wider organisation and external bodies. This had left subunit members feeling a degree of cynicism about the process:

*I think at the Course Team level, and I'm not cynical about that... I think at the Course Team level, that's the forum in which you have the debates about what's working and what isn't, where we are going and how we need to improve, but I do think that associated with that there are lots of bureaucratic procedures which don't really add very much to the process. I think that's what I am saying really; they are necessary evils you have to sort of do because Government and Regulatory bodies require regular reports on quality, when in actual fact it's not adding anything, it's just a kind of accountability process. [ENG1g].*

Similarly in ENG2, the structures for evaluation within the wider organisation had appeared to hinder the debate concerning forms of evaluation within the subunit. Despite having opportunity to prepare questions for part of the central evaluation, little input was now forthcoming from the subunit as a team. One respondent noted that there was little internal discussion or debate about these issues, where the process of decision making appeared absent. As will be seen below the monitoring and assessment side was still the task of the subunit, but the central evaluation had been seen to become a routine and the group had

become decoupled from it, even though it might have provided opportunity to explore further some of the issues surrounding course development:

*When you look at the external evaluation there is a standard evaluation questionnaire that goes out from [NN – Institute title], we are however given the chance to add Module specific questions to that, which would allow us to try and address the kind of questions that we're raising there about purpose and achievement, but the kind of debate that I've indicated would typically not take place, once the Course, it's about Course creation and as I indicated once the Course is produced the team changes and there's just a couple of us who take responsibility for seeing it through. [ENG2m].*

The same respondent went on to express how the purpose of the evaluation process had also changed over time. It was noted that generally the “evaluations are fairly fixed term, [and] fairly fixed in their form, partly because so much of it rests on the monitoring and the annual course report”, however due to financial pressures on the faculty the subunit was placed in, emphasis had changed much more towards budgetary control. This had tempered the evaluative work taking place and meant that decisions were much more strongly based on viability of programmes rather than quality of content. This had particularly affected the master programmes as they did not carry the same funding as undergraduate degrees. This shift in emphasis has created scepticism amongst programme members to the evaluative process, feeling that decisions at higher levels were not based upon reports and findings from the subunit.

In these groups, then there was a sense of what can be described as “dismissive submission” within the framework for evaluation. In many ways the subunit members in NOR2, ENG1 and ENG2 expressed a resignation with regard to implementing the demands placed upon them, even though they often disagreed with the focus, character, volume and ultimate utilisation of the data collected. This meant that decision processes with regard to the quality assurance system were characterised by ritual responses, where data collection was less likely to be thought relevant to the within group processes related to programme assessment and development. Furthermore in England the respondents, while agreeing to the importance of evaluative systems, struggled to accept newly developed standards and demands for instantaneous impact related to the subject field and found it hard to build these processes into their working methodology. As will be seen in the next section, decision responses made in NOR1 appeared to contrast with those in the other subunits.

Theme	Descriptive element
Dismissive Submission	Collective inaction
	Central initiation and download
	Critical rejection
	Decoupling and disengagement
	Individual interpretation

### 9.3 Decision responses to demands on the subunit

The previous section dealt with the perception of the decision processes within the subunit, and the extent to which they took place in concert. While this section on decision responses overlaps with the previous one, the focus is placed on responses to demands placed on the organisational unit. The data offers an interesting comparison to that presented in the previous section, as respondents were asked to consider the pressures placed on them related to evaluation and how they had responded. The framework of demands was outlined in chapter 7 and was seen to be both internal and external to the wider organisation. The data presented outlines in more detail the increasing workload and challenges to time available that respondents had referred to in the previous section. Focus is firstly placed on reflections over the quality assurance systems that the wider organisations had adopted, their impact on the subunit and how they had responded to allied demands.

#### 9.3.1 Responses to internal demands

Emphasis in this section is upon the evaluation frameworks within the respective organisations. Discussions surrounding the decision responses of the subunits were based on the perceptions of the goals and intentions of these frameworks. Respondents' perceptions were related to seeing the new systems as obligatory to follow, increasing their workload, heavily bureaucratic but being reductionist in nature and adding little to the knowledge base.

Decision responses	Related issues (interlinked)
Internal evaluative configuration	Reductionism
	Increased bureaucracy
	Overload
	Legitimatising
	Institutionalisation

As was outlined in chapter 4 a result of the Bologna process has seen HEIs required to implement quality assurance systems that satisfy the "Standards and Guidelines for Quality Assurance in the European Higher Education Area" (ENQA, 2005). As a result systems across the HEIs in this study were seen to bear an isomorphic like resemblance. Interestingly within the Norwegian subunits much more focus was placed on discussion of impact of NOKUT, the Norwegian quality assurance and accreditation body, whilst the members of the

English subunits rarely referred by name to the QAA. But across the different subunits there was agreement that the impact of quality assurance systems had required a response from all organisations. However, there were noted to be different initial requirements for programme teams depending on both the status of the HEI they were placed in and whether the programme had been put in place pre or post Bologna. As one respondent from NOR1 put it:

*The demands from NOKUT apply to everyone. But it's clear that the organisations, the colleges and universities, come out of the process differentially dependent upon how far they've come, so there is variance [based on organizational characteristics and history]. [NOR1i].*

When asked about their response to these demands related to development of the programme and quality assurance, another respondent at NOR1 noted that the majority of their plans and procedures for the programme and decisions about evaluation had been put in place prior to the Bologna related "Quality Reform":

*[the demands] actually came a little later, because of the Quality Reform; those pressures came a little later. I remember we spoke about this in our group, about whether we actually had procedures and routines in place from before. So for us it wasn't just the quality system, but also in relation to supervision and follow up of students, much of it was in place already. So it wasn't much of an adjustment really [NOR1f].*

As was seen in the previous section, the subunit members from NOR1 reflected over both their interest in developing and using evaluation and making decisions within a collegial setting. As a result these processes were already in place before the demands for quality assurance came. Later the subunit had used time to understand the new demands placed upon them and how they fitted in with their practice. Interestingly though, despite their being a sense in which these processes were already in place there was a sense in which they were still experienced as demands upon the subunit. The members continued to try and influence these frameworks where they were not in agreement. There was a sense in which they had lost control over evaluation where they perceived the purpose to be changing. The members of the academic group desired to change what appeared to be developing into a ritual:

*The systems just have to be followed in a way, but you can to some degree influence them. You can influence how you implement a formative and summative evaluation. But I feel that it's more about us showing them that we are doing something, it's that they are interested in. [NOR1c].*

But there continued to be a lack of knowledge whether or not the evaluation reports had been read and what information was really considered as important. One respondent who had had a leadership position outlined how the formal evaluations with their more general data presentation were those sent to the



higher levels. These included information about student numbers on the courses and throughput. This data was considered to be limited. The more in depth data collected was only used internally. The respondent reflected that the Institute was pleased with the programme, but their measure of success was the general data on production rather than on qualitative data:

*We have a high legitimacy in relation to quality and the number of students etc. But this isn't something they observe based on feedback, they are looking at influx. The leadership of the Institute doesn't go in and read our feedback; I don't think so, because we don't actually present any of this data to them. We just use it internally. [NOR1g].*

The subunit members in NOR2 also referred to the fact that they abided by the formal demands perceived to be coming from NOKUT. Although these demands were not seen to be a major problem for the respondents, the move to greater formality was considered to be time consuming and brought challenges into their ordinary master programmes and commissioned programmes. In addition there was an increased bureaucracy within the wider organisation that they needed to relate to:

*We are also at the mercy of the bureaucratic system within the organisation. ... This has been a nightmare in all the years that I have worked here. I mean, I came from bureaucracy but my experience is that [this organisation] is much more bureaucratic than that; we felt we had room to manoeuvre there...*

But the respondent felt that meeting all these demands had begun to take its toll on the subunit, even though they were attempting to respond by adapting their processes:

*We satisfy the formal demands but we use too many of our resources for my taste. This affects my capacity to work; it affects the profitability of [the organisation] and our capacity to sell the programmes. But we are learning, we are a learning organisation and the organization will streamline its quality assurance systems and I'll get to employ more members of staff to run things... [NOR2e].*

In England there were similar reflections. In the subunit at ENG1 discussions surrounded the mismatch between the ideas of the academic staff and how they would like to evaluate, and the systems that were now in place. One respondent outlined how they would have liked to adopt a more innovative approach to assessment and evaluation within the programme but recognised the difficulties in attempting to introduce and change something like this. The respondent therefore had accepted and followed the “normative procedures” within the “confirmative, bureaucratic processes” of the organisation, while at the same time attempting to balance this with a more liberal approach with the course participants [ENG1b]. In ENG1 another respondent exemplified how the

increased external demands and pressures had caused the subunits to adapt and refocus their evaluation activities. The respondent outlined how the external demands for data had forced the subunit to move away from an informal system based on their professional experience and collecting general participant feedback to a more formal system. The new workload had to a degree stopped them doing both. But more than just more formally and systematically gathering data regarding perceived satisfaction and usefulness, there were additional demands to gain impact data, despite the complexities and difficulties such activity involved.

Another respondent from ENG1 reflected over an evaluation seminar that the institution had organised as staff began to feel uneasy about the amount of time they were spending on evaluation combines with the lack of information they were collecting from students in particular. The amount of feedback required was still considered to be inhibitive in terms of data collection. Having changed to a formal system it seemed like they were gaining less information than before. The aim of the seminar had been to develop new techniques, but the respondent noted that they were again given more tasks than they could feasibly implement:

*...the problem is of course, that the first speaker in this session had all of these ideas about different ways to implement feedback systems, and all you ever do is evaluate the course, and all of your time is spent evaluating and so we obviously haven't got the time or resources to do it so we have to try the ones that we think can work without too many resources going into it. [ENG1h].*

Another respondent had recognised the challenges to trying to meet all the various demands for evaluation within one model, within this limited time frame:

*I think that there's a multiplicity of purposes really and because there's the institutional purposes, which are very bureaucratic and they require various things under certain columns etc.; there's our own teaching and learning purposes, in that are we... delivering this to the best possible standard that we can possibly do in this particular format, because it's quite challenging; and then the third things is, because it's specifically ... about applied educational leadership and management, is all the assessment that we do on the course actually helping our students to apply what they are learning theoretically into their own places of work wherever they are... [ENG1a].*

I will deal with the particular purposes raised by the respondent in ensuing sections. But the main point of importance here was that the combination of all these purposes had become a deeper problem as the system had become more rigid. When the system was more internally grounded, the academic members of staff could use time to sort the various forms of data and proceed to deal with them accordingly. The new systems offered little flexibility for programmes that didn't fit in with the overall pattern for evaluation. The respondent noted that

this was also highlighted by the fact that a master programme was delivered as distance learning, which led to additional issues, one of which in this particular case was that it came under more than one framework within the wider Institution. The systems of course were not uniform, which meant the respondent had to report more or less the same information twice but in different presentations:

*I don't know if you realise, but my course actually fits between 2 institutions... so we have to fit into their guidelines as well. In fact that is what I was [just] doing... I was just reading the annual review and seeing if it I had to make any changes to it...They do more or less dovetail together, although the formats are different; we just have to re-jig the information for different formats. [ENG1a].*

The respondents in ENG2 had similar experiences, recognising that the drive to formality had left little time to reflect over evaluation in a way they had been able to do previously. The respondent described the academic year as now being “punctuated” by “moments” of evaluation which needed to be adhered to, but reflected further that the group was trying to begin to think about how they use these processes to influence the formal evaluative frameworks, noting that “it’s also about using those as levers for change isn’t it” [ENG2n].

The various subunit members were collectively sceptical to new ways in which quality might be defined and published within the formal frameworks for evaluation based on external demands for quality assurance. There was a growing concern about who would have access to such information and how it might be used, and across the subunits respondents were therefore defensive in terms of the format of evaluation activities. With some adjustment, as well as the continued freedom to comment on the data, there was a suggestion that they might see some improvement. However, currently the frameworks could only be seen as being utilised for control and legitimation purposes:

*In a way it's all about it being very much simpler and it's a system, and we are very much taken by systems and structures. And the final evaluation is part of that as a legitimating aspect to check that we have done what we are supposed to. It's nothing more than that. [NOR1e].*

With regard to a question about the form an evaluation takes, in NOR1 it was reflected that the demand to evaluate had to be adhered to but that there was an opening to discuss how much effort should be placed on doing it and what impact it would have on programme development. Despite the general agreement across NOR1 about evaluation they might disagree over the extent to which policy should be implemented:

*It's just part of the Quality Reform that has been forced upon us and that we have to do it and just accept the fact. We can of course discuss how much effort*

*we should put into it and how much importance we should give it. And on that point we are a little in disagreement.*

The subunit members were more concerned about the appropriateness of task rather than just rejecting further tasks. The systematisation of the data into a more formal system brought further challenges. Respondents also referred to the limitations of the new system. While they were being asked to report more, there was a growing impression that the type of data required did not reflect the idiosyncrasies of the programmes and present a fair picture of progression. With the relative reductionist nature of the reports sent upwards in the system respondents across the subunits spoke of the importance of reconsidering what the purpose of evaluation was. These factors were exemplified in a response from ENG1, where the respondent spoke of the way data was now to be presented more quantitatively. The respondent from ENG1 was outlining the framework for evaluation and quality assurance, discussing how this focus on presenting data as statistics was limiting in terms of how to present the programme. Instead of presenting a qualitative summary of what had been achieved and an evaluation of the programme in terms of its progress, the shift in focus meant that one was always commenting on the negatives or areas of underachievement and this was considered to be problematic:

*You have to give a commentary in the annual report saying how many students have withdrawn, failed, etc, etc, you have to give the statistical figures. And obviously if they are looking bad then you have to give a commentary suggesting why this might be the case. [ENG1a].*

The frameworks were felt to be more of a control mechanism than evaluation system. Another respondent from NOR1 had also experienced how information to be used internally with course members needed clarification when summarising and reporting to higher levels. In this case the respondent had sought and taken the opportunity to explain to the Faculty board the background behind comments that had come from programme participants, offering an interpretation that the format otherwise didn't allow for:

*I have been part [of a group] that has presented comments related to the annual report to the Faculty, because there can be types of information that I feel are fine to share with course participants but that can be interpreted differently if you haven't been part of what is going on and don't know why the relatively critical assessment I ask students to give has arisen. [NOR1i].*

This was noted as balancing the different purposes of evaluation within the organisation, for control and for development. But as will also be seen in the section below on external demands for impact, the Norwegian HEI system was recognised to traditionally allow a great deal of freedom, but only as far as there was perceived to be relative success:

*It has always been like this in the Norwegian University system, but it's clear that you don't have that mandate if it goes wrong [NOR1i].*

Another respondent recognised that this could develop tensions; raising the problem of loyalty to the system against the individual's own view. But, there was still semblance of freedom within constraints, and the constraints seemed more about doing the task rather than detailed control:

*It's my opinion that when we are working in an institution like this then of course you need to be loyal to what the authorities say, if you can't do that then you don't need to work here. It's the same as being a headteacher, it's a dilemma, and you can disagree personally. But I think there is enough freedom to act. [NOR1e].*

This section has dealt mainly with the introduction of greater formality into evaluation frameworks, requiring the subunits to take decisions at programme level of how they would respond. The examples here suggests that greater formality in the system was perceived to have greatly increased workload rather than merely rationalising frameworks, even though the respondents across the subunits did not feel that they were generally getting any qualitatively better data. As a result respondents spoke of feeling forced into coping with the extra demands while attempting to find creative ways to meet them. At the same time they were sceptical to whether information was read by their leaders. The reductionist nature of the data provided did not convince them otherwise. These responses were grounded on the perception that data was not really utilised at higher levels. An important area that the formal frameworks were meant to address was increasing the amount of data from students, giving them greater voice within the system. It is this subject that I turn to next as I refer in more depth to many of the specific demands and issues subunit members needed to resolve.

### **9.3.2 Responses related to participant demands and participation**

As was seen in chapters 4 and 7, student voice and participation have become a key part of HEI frameworks for evaluation. Respondents were asked to discuss their responses to the demand to include students more actively when evaluating the programmes, as well as to consider the demands students placed on them through their feedback.

Decision responses	Related issues (interlinked)
Participants' participation	Reductionism
	Evaluation overload
	Consumerism
	Academic / voice conflict
	Formalisation

Respondents from across the subunits were interested in improving student participation within the evaluation of their programmes. They had registered a lack of participation over time and this had shown little improvement now that the processes have become more formalised. In addition to the challenge of getting participants to regularly participate in the formal procedures, the new systems had also presented difficulties with interpreting the information gathered. As a respondent from NOR1 related, students did not always see the purpose of what they were being asked to do and centrally designed forms were often difficult to interpret:

*But what I miss is [an explanation] of what the purpose of this feedback is. Is it, in a way, to express their discontent or to express their satisfaction, or is it to contribute to the improvement of the programme? They could have done a better job with the form, but otherwise I think they took it seriously... but these are busy people, school leaders and teachers, so it isn't certain they prioritised it. So what does it mean that there was often a low response rate, does it mean perhaps that they were satisfied? We couldn't measure that. And those who answered were perhaps those who were most dissatisfied? I just don't know... [NOR1g].*

This point of difficulty of interpretation was also recognised earlier in responses from ENG1. It was also problematic for the subunit in NOR1 that they were getting so much positive feedback, but in addition they had a problem with the type of feedback that they were getting and needed to report on. The data were only general and not specific, making it difficult to interpret:

*We are lucky that we have so many satisfied students. But there must be things that they aren't satisfied with? But the response we got from our module in the middle of the programme was amazingly, very positive. So we need to challenge them a little, aren't you being a little too kind with us now? You need to be a little more precise, it's all too slushy! It's not always like this though, evaluations can be a little tired, or over positive without anything concrete coming forward. So it's a challenge to get specific comments that you actually do something with. But it's all about how you ask, as well. [NOR1h].*

With so many degrees of freedom the subunit members had struggled to assess the data collected. The respondent raises important issues here, what exactly is being evaluated, and what might the lack of participation suggest?

As was seen in chapters 7 and 8, the institutional QA system of which the subunit NOR1 is a part required that students participate in evaluations of the various modules undertaken as part of their study programme as well as a summative evaluation. One respondent considered the dilemma this had caused and the decisions made to address it. It was noted that should the subunit follow the rules as they were stated, it would require those on the programme to participate in many evaluation activities. The respondent (administrative)

outlined how they had decided to deal with this when realising that the frameworks did not stipulate the form the evaluations needed to take. Therefore they chose an option that would be perceived as requiring less work for the students:

*If we implement a written formative and summative evaluation for every module, then that would mean 6 to 7 evaluations from the same student group within a year. They would never accept that. Put it like this, it would kill student interest in completing written evaluations. So we have chosen, in a way, to run reference groups. There are [central] demands for the topics to be discussed and how frequently, but not really to their form; so we've chosen reference groups [NOR1a].*

Another interesting example of how a subunit attempted to implement the demands evaluating participant satisfaction through the gathering of feedback placed on them came from ENG1. Returning to an account of how one respondent referred to a particular course team working within the subunit that had struggled to operationalize and implement the demand for student feedback, it can be seen that the respondent reflected additionally over what was perceived as a timely intervention from the organisation, as the particular course team had in a way given up on the task which they had interpreted to be too complex:

*Well, I suppose we really only started to think about it, you know we've been doing this and we thought well we were not really getting the respondents, we sort of gave up on it. So then this course came up from the [NN –part of org] which was really very timely. So I went to that with a few other people from the [NN –part of org] and from there I brought back from what I had learned from that meeting to a tutors group with the whole course team, and discussed all of that there. Then between us we came up with various ideas that everyone thought would work for the team, for the group. [ENG1h].*

This course team had also struggled to implement the demand emanating from the central organisational system, finding the requirements to be too extensive. The respondent had also noted from the methods meeting that these reactions were widespread in the organisation and other subunits attempting to translate the same requirement had come no further. These comments were interesting with regard to the decisions regarding the evaluation model to employ. While the framework was there to be followed, the unit were presented with the opportunity to interpret and develop their own models, which was reiterated by the intervention group. But there was still the sense that the purpose of these activities was felt to be about satisfying the system rather than improving the programme. It was at this point that they had realised that they could merely adapt existing models to meet the demands as long as they addressed the issue of providing the necessary information:

*Because all of us thought, you know, well it's great to have all of these ideas but we'd just spend all of our time evaluating, so it's ways that we can actually... and you know the student rep idea has really worked very well. And then one of the other things that the tutors said was if they get any other comments or anything like that, they'll forward it, in fact some of the things we were doing already, but formalising it a bit more so that we keep things together. So in a way, that was discussed in the context of not just evaluation, but how what is fed back in evaluation can be implemented in terms of improving the course [ENG1h].*

The problem of how to use the feedback was an issue that the subunits needed to address seriously. Across the subunits respondents discussed that they struggled with these tensions. They collected more data and were not always aware how to interpret them due to flaws in the system. They still however needed to report back to their students and give some idea of how they would follow things up. These reflections were again described in Norway at NOR1:

*We had done all this before, but first and foremost internally, and we reported on the consequences. But now everything has to be more formal. That's the demands from the organisation. So we have their evaluation forms and we have to report what consequences student feedback will have. [NOR1e].*

*Tensions: What they think and what we know...*

But this leads to the next problematic area, once the data is collected how to weigh up participant demands against the purposes and context of the programme. This issue was described as a challenge between what students think and what the teachers know. As has already been seen, following the demands meant that the subunit members needed to focus more upon gathering participant feedback. The subunit members from each of the four institutions suggested that they were often in agreement with what the students highlighted. But there was also a sense the increased focus on feedback had created expectations in terms of changes to the programme that challenged the basic premise. These thoughts were exemplified by a respondent in NOR1:

*Furthermore it's not the case that we just accept every suggestion or that we can change the programme based on what the students want. There are a number of criteria that decide what we plan to do, not just the experiences of the students... It's also the case that they formulate things in an unfair way. So we always make assessments of how we will interpret these kinds of surveys. [NOR1i].*

Because of this, it was more likely that the main content of the internal programme was influenced more by the interests and opinions of the programme group than by issues raised by participants:



*“it isn’t the students who have said that they want this, in fact they don’t want it really. It is us who think it is right; we, as professionals who think it is exciting and so we put it in”*

There was then perceived to be a dichotomy between what the participants want and what the programme designers perceive that participants need; the rationale for the programme. The evaluation responses are weighed against this formulation. In each of the subunits the respondents noted this to be a difficult decision process, challenging what and how they evaluated, knowing that the results had impact on their relationships with the participants as well as forming reactions from higher up in their system. They reiterated that the problem was not in receiving criticism but rather the expectations for change and speed of it, as well as the overall perceived limiting of their professional judgement. In NOR2 one respondent also reflected over how this evaluation data was weighed against self-evaluation of how the programme was going:

*I don’t look at anything else; I’m not really interested in it. Because if I have them where I want them, then I’m safe. Whether the reports focus on organisation or how much they liked the lecturers and all that, that’s not important. I use evaluation to a greater degree as a point of reference to see the overview: is it going well, ok or poorly? If it is only going ok or poorly then I do something about it. If it’s going in that category that they say that it’s going well and I know it’s going well then I don’t do anything about it. [NOR2g].*

Another respondent followed this up by recognising the context for data interpretations to be a professional one:

*So we are constantly discussing this, how are satisfying the students and contributing to their learning. So we are interested in learning output, but not necessarily as they feel it there and then, but rather how we think things should be. [NOR2e].*

But this approach also left its challenges with a respondent recognizing that the subunit was poor at utilizing evaluations in programme development as well as poor at discussing with students:

*“You’re right, it is a weak point, one of the weak points. Evaluation just gives us some feedback that something should have been better here, but that is enough for me really. I get just the barometer I need; my starting point to do something. But it doesn’t tell me much, maybe if we went into a process it could help me to do something better. But then we would need to go into a dialogue and we don’t have a very good tradition for talking to students”*

This meant that despite the intention of increased voice for participants, the main focus of the evaluation was more often than not the academics themselves. Another respondent recognised that this contributed to the process becoming

ritual like, but this was also based on the experience that the responses from the students were more or less the same, year in and year out. And as the respondent pointed out “you start to ask, how much time am I going to use on this?” This comment was reinforced by another member of the same unit:

*I think maybe it's a lot of work for very little result. Is that awful to say? I'm not saying that student voice is important, not at all, and it's very interesting, I just think we could have a tool where we can see what they actually say and how many say it, and that would provide the basis for a good discussion. [NOR1e].*

The interviewees generally did not feel that they used the data and feedback to improve the course as much as has might have been. Discussions and decisions about the programme improvement were taken in a more summative fashion, and based against an already existing plan in the organisation. This was either as a result of or in addition to the fact that there was disagreement about how to interpret them, and subsequent changes would be applied to successive rather than current students. The interviewees were not saying there is no feedback, but rather that one might have expected the programmes to have given more emphasis to this given the subject area. This was recognised by a respondent in NOR1 when asked to consider what evaluation influenced the programme:

*So it's really as much about the interests of the academic staff forming the programme. But these are influenced by meeting the students; definitely. So we can be influenced. Because of the student group we have with active school leaders we have to be aware of what needs they have. So we have to balance this constantly; what is useful for them to make them better school leaders.*

This section deals with an interesting area of reflection with respondents focused upon the interpretation of evaluation data. Discussion developed with regard to instances where student demands, that might change and develop as the programme progressed, conflicted with parts of the underlying rationale of the programme and the way the subunits evaluated. This draws the discussion towards the basis for evaluation and perception over programme approaches and their basis. This is also linked to idea of academic anchoring, which is outlined in section 9.3.5 below. I turn next however to deal with the issue of decisions made about impact evaluation.

### **9.3.3 Decisions about ascertaining programme effects and impact**

Beyond the requirement for greater systematisation within the organisation, as was seen in Chapter 8 the subunits in England faced increasing external demands at policy level and from funding agencies to begin demonstrating programme impact. At the same time it has already been shown that many of those interviewed had worked or were working with the NCSL on the evaluation of their programmes as external evaluations. Respondents had noted the demand to approach the issue of impact and value for money in the tenders for evaluation from the NCSL.

Decision responses	Related issues (interlinked)
Ascertaining programme effects	Impact focus
	Formalisation
	Academic resistance

Respondents related how discussions about this issue were coming on the agenda more often, and now not only were the members forced to think about the issue when acting as external evaluators, but also in terms of their postgraduate programmes. When asked whether this area was becoming more important and whether they were at the stage of making firmer plans and decisions about it, one respondent related how they were considering how to follow up their students with longitudinal data collection, but that current practice was still very limiting:

*we were talking about this at a meeting the other day, about going back to our students so long after the course and talking to them in terms of impact and how you measure impact and all that, there's a huge amount of literature on it as you know. But, I mean, from my personal point of view I look at impact in terms of their development over the years that they are on the course. So, from my point of view I can't judge whether they've had an impact on their school, or their college, or whatever it is their involved with ... I know that the course team there are having the same sort of discussions about that more generally [ENG1a].*

The various team members recognised this would mean a change in emphasis of their evaluations; moving from measures of satisfaction to impact reflections:

*Yes, well that's the other question we've thought about... It actually addresses students who have already finished and we did talk about that. ... the evaluation at the moment isn't just about the experience, it's also about how useful and how practical have you found it because they are all practitioners anyway. ... We did think about [doing] that in future; having a follow up 2 or 3 years afterwards, seeing how has it impacted on your career, or your life basically, doing this here. But we haven't got further than thinking about it yet. [ENG1h].*

In section 8.4 it was noted that when considering the structures and approaches of the current evaluation model in ENG1, one respondent considered that the subunit had relied too much on intuition, or “gut feeling”. However, the respondent also felt this was justifiable considering the responses they had received from programme participants. It was recognised that the outside pressures had led to rethinking in terms of the evaluation models, especially with regard to impact data, but there was still a distrust of models and a tendency for academics to want to place emphasis on their own professional judgement instead of the collection of the so-called “hard data” that were becoming sought after by external bodies. Building upon reflections outlined

earlier, one respondent was asked for the reasons why it would take longer to develop more sophisticated evaluation models for postgraduate programmes. The respondent outlined two main reasons, competence and model design:

*I think probably lack of understanding, I think lack of building evaluation in at an earlier stage and I think this UCET<sup>180</sup> discussion has kind of helped to sort of focus people's minds much more about ok, we do need to start getting data about...[ENG1g].*

The respondent recognised that their decisions were based on upon the competence they currently had in this area and that the postgraduate programmes were not organised in such a way that would easily enable such data gathering. But the respondent also recognised that the discussions were now more on-going, and that it was becoming hard to ignore the demands, especially as they were tied so closely to funding.

*But the TDA<sup>181</sup> with the funding that they are now giving these programmes have really made universities think much more sharply about what impact, if any, their programmes are having in a way that's suggested, what back at the ranch really, because we haven't really done much on that....*

The tension between using professional judgment and requiring “hard data” was noted across the four different institutions, even though in the English setting it was now becoming more heavily tied to issues of impact and the members perceiving a clear demand upon the decision about what and how to evaluate. But in ENG1 it was also recognised that to perform such evaluations the need of extra resources by both individual personnel and across the team would be significant. There was an extent to which the process was being resisted as far as possible, until a solution that would satisfy the various groupings was worked out:

*I think people are reluctant to take it... perhaps reluctant to take it to higher levels because there's all sorts of implications and I think there's a limit to how much time you want to spend on doing impact evaluations; [people generally are asking] why have we got to spend such a huge amount of time? It's a matter of getting a sort of lean, cost-effective process by which you can do impact evaluation and I don't think we've got there yet. I think we are getting there but I don't think we're there yet. [ENG1g].*

At the same time there was also a sense in which those programmes not requiring external sources of funding had not fully dealt with the growing pressures in the environment for impact data. When asked about how the external demands for impact had influenced the programme and caused a change

---

<sup>180</sup> Universities Council for the Education of Teachers

<sup>181</sup> Training and Development Agency

in the way evaluation was approached, one respondent from ENG1 noted there to be an institutional inertia

*One of the things is that I think if you are talking about universities you are talking essentially about very conservative organisations. And so I think by and large innovations in assessment come quite late to universities. I'm sure you'll find individual examples which don't accord with that, but, by and large, they are quite traditional. [ENG1b].*

As was seen in section 7.1.1., the response of the group in ENG1 was try to adapt models using “creative thinking” about a process that was considered to be “naïve and unsophisticated”. Similar discussions had taken place within ENG2, and there was an increasing interest and curiosity in the efficacy of the programmes on offer, while again recognising the complexities of such an activity:

*I think longer term impact is also important and you know we've looked at that a bit, but it would be interesting to do more research on impact some way down the line, of course that's difficult to disentangle from other variables, but I think that's the 64, 000 dollar question really of any CPD provision, whether it does have any long term impact and usefulness in people's professional settings. [ENG2k].*

The problem, according to the respondent, had been that the process was rushed in too soon as a political decision before it had been fully tested and assessed. In developing models to assess how participants had implemented what they learned from the programme they too felt the emphasis on understanding direct effects on school outcomes were based on poor understanding of methodology, which was however beginning to change:

*I think particularly earlier on the external pressures were perhaps a bit, what's the word, perhaps asking for rather simplistic lines of causality between professional development activities and direct effects on pupil performance, and I think that has come from our Department for Education and Skills as was, and other Government bodies, but I think more recently there has been an appreciation, both externally and internally, of the complexities and subtleties of tracing the effects of CPD on practice and on pupil's thinking and their organisation, yes. [ENG2k].*

Despite their reservations, the subunit had chosen to develop a methodology that would begin to address the issue, but this was still in its initial stages and mainly related to the project work produced on the course, comparing pre and post programme data collection based on individual reflection:

*So the main tools that we are kind of using to look at what impact does a teacher studying this course have, is that we look at the projects they do and what we are trying to do is to frame those projects slightly differently so that there is*

*increased emphasis on the students, having undertaken a needs analysis at the beginning and then revisiting that and actually talking a little bit about what it is that they feel they personally gained so that we make that much more explicit. [ENG2n].*

As was seen in chapters 7 and 8, the discussion about programme impact had not yet come to Norway in the same way as in England. However as it was becoming a key issue in the field of school leadership I asked the members at both NOR1 and NOR2 what, if any, discussions and decisions they had made considering these developments. Interestingly because the pressures had not yet materialized as demands the subunit had begun to discuss the issue and consider how it might be done, particularly as members began to feel that they had not really been able to research key theories of the programme in the practice field:

*Take the idea that you get a more analytical view from better knowledge of an issue. It's not certain that it makes you a better leader in practice, but I think that we believe that if you have more knowledge and an analytical perspective on things then it increases your skills set and ability for action. When you make decisions, for example, you'll see more alternatives for how to do things. But we haven't measured this in any way. [NOR1f].*

As another respondent put it “we don't know if they become better leaders” from the programme, “we can believe it, we can suppose it might be true, and we even know some who are better leaders, but we haven't followed it up”. But there was again then a curiosity to see how programme theories were applied in practice and what “effect” they might be having. The idea had now reached the group discussion and focus would be placed on developing a research project rather than merely some kind of evaluation tool:

*We've talked about it in our academic group, that it would be interesting to do, but it would require a separate research project. [NOR1f].*

But this idea had become more relevant now that the first students who had completed the programme were now back in the practice field and the subunit felt the first response would be to see if and how they were applying the knowledge they gained on the programme. But the respondents were very clear about the way this process was not to be approached. As a team they were strongly skeptical to the ideas linked with impact that they had seen to be developing in England:

*[I struggle with the idea] of measuring an effect in itself, that is, to take a leader who hasn't studied here and compare them [with one who has]... but it would be fun to follow a group of students, interview them before they started and follow them up, related to how they reflect upon their own roll of leader, before and after, that would give just as good a picture. If you try an effects measurement then there are so many others things influencing, which maybe have nothing to*

*do with this at all. But everything is about measuring objective effects at the moment.*

In NOR2, however, the members were a little more open to attempting to measure effects on practice and there was a disappointment that there had currently been too little focus both from the policy makers. As a result the subunit members had taken the initiative to develop a framework to be used:

*I think that we know too little about this: And we've tried to start a research programme into this, but it's not an area that gets so much attention in Norway [NOR2g].*

They were also interested in seeing some kinds of effects of their programmes, and whether practice was improved but were more open to measuring different variables from the academic field itself on improving models to be used

*We need to show more interest in looking at the effects of own teaching. Is there any impact from the school leaders we educate, do they improve, that is does our programme lead to any value added effect for school leaders? And there has clearly been too little research on this... and very poor research at that. [NOR2e].*

When asked whether such measurement was possible, another respondent answered "yes, of course". But they were agreed the members of NOR1 that this was a research project rather than evaluation issue. Rather than avoiding the variables the subunit had discussed how they would initially approach the issue, first by identifying and recognising the complexity of variable patterns and measuring them within a quantitative econometric model while accepting that links between the varied processes might appear weak. The intention was to develop these models over time. It was interesting though that the members of NOR2 had decided they wanted to follow patterns similar to those developing in England, taking their inspiration from there:

*But it's clear we think that this type of research is lacking and want all the data on the table such that it could be used to improve school leadership training. And we know there are a lot of programmes for school leaders in Norway that are not interested in getting the leader to influence the value added... [NOR2e].*

### 9.3.4 Decision responses to commissioners

Decision responses	Related issues (interlinked)	
Commissioners	Balancing tension	Bargaining
		Buffering
	Adaptation	
	Academic anchoring	Projecting competence
Protecting professionalism		

In this section I outline the various responses taken by the subunits as they found themselves in a position of having to balance demands between their Institution and the external bodies commissioning their programmes, as well as giving greater voice to programme participants. There was great variation between the bodies they had dealt with in terms of competence and subsequently interest in the implementation and evaluation of the programmes. The subunits had often been involved in bargaining processes while attempting to buffer the demands from the task environment. Interestingly, the result of greater commissioner involvement had meant in NOR1 some renewed reflection over evaluation models and the possibilities for developing more formative approaches. A subunit member reflected over these processes recognising that there was often a greater external challenge to programme thinking due to the proximity of the commissioner to the process at hand. But the members felt that they could base decisions on their professional basis. As has already been seen this was considered to be relative to the degree of commissioner competence. It was here the decisions were affected by a negotiation process, but the members saw this as positive, exhibiting a dynamic part and propensity for feedback:

*If we consider the in-service programmes, because I think this is really interesting to have alongside the Master, because you are in close contact with the school owner. There is a 'triangle' here, with the school owner and their demands and expectations, and us as professionals, and so there's also the academic field, that is, what we know. So we are in discussion or negotiation if there's something we don't agree with. [NOR1k].*

It was therefore recognised within NOR1 that the multiplicity of demands and constraints made the roles difficult to distinguish between and relationships hard to map:

*So there are many different relationships that affect this, it's not just the internal and what we do here. This is the action plan for the programme, and if you look at the point about education you can see we have a separate area the focuses on improving quality and throughput. This is an area all educational programmes have and if you look at the strategic plan for the institution you'll find the same thing. [NOR1k].*



There was also the impression that a response would still be discussed within the programme group as to the extent that the format would be adapted to suit external demands:

*So then it is a question of whether we adapt the system such that it matches their criteria [NOR1i].*

It was felt that the design and model for evaluation employed currently did meet those needs however. These responses are interesting. Members of the group reflect that the evaluation grew out of a perception that greater external assessment demands would develop, but at the same time the group appears to have acted to shape their response to fit their own goals, adapting the model. The initial model has been developed incrementally to fit to the demands that have come, the impression being given that the main aim is to suit the programme groups own needs as well as satisfying those of the formal QA system. Responses suggest a high degree of interaction and cooperation across the group, and tight coupling at the micro level. In receiving reports from administration as to the general demands within the system's hierarchy the group can then adjust their model to meet those needs.

In NOR2 it was recognised in one case that greater involvement of commissioners had led to a more constant form of external - formative evaluation, where the was group required to negotiate more and reflect over the student experiences, with the impression of being resource dependent. Again the image of a triangular relationship was used, but here because of the role and influence of student evaluative feedback on the commissioner:

*[commissioned programmes] are much harder; there's a greater focus on them. This is partly because we are not just relating to students but also to commissioners, but also because the commissioners are relating to their students. So if the commissioners' students are dissatisfied, the first ones to hear it are the commissioners. So we have to negotiate this Bermuda Triangle [NOR2e].*

Interestingly though combining the basic organisational evaluation methods for postgraduate programmes with the commissioners' evaluation had provided the subunit members with much more insight, so despite a potential reduction of freedom and greater sense of accountability, the extra input was seen as helpful:

*We use standard methods for Master programmes here. But with our extensive contact with the commissioner... they have their own evaluation of our programmes. So we get to be thoroughly evaluated. [NOR2g].*

And this relates to the reactions from the respondent in NOR1 above, concerning the basis on which the subunit members responded to external demands. In the previous chapter it was recognised by a respondent from NOR2

that the subunit needed to balance tensions arising between selling a programme and being part of a particular institution. And related to the comments above, there was a level at which the subunit members would not give in. The respondent developed this point in terms of responding to tensions, relating demands and complaints to their own academic identity:

*And clearly we can't, even though we might wish we were more operative, we can't change our identity, it's not possible. This organisation as an institution, where I'm a researcher, with my position – I have a ballast, and if I can't use this insight and understanding in the programme then there isn't a programme, I wouldn't get involved. So clearly we set the premises [NOR2g].*

This idea links to a concept of academic anchoring, outlined in the next section, which is seen to be part of the decision response of subunit members to external demands and pressures for evaluation.

#### **9.3.4.1 Academic anchoring**

Across the four subunits there was reflection of the role of professional identity in the decision process concerning evaluation. This was tied to initiation, design, implementation and interpretation of data. Respondents spoke of the importance of their academic role and experience when make decisions. Some brief examples are outlined below.

In NOR1 a member spoke of the issue of impact and how the subunit used their professional academic position and experience to respond to any introduction of such demands. There had been an increase of interest the quality of results produced in schools, but such effects were well known to be difficult to measure. The respondent argued that if there was an attempt to introduce demands for effects measurement then the subunit members were well able to respond academically to any such demands due to their academic competence in this area based upon a solid, active research foundation:

*So I think it is important to legitimate that (we) have an academically solid programme, which is based on research on leadership... And this is the point of research based education, that we are actively engaged in the programme but also that we keep ourselves up to date on the research front. [NOR1k].*

These are interesting reflections, showing what the member perceives to be the ultimate base of how a decision will be made. In a similar vein, an example came from NOR2 where a respondent reflected over how the academic anchoring outweighed evaluation and was the basis of decision making for the members of the subunit:

*But, you know, evaluations like these are not the best arena for learning. We primarily learn from ascertaining whether things are going well, ok or poorly. So I don't learn much from them. How I learn to develop the programme comes*

*through my academic reflection; from the academic discussion, from literature, research and being in the classroom. And also I suppose in dialogue with the commissioners. [NOR2g].*

In ENG1 there was also agreement on this topic, as a respondent compared their basis for change with the new formalised evaluation systems recognising that professional identity with colleagues was the most important factor in decision making:

*I think quality essentially comes back to professionalism of people and that people working as professionals have a responsibility to ensure that what they do is as good as it can be and I think all of this wider bureaucratic apparatus just sort of stops you doing that. So that, you know, I'm quite, I'm becoming increasingly cynical about quality processes in organisations, which is shame because I think there's potentially a lot that could be usefully deployed. But I think quality essentially hinges upon the individual and his or her approach to the way that they do things. [ENG1g].*

In ENG2 reflection was made over the survival of programmes, related to an evaluation that was to be done to look at the quality of a programme. Here the respondent reflected over the nature of the evaluation that was performed centrally. When some of the programmes had come under criticism, the team had to argue for why the programme should survive in the way it was. Here the respondent was clear that when pressures came, calling for the development of popular programmes that would become more heavily subscribed, then they responded by appealing to their academic credibility:

*you know, it's a fight to justify, but why I went onto the Masters, you know we are a university, if we don't have that balance with our own research, [then] my credibility as an academic would be brought into question. I think, I wouldn't be happy if we were just out there turning out courses that draw in loads and loads of people. So at one sort of crass level, yes it's the bottom line, and I think that's true, you know talking to colleagues and friends in other places there is more pressure on everyone these days than there used to be, so the idea of producing courses just because they have academic credibility, I am not saying it has completely gone, but it's harder these days. So, in terms of the evaluation of the Masters, as I say, over the years that I have been here, starting with this [external evaluation], criticism from outside, we've done some things about addressing that, but there's been the criticism from the powers that be within the [NN – org name] about having too many courses... [ENG2p].*

But another respondent also recognised that academic anchoring could also be symptomatic of academic inertia, whereby a certain way of doing the programme had become institutionalised and was defended as the accepted way of doing things:

*there is quite a lot of history of consistency there if you like and there's a culture for the kind of course that we think is appropriate, which may be creating us difficulties, maybe we've got too much stuck in a rut, I don't know. [ENG2m].*

This reliance on professional academic identity was important across the subunits and appeared the strongest underlying defence and reference point in the discussion about evaluation. I will deal with these issues again in the next chapter when analyzing the responses in more depth.

#### **9.4 Summary**

It is interesting to see how the various subunits dealt with the pressures upon them and the processes that took place within. In NOR1 the subunit had developed a collegial based decision making system to try and address issues of which evaluation was a major area of discussion. This was practiced on all areas and the group members considered that they had developed into a tightly coupled unit. However, members were also aware that these processes were time consuming and were beginning to consider ways of redeveloping their method of working without losing aspects of their collegial form. The subunit at ENG1 had experienced that the pressures of time and increasing internal and external demands meant that they could afford less time to such discussion about evaluation. They also experienced that the institutional system had become increasingly bureaucratic and they suffered from a lack of feedback. The subunits at NOR2 and ENG2 both exhibited signs of becoming increasingly more decoupled from the central organisation system especially related to evaluation. Much of this is due to the way it is organised and the lack of possibilities for input. But this meant that they were also less likely to discuss evaluation and make decisions together in how to develop and move forward, relying on their individual professional judgements within their own areas of responsibility. This is not to suggest that their judgements were necessarily flawed; each of the interviewees had experience of evaluation and academic staff had long careers in teaching and research. But they had also experienced a lack of feedback from the levels above them in the organization and this had led them to use little or less time on developing evaluation frameworks beyond their own self-evaluation and formative discussions with programme participants.

The issue of commissioner competence was also an engaging topic for the respondents in the Norwegian institutions. The rapid growth in demand for postgraduate programmes resulted from the increased funding available from central government for local and regional authorities, who as employers were entrusted with commissioning programmes for their own areas. As has been shown this created tensions in terms of negotiations surrounding programme provision, implementation and evaluation, with variation in interest and expectations for outcomes. Additionally as has been seen throughout the data chapters, the subunits needed to balance demands and pressures from their institutions with regard to matriculation and quality assurance with the external

expectations, or lack of them, with regard to intended and expected outcomes. At the same time, association with perceived competent external commissioners led to subunit members being able to adopt more formative, improvement directed evaluation models which appeared to lead, after reflection, to both instrumental and conceptual changes considered to benefit current programme participants, bringing more immediacy to the process. Although there was recognition that this could create challenges resulting from resource dependency, on the whole subunit members considered they maintained a strong influence over the processes. Interestingly, there were not perceived to be any major external demands for evaluation attached to funding process in terms of outcomes. At the same time in balancing the local demands where they existed, the HEIs had to consider the basis they would negotiate with the different stakeholders in the task environment based on their self-assessment of the programmes delivered balanced against perceptions of commissioner and programme participants.

## **10. Discussion of decision processes thought to influence the design of evaluations**

As was outlined in chapter 1, the overall aim of the study was to investigate the decision making processes surrounding the design for evaluations of postgraduate programmes in school leadership development. These processes relate to the perception and interpretation of demands the programme providers considered that they faced. The focus is on understanding how these processes take place within 4 different settings across two countries. It was recognised from the outset that the processes under study are not thought to be exhaustive nor is it maintained that all the variables have been isolated. Rather, focus has been placed on exploring these topics and introducing new areas for investigation that have only to a lesser extent been applied to evaluation research, namely with regard to understanding decision processes about evaluation through the application of decision making theories. It was considered that deeper investigation is required into the organisational decision making processes that lead to a particular choice of model, and how and why this choice is made by the evaluating group. However, as was stated in chapter 5 the organisational decision theories are perceived to function in combination, rather than considering that any one model can explain all behaviour. The context of this study focuses upon decision process within HEI subunits evaluating their school leadership development programmes. By investigating processes taking place within the particular subunits focus is placed on the framing of evaluation demands. In this chapter I consider the themes drawn from the data outlined in the previous 3 chapters related to the theory outlined in earlier chapters.

The framework for the areas of investigation for this study were developed from a reapplication of Stufflebeam et al.'s (1971) categories of problems related to evaluation decision making, as was described in chapter 5 and summarised in table 8. These areas were considered by the authors to be important for understanding decision processes related to evaluation. Five interlinked areas of investigation were outlined. The first category is focused upon definitions of evaluation as understood by the subunit members, as well as their perception of understanding within and across their group. The second category is focused upon perceptions of demands for evaluation upon the subunit, attempting to discover the range of pressures faced and how these were interpreted. The third category is focused upon reflections surrounding the designs available and utilisable within the organisation. This category also focuses upon the purpose of the designs and the perceived degree of agreement about the basis of the models utilised. The fourth and fifth categories are focused upon decision makers and decision making processes related to choice of evaluation models. In these categories it was considered important to identify the different decision making roles related to evaluation within the subunit, as well as investigating the decision processes by analysing the reflections of respondents in relation to

theories of organisational decision making. These decision processes are related to the different theoretical templates of organisational decision making.

Through analysis of the data, three interlinked areas were identified to illuminate the research problem: demands, designs and decisions. As was outlined above, these processes are thought to be interlinked and recursive, but are dealt with separately for ease of presentation. The data were presented in chapters 7, 8 and 9, focusing respectively on the responses given by subunit members with regard to demands placed upon them, potential and actual designs available for implementation and decision making processes underlying their choices. This chapter is also divided in the same way, as I relate the findings to the theory outlined in chapters 2 to 5. I deal firstly with reflections over demands on subunit members.

### **10.1 Demands**

In chapter 7 focus was placed upon the first research sub-question, investigating how the members in four different HEI subunits offering postgraduate programmes in school leadership perceived the pressures and demands upon them with regard to the evaluation of their programmes. The sources of these demands were found within the task environment: that is, from policy makers, agencies, commissioners and programme participants; from the wider institution and also from within the subunit itself. This framework of overlapping tensions and pressures presented a complex set of demands that the subunits needed to negotiate when implementing evaluations. The responses made by members of the different subunits are outlined in table 13 further below.

In chapter 2 I outlined the academic field that subunits' postgraduate programmes were operating within. There was noted to be a significant policy focus on improving competence of school leaders, in these cases within and across the OECD, which in turn had influenced national education policy. Drawing on the work of Bush *et al* (2006) it was noted that the focus of policy has increasingly become concentrated upon understanding how competence development has an impact on improved pupil outcomes. Supporting evidence should inform current and future policy. In addition focus is placed more heavily upon the role of the head teacher as leader of the organisation. It was further noted in chapter 2 that these policy foci have led to a wide discussion within the field of educational leadership and management about the variety of programmes on offer; their purpose and perceived efficacy. This discussion is partly centred upon the question of evaluation. If one should look for evidence of programme impact what kind of evaluative frameworks should be used and in what way should they be implemented. These demands were observed more strongly in England, where it was respondents noted that the National College for School Leadership (NCSL) was setting an agenda for understanding the school leader role, in addition to developing frameworks to investigate programme impact. Both these areas were considered by subunit members to be

reductive in form. In Norway, policy had focused more upon decentralisation, where local and regional authorities were given the responsibility to commission programmes. At the time of data collection this provision was subject to little central control. As a result, programmes were generally found within the higher education sector, where commissioners chose between different portfolios on offer.

In chapter 4 the organisational context was outlined. The programmes under investigation in this thesis are found within the portfolios of higher education institutions. The subunits as parts of higher education institutions face similar demands stemming from the Bologna process, part of which is focused upon the development and regulation of evaluative frameworks. It was also noted how there has been a general policy move within Europe towards greater standardisation (Brunsson & Jacobsson, 2000). Standards are increasingly seen to be downloaded into organisations as a supply side function (Furusten, 2000), changing from interpretable guides to measurable targets Kushner (2001). The “customer perspective” has at the same time widened (Øvretveit, 2005) and with regard to the programmes under study here one might perceive government, school owners, teaching staff, parents and pupils as indirect customers or recipients of their benefits. Subsequently the degree of impact is extended. This has become more evident within the field of higher education, notably in relation to quality assurance and its regulation (Stensaker, Rosa, & Westerheijden, 2007: 253-4). Stensaker et al. note that control of these systems has been developed through the setting up of organisational bodies, as well as increased moves towards greater legal control. Quality assurance developments were seen by Henkel (2002) to be part of the progression of NPM focused on increased efficiency and greater control. In section 4.2 it was noted how the shift to evaluative control has collided with the traditional technology of H.E., where academics set the agenda and quality criteria (Kogan, 2004: 6) based on what Vedung described as “professional, mostly unwritten and tacit, quality norms in self-evaluations and against quality norms of their peers” (2003: 42). Henkel (2002) notes that introduction of QA principles into the academic arena will be tempered by institutional values and traditions. Researchers into higher education have also called for more investigation into how academic staff interprets quality assurance initiatives (Nasser & Fresko, 2002; Newton, 2000; Westerheijden, Hulpiau, et al., 2007). The subunits in this study find themselves within the framework of these two major influences.

Within these two major contexts there are different areas of interest within the data with regard to demands and pressures faced with regard to evaluation. I have chosen to focus on two areas in this section. The first area deals with how the subunits interpret and balance the varying tensions that they face. The second area deals with their perception of the purpose of these demands, the perceived use of the information gathered and their application across the different contexts.



### 10.1.1 Balancing tensions

As was noted above, each of the subunits was operating across complex and sometimes conflicting contexts. Respondents in England discussed the changes in the field of school leadership development both as programme providers “competing” with the NCSL and wider educational policy, and as HEI subunits “coping” with increasing demands for greater formalisation of their assessment activity and additionally many operating as external evaluators for NCSL programmes. Evaluation was a key meeting point in this arena. There was a felt tension between an educational sector and the field it finds itself in. It was important for respondents to outline how HEIs had traditionally different forms of evaluation based on the professional judgement of the academic, and yet now under New Public Management and Modernisation policies, external pressures were developing to form a clearer link between funding, implementation and achievement. In England there was a sense of resignation amongst respondents that these bureaucratic processes were fixed and determined, and there was a notable shift from programme control to fiscal accountability. It was against this backdrop that the call for a demonstration of impact was becoming more integrated across the two contexts. The subunit members found themselves placed between the two demands, as illustrated in figure 11 below. In addition, students were noted to have become institutionalised into a new role, that of customer or consumer<sup>182</sup>. This role was clearly evident in both impact and quality assurance processes. Interestingly the students on the programmes in this study were by and large professionals in full time positions. Respondents across all the subunits recognised that in many respects it was the change of role given to the participants that was important; the issue of student voice had meant that providers needed to become more proactive in engaging a response as well as following up the feedback that they gained.



**Figure 11: Balancing contextual tensions**

The responses from the subunits in England noted that the tensions from what originated as policy demands had become difficult to balance against the evaluative processes taking place within the sub-unit. More emphasis on quality assurance and formal reporting gave the impression that less “evaluation” was being done and that tensions increased exponentially with the more demands

---

<sup>182</sup> as Dahler-Larsen and Krogstrup suggest, programme participants’ perception and rights are moving from users to consumers (2000: 296)

that were placed on them. This was allied to the fact that the members perceived both demands and institutional responses as bureaucratic. The interviewees across the subunits discussed how alongside the introduction of quality assurance there had been increasing pressure for accountability within their evaluative systems. As was noted in section 4.2, Becher and Kogan (1992) have outlined 3 “potentially conflicting” modes of accountability that influence the internal units in HEIs: public contractual/managerial, professional and consumerist. Respondents in both countries, but particularly in England, noted an increasing pressure from both policy and market demands, while they focused more towards the professional accountability associated with their academic field. There were nuances in the responses however, recognising the academic field not to be in full agreement with regard, for example, to the potential to ascertain impact of programme effects. In NOR2 there was also a greater openness to the importance of “customer feedback”, but this was qualified against professional judgement seen to be the most important factor of appraisal. The key point was the continued identification with professional accountability as the arena to ascertain quality of provision.

#### **10.1.2 The extra level of commissioner and the influence of government bodies**

In Norway the role a commissioner appeared to raise further questions in both the ways that evaluations were carried out and how eventual findings might be used. As was exemplified in the response of one interviewee in NOR1 at programme level the academic in charge was responsible for development, but they might need to negotiate with commissioning bodies. This negotiation was a recent aspect in the development of programmes and brought new challenges and tensions to decision making routines. Similar responses were given at NOR2. As will be seen in further sections, the level of tensions was dependent upon the competence of the commissioners. Inevitably, commissioners were keen to see local improvement within their own schools that might be linked in some way to the quality of provision their ‘employees’ received on the development programmes. But with little experience in and knowledge of educational leadership any demands they had related to ascertaining impact could be rebutted by the academic staff. Interestingly here in both Norwegian subunits there was evidence of a renewed vision for programme providers, engaging more closely with programme participants as they gained greater access to the workplace. At the stage of these interviews however the groups recognised the limitations of the feedback they received and the limited quality of the evaluations undertaken by commissioners due to their limited competence with issues surrounding school leadership. However, in both NOR1 and NOR2 subunit members had experience of programme commissioners with clear goals and related competence. The results of evaluations drawn from the quality assurance system might provide an overview of attitudes to the programme and what was interesting was the way in which this limited information might be interpreted by the commissioner in their review of the programme. This appeared to reinforce particular emphases of programme evaluation, for example

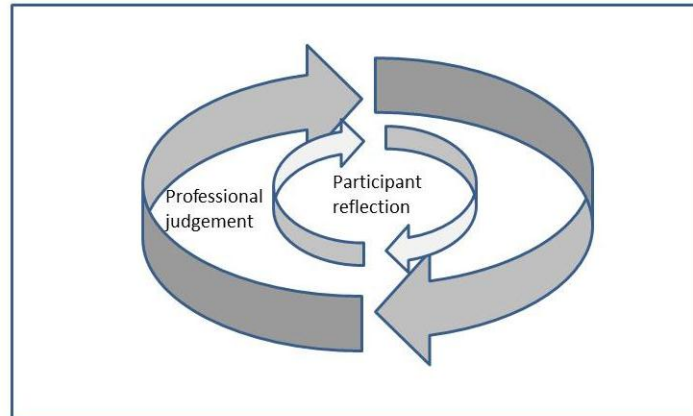
focus on participant satisfaction. Sometimes the commissioners implemented their own evaluations to assess satisfaction. This was exemplified in a response from NOR2, where it was recognised that these evaluations were satisfactory in of themselves but they were felt to provide less information about the programmes than the informal formative evaluations that the programme leaders themselves undertook.

In section 5.6.1 reference was made to Thompson's (2003) research into decision making and the impact of demands from the task environment. Thompson recognised that under circumstances where the task environment lacked the expertise to assess or where causes and effects are difficult to ascertain, organisations/units seek extrinsic measures of judgement, and additionally drawing on the work of Simon et al (1954<sup>183</sup>), noting when any outcomes of an input are outside of the control of the provider organisation, then outcome assessment would be resisted (2003: 92). This line of argumentation will be dealt with further in the sections of this chapter and particularly in section 10.4.1 when considering the degree of internal coupling.

An ongoing challenge for the subunits that will be observed throughout the various sections of this chapter involves the positioning of participant reflection and feedback within the evaluation framework, increasingly since the introduction of quality assurance frameworks. Responses from each of the subunits reveal that subunit members value the comments and reflections of programme participants but they reacted to the way they felt this data was to be utilised. The subunit members in all groups declared themselves willing to discuss and follow up this data and positively encouraged participation but at the same time saw the necessity to analyse and interpret it with their students within the framework of their occupational community. This relationship is presented in the figure below.

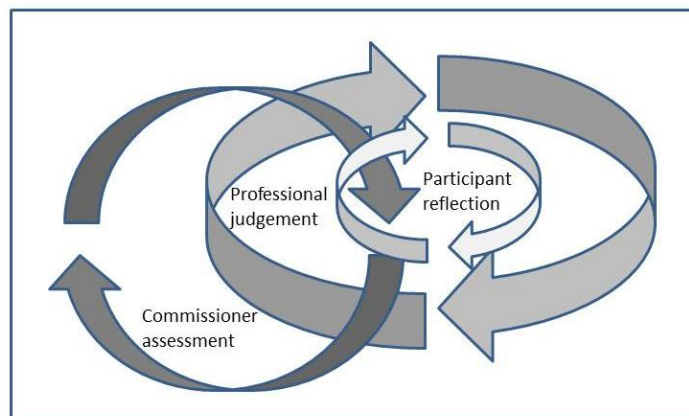
---

<sup>183</sup> Simon, Herbert; Guetzkow, Harald; Kozmetsky, George; and Tyndall, Gordon (1954): *Centralization vs decentralization in organizing the controller's department*. New York: Controllershship Foundation.



**Figure 12: The framework for reflection over participant feedback**

Interestingly it was noted how the effect of commissioner involvement in Norway had amplified the evaluation framework of the programmes, even though at the time of data collection these processes were still considered to be at a rudimentary phase. Commissioners' expectations varied in expected effects for the programmes, as did their competence to assess programme impact. However, the subunits members reflected upon how their own professional judgements of the programmes had begun to be impacted by the exposure to workplace application and commissioner reflection even if they still considered this at an elementary stage. At the demand level the subunits investigated in Norway had experienced how more concrete demands had required them to perform evaluations more frequently, with the data needing to be presented in a more concrete way and how there was often a broader stakeholder group that they needed to relate to, examples being given of commissioning bodies meeting with programme providers and representatives from the programme participants. Whilst previously their evaluation had been based upon professional judgement building upon interpretation of participant reflection, they were now exposed to a twofold increase in data, that is, the commissioners assessment of the programme's impact based primarily on interpretation of participant satisfaction and in some cases also on an attempt to assess impact on the workplace, and additionally the collective reflection of programme participants from a specific group, and in some cases a number of senior members of leadership teams from a particular school or educational workplace.



**Figure 13: The framework for reflection over participant feedback when involving commissioners**

Although, as I have stated, these varied in the level of reflection, the point is reiterated the subunit members discussed the influence of such “prompt” processing over their evaluation process. This was not reported to have been an uncomfortable process for both subunits, and the members related to the fact that those commissioning the programmes continued to have respect for the providers’ professional underpinning. The challenge as was exemplified in subunit NOR2, was to move commissioners on from assessment of degree of participant satisfaction to assessment of degree of application in the workplace. What is interesting here is that processes were affected by an assessment.

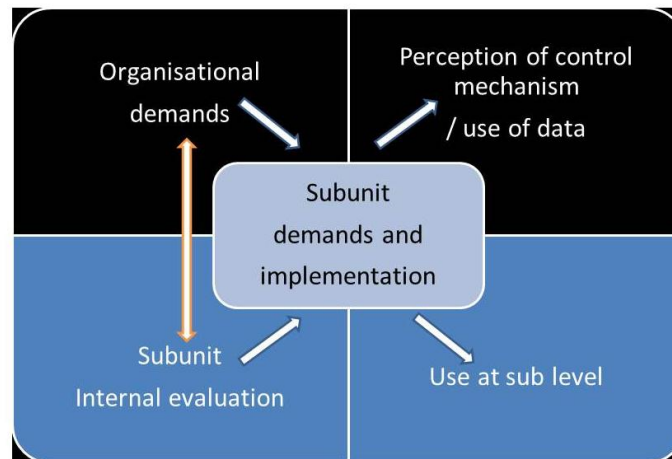
As was referred to above in more general terms, in England the forming and rapid development of the National College for School Leadership (NCSL) had had an interesting influence on the HEI based programmes. While the programmes in England were not subject to the NCSL, there was recognition that it was driving the wider field of understanding about school leadership and its role, and ultimately evaluation techniques. With national research funding being filtered through the NCSL and its programmes operating in competition the members of the subunits felt under great pressure to conform to the national standards and methods of working. The majority of respondents from England had been engaged at some point as external evaluators for the NCSL, and continued to deliver applications to tender. Respondents spoke of the strong influence of the NCSL but considered that the evaluative frameworks were simplistic and reductive. This matched national demands, that programmes could demonstrate their effectiveness. While the units reacted against this, they also recognised that they were being driven in a similar direction if they were to survive.

### **10.1.3 Internal pressures and influences**

Interestingly it was within NOR1 that respondents discussed the pre-eminent influence of internal demands and pressures as framing the evaluation process in relation to the wider organisation. More than merely meeting the requirements of different mandators, the subunit members spoke of contributing to development of their own programme and of the wider institutional frameworks for evaluation. It was the latter point that distinguished the subunit at NOR1 from the other units in the responses. While the members of the other subunits highlighted the importance of their self-evaluation processes, these were seen to be decoupled from the wider evaluative events in their respective organisations. There was a strong utilisation focus within all of the subunits, but the subunit at NOR1 had the intention of ensuring utilisation within the wider organisation.

Amongst the other subunits there was moderate to weak internal pressure to develop their evaluation processes. Whilst subunit members from NOR2 and ENG1 and ENG2 spoke about the importance of academic discussion related to programme improvement and their frustrations over the quality assurance systems and other external evaluation frameworks, there was little engagement with and little discussion about the improvement of the institutional frameworks and no reported response that matched that in NOR1. The variations are presented in the figure below. The lower part accounts for within subunit demands for evaluation, while the upper part relates to the demands from the wider organisation. Factors affecting the subunit from the wider organisation include perception of evaluation as a mechanism for control and perception of use of data. Perception of data utilisation is also important within the unit.

As can be seen in the figure there might not only be a notable variation in the way that organisational demands are implemented, but also in the way that the subunits engage with the organisation over these demands. I will return to this idea further in section 10.3 when dealing with decision patterns and responses, but in NOR1 in particular there was a sense in which the subunit engaged in a different way with the demands placed upon them. This process of moderate engagement with the institutional frameworks appeared to narrow the gap to the internal processes of the subunit, bringing greater proximity between the evaluation processes at the different levels. As has already been outlined, this did not appear as evident in the other subunits apart from NOR1.



**Figure 14: Demands at subunit and wider organisational level**

#### **10.1.4 Evaluative information as signal and symbol**

As was outlined above, the second main area of interest arising from the findings is related to the perception of the purpose of demands to evaluate, focusing on perceived and applied use of information gathered. These arguments stem from a reaction to rational theories of choice and decision. In section 5.5.1 I outlined critical responses to the rational theory of decisions and these were subsequently related to evaluative processes. Feldman and March (1981) discussed how information gathered can function as signal and symbol. They observed 4 major problems relevant to this study with regard to information gathering and rational decision making in organisations. They noted that available information in an organisation is systematically incorrect, that it is often gathered in an instrumentalist perspective for legitimising purposes, that it is often presented as quantitative in form to emphasise objectivity, and that it is often gathered in a “surveillance” rather than decision mode”, having no “apparent immediate decision consequences”, merely monitoring the environment. They noted that organisations collect far more information than they need or can use, and additionally offer incentives for doing so (1981: 174). Interestingly, Feldman and March concluded that the idea of “information utilisation” is considered to symbolise that a group is committed to rational choice (ibid.: 182). Munro recognised similar findings in research into the impact of auditing in the field of social work, noting how reductive measures of impact take over as performance indicators, seemingly as a way of rationalising current and future decision making about programme quality and efficacy. In doing so, work across the public sector is increasingly focused upon surveillance of quality control systems rather than the practice of work undertaken (Munro, 2004: 1093). Dahler-Larsen (1998; 2006a) has also drawn on these ideas, relating them to evaluation theory and practice.

### *Signal*

Reflections from the subunits concerning the implementation of quality assurance systems and evaluation frameworks were in line with the ideas of the signalling effect of information gathering (Feldman & March, 1981). As information gathered is little utilised, organisations might be perceived to collect it with the purpose of matching environmental demands and at the same time signalling their seriousness. In addition respondents across the units revealed a perception that the processes were based on a surveillance idea rather than a decision based one. As was outlined briefly in section 5.5.1, Feldman and March referred initially to the idea of organisations monitoring their environment, a kind of “thermostatic linkage between observations and actions” (1981: 176) whereby they seek to understand what expectations will be placed upon them, although this is more akin to collecting “gossip” than an “explicit calculation”. The subunits reported little feedback in any of the organisations as to the results gained from evaluations, and the groups could not say how the information gathered was used. This was similar to Munro’s (2004) findings, as one respondent in ENG2 pointed out, there was only interest in the wider society in what works, and not in what doesn’t and why it doesn’t. These observations are also similar to findings in Dahler-Larsen’s research (2006a: 134).

Additionally respondents described a situation that had developed where what is important is defined and stated by the customer: which might be the participant themselves or in the case of the commissioned programmes in Norway, mediated and / or moderated by the commissioner or other external body. In England the signal appeared to the respondents to be to all intents and purposes that ascertaining an understanding of programme impact was the most important evaluative process. This was also evident in some cases in Norway, but as more of a direct demand from specific commissioners; in NOR1 this was exemplified by a local authority looking for evidence of school improvement in terms of team development while in NOR 2 another was looking for improved pupil outcomes, which they maintained would come from a certain type of leadership.

### *Symbol*

As well as a signalling function, information can also be gathered to symbolise that an organisation is competent but without the process of information gathering providing any greater knowledge to the organisation itself (Feldman & March, 1981: 177). This again suggests a loose-coupling between information and its utilisation, but now with the purpose of presenting an image to the wider environment, similar to Simon et al.’s idea of extrinsic judgement. Dahler-Larsen also recognised that such approaches could spill over into an organisation’s evaluative framework (1998: 134). None of the subunits in question reported receiving consistent feedback in relation to the evaluations they were asked to perform neither did they expect any unless something should go drastically wrong. They were unsure therefore, especially internally, as to the



purpose of the information gathering required as part of the evaluation process; beyond that it was a formal demand. As Reichert (2007) noted, for development to take place there is a precondition is that individuals trust that QA will offer benefit and be followed up. The symbolic effect of performing the action was considered to be more important. However, they generally felt supported in the sense that their programmes were considered successful. Any focus on longer term development and improvement was felt to be missing from the quality assurance frameworks. The systems were therefore felt to be outward focused to meet external demands rather than inward focused on development and improvement, or as one respondent from ENG1 described it, “*it’s just a kind of accountability process*”.

These findings are not unproblematic. While this type of surveillance takes into consideration the organisation’s relationship to its environment, there are question marks raised over the way that an organisation treats evaluation internally. Evaluation researchers have also noted the increasing acceptance of surveillance and control mechanisms drawn from wider society (Norris & Kushner, 2007). It could be suggested that this fits with the idea mentioned above that responses were only expected when courses failed to live up to the demands placed in terms of quality, considered to be an external demand<sup>184</sup>. This might relate to poor feedback which might affect future applications to the programme or perhaps worse still dropout; which in the field of practitioner focused studies is not unlikely. But it was unclear how this information would be used and what influence it would have over future decisions about the programmes. In each of the subunits members reflected over the perceived importance of these data for the educational authorities, but also questioned whether the data collection processes could explain the complexity of why students dropped out. There were further question marks placed over the reductive nature of data collection and lack of opportunity to offer an explanation to those at higher levels in the organisation. In NOR 1, for example, it was recognised that the subunit as a group desired to account for anomalies with student throughput. These anomalies needed to be explained at Programme board level and Institute level but the subunit members needed to push for opportunities to do so to ensure that their reasoning was heard. The responses from the other subunits, furthermore expressed disillusionment with the processes. As members felt that they were not listened to and they merely gave up trying. This led to the micro frameworks of evaluation becoming increasingly decoupled from the central system. As will be seen in section 10.3 this led to a further form of “dismissive submission”.

A summary of the recognised demands for evaluation is presented in the table below. As is noted in the section for policy makers, in England there was considered to be a strong set of demands, more widely focused upon evidence

---

<sup>184</sup> An area exemplified in the responses was the impact upon ‘throughput’, or overcoming the danger of dropout.

and impact. This was felt in terms of the subject area and the more generally in the broader field of higher education. In Norway these demands were felt to be much weaker, and described more as pressures.

**Table 13: Summary of demands upon subunits**

Perceived demands upon the subunit with regard to evaluation						
Subunit	Task environment				Institutional	Within-unit
	Policy makers	Commissioners	Students	NCSL		
NOR1	Weak pressure for evaluation, but developing moves towards evidence-based policy. Bologna related processes introduced moderately. Focus upon formal activity.	Adaptation to commissioner needs. Problem with 'competence'. Some requirement of visible change.	Context-bound and group dependent. Dependent on level of study. Students that demand evaluations, 'very clear voice'.	n/a	Implementation of Bologna related processes. Programme validation, systematised framework. Throughput focus. Variable pressure to implement quality assurance framework.	Strong within-unit demands to improve evaluation competence. Collegial discussion and feedback. 'Culture for quality and learning'.
NOR2		Adaptation to commissioner needs. Expectations of improvement. Year to year assessment.	Evaluate based on personal preference. Demand that responses be acted upon.	n/a		Limited collegial discussion and feedback. Little discussion about evaluation. Individual self-evaluation.
ENG1	Strong policy pressure for evaluation and assessment. Evidence based policy and linkage between funding and outcome (impact). Bologna related processes introduced strongly. Focus upon formal activity.	n/a	Varied perceptions of quality amid different preferences.	Central role in 'training' of school leaders. Influence over evaluation models. Considered reductive. Focus on 'what works'.	Implementation of Bologna related processes. Programme validation, systematised framework. Throughput focus. Strong top down pressure to implement quality assurance framework. Bureaucratic demands.	Moderate within-unit pressure about evaluation to improve programme. Some collegial discussion and feedback.
ENG2		n/a	Evaluation based on past experiences with institution.			Moderate/weak within-unit discussion about evaluation to improve programme. Limited collegial discussion and feedback.

The reflections of the subunit members concerning the complex web of demands surrounding evaluation appear to give support to similar reflections made by Dahler-Larsen, that evaluation stemmed from a “normative demand for the development of evaluation procedures rather than a rational demands to improve programme function and goal realisation”<sup>185</sup> (1998: 92-93). Dahler-Larsen interpreted this to mean that institutionalisation of evaluation procedures creates a problem for the “explication of programme goals and evaluation criteria” (ibid.). Attention is furthermore drawn away from other organisational values (1998: 95) that might be of fundamental interest to an academic group. Such an

<sup>185</sup> My translation from Danish.

approach is thought to cause problems for the design and development of evaluation models. It is to this area that I next turn.

## **10.2 Definitions and designs**

In chapter 8 focus was placed upon the second research sub-question, regarding the evaluation frameworks available to subunit members, relating the responses of the different subunit members to evaluation definitions and designs. These findings are linked to the discussion of evaluation in chapter 3 and related material in chapter 2 considering the debate surrounding evaluation of school leadership programmes. Additionally it is related to discussions concerning quality assurance within higher education institutions as outlined in chapter 4. Chapters 2 and 4 recounted much of the backdrop for these programmes, and it was noted that demands for evaluation were much stronger, centralised and more tightly controlled in England compared with the looser, more decentralised and loosely controlled frameworks in Norway. This can be related to Karlsson's (2003c) findings, as outlined in chapter 3, where England was considered to have a more positivistic tradition for evaluation than Norway, seen to be based more upon developing the democratic ideal. However, as was noted in the previous section, the implementation of Bologna related policies had seen the beginnings of tighter control over evaluation processes in Norwegian HEIs than had been previously experienced. Such developments have taken place in what has been seen as an age of accountability, with the advent of New Public Management (NPM) and Modernisation policy processes. Evaluation has been a central part of these policy processes, with greater focus on performance management (Dahler-Larsen, 2007) and greater shifts towards control, measurement and modelling (Hood, 2005; Dahler-Larsen, 2005b).

Against the new focus upon and impetus related to evaluation questions have also been raised with regard to whether this has resulted in increased utilisation of data. As was seen in chapter 3, utilisation of evaluation data is seen as an important issue. As a primary purpose of an evaluative exercise, many authors have attempted to understand why there is so little utilisation of findings (Weiss, 1972; Alkin & Taut, 2003; Kirkhart, 2000), particularly in an era of greater accountability (House, 1993). This is part of the evaluation paradox, where evaluation is considered to become a mantra rather than a rational response to a request for information and evidence (Læg Reid *et al.*, 2004). Evaluation can therefore become institutionalised in organisations as little more than a ritual reflection (Dahler-Larsen, 1998). Greater understanding of these processes and what moderates them is thought to move evaluation research beyond focus on method proclivity and involvement of stakeholders (Christie, 2003), towards understanding decisions made. Interest in understanding models of evaluation stems from Stufflebeam's (1983) research, where investigation of the choice of model was related to ascertaining underlying practices, traditions, expectations and experiences.

The data presented in chapter 8 was related to two major topics: the underlying frameworks of evaluation and respondents' attitudes to them, and a discussion over current models, their purpose, focus and limitations. The latter area also includes further reflections over the introduction of impact focused models.

### **10.2.1 Attitudes to evaluation**

As was outlined in chapter 8, in attempting to understand the underlying frameworks of the evaluation models used within the subunits, questions were asked with regard to members attitudes to evaluation, reflection over influence and impression of current models employed. The purpose of this was also to gain some understanding of the impact of evaluation theory on enacted models as well as to see whether the fact that as the programmes themselves focused on evaluation techniques this would have a greater impact on the evaluation models within the subunit. The intention was partly to follow up Christie's (2003) research, as outlined in section 3.2, concerning the influence of evaluation theory on practice. There was noted to be generally little reference by respondents to specific evaluation theorists, and this was common across the subunits. However, some members of NOR1 reflected upon the influence of theorists they had experienced through the programme. Christie's findings had distinguished between the more experienced external evaluators and internal evaluators. Christie found that those with less evaluation experience, usually internal evaluators, tended to follow the frameworks set within the organisation, as opposed to the more experienced evaluators who felt able to map out their own pathways. An additional finding arising in the higher education subunits in this study was a more notable difference based on experience or longevity within the wider organisation more than on evaluative experience per se. Amongst the respondents even those who were newer in their organisations had had wide experience of evaluation from previous roles and positions. However, those who had had less experience in the wider organisation appeared to follow more closely the framework set by their organisation, although this appeared to be moderated, in line with Dahler-Larsen's (1998) findings, by the extent to which evaluation was discussed and agreed upon within and across the subunit. Experience was described more as a development of *evaluation vocabulary*. There was an expectation that this was part of the responsibility of a higher education academic, to develop one's own repertoire.

Interestingly there did not appear to be a propensity for quantitative methods in the work of the academics interviewed in the subunits under investigation, particularly with regard to evaluation methodology. The role of professional identity and relationship to the field of study related to these occupational communities also seemed important in this organisational setting. This created challenges for the subunits when it is considered that the quality assurance frameworks in HEIs have been more heavily centred on the production of quantifiable data focused upon output statistics. The relative strength of the subunits as occupational communities appeared to be challenged by the effects of institutionalisation in the wider organisation combined with policies

introducing greater standardization across the system. I will return to this latter point in the section dealing with decision processes below.

### **10.2.2 The QA system: formality and informality; accountability versus improvement**

Responses across the subunits were similar with regard to a shift in perspective of evaluation frameworks and convergence towards accountability models. As was seen in section 8.4, there was also a similarity in the configuration of frameworks including module evaluations, both formal and informal; student feedback frameworks, use of participant representatives and reference groups; monitoring of programme delivery and progression; regular course team meetings to assimilate feedback and assess implementation progress; involvement of external examiners; annual course reviews; summative programme evaluation; and periodical cross programme evaluations. These processes were recognised to reflect the ENQA standards and guidelines referred to in Chapter 4.

In addition, the demands for and implementation of quality assurance systems within higher education institutions had led to a clearer division between formal and informal evaluative frameworks. What was different, however, was the perceived interplay between formal and informal evaluation processes within the different organisational structures. This finding is considered to be of importance to the overall research question, reflecting that despite the similarities of demands on each of the organisations there are different decision responses to the demand to evaluate.

A further point of importance regards the “tension” between accountability and improvement focus. The theoretical findings in this area were outlined in section 4.3. Harvey’s research has suggested that accountability demands will likely merely only be complied with by academics, rather than replacing or transforming current behaviour. It has also been claimed that accountability and improvement are perceived incorrectly as two different ends of a continuum, rather than interrelated areas (Harvey & Newton, 2007: 232). Any real “tension” is thought between perceptions of quality at management and operational levels (Newton, 2000: 155). The findings in this study are considered to comply with Harvey and Newton’s research. Members across the subunits were more dismissive of the way evaluation frameworks and data were interpreted by the leadership than they were of the data that was to be collected. They rejected the reductive way the material was construed and the lack of feedback that was forthcoming. In NOR1, in particular, some members had reflected that the QA system had given them more evidence about the programme, while in ENG1 there was perceived to be a problem with the decoupling of formal and informal evaluation, seen as a challenge to the professional integrity of the academic staff. In ENG2 one academic member of staff recognised and affirmed the different purposes of accountability and improvement, but was more concerned that commitment to improvement meant findings would be used “*creatively and*

*constructively and not just used as an accountability tool to say its ok, or it's not ok*". I will return to these responses in the next section. The next subsection, however, deals with the way subunit members observed the applied models in their organisations.

### **10.2.3 Focus and limitations**

The findings outlined above will therefore be important in understanding decision processes. Another area to be dealt concerns that of perception of model focus as well as discerned limitations with it. When discussing design focus, there were three key areas of importance identified by the respondents. The first is that of *participant perception* and the second of *deliverer reflection*. These two areas are considered to be heavily interlinked. In line with the research into quality assurance frameworks outlined in chapters 3 and 4, subunit members described the evaluation systems to be focused upon student perceptions of satisfaction with the programmes; one member of NOR1 referred to the system as, "*a measurement of students' perception of what they have learned*". In NOR2 another respondent highlighted the reflection at micro-level over own practice. The data findings from evaluation processes allowed a reflection over own teaching practice, and its impact on student learning. Such an approach was also in development at NOR1. Responses were also highlighted in the English subunits, but here the reflection was that the level of focus and questioning was generally too broad to offer useful data for future use at programme level. There was a strong recognition again that the focus on the student was to give increased voice to programme participants.

A third area relevant to the responses from Norway concerned the *role of commissioners* and their interest in these areas. With an added layer of evaluator interested to some degree on the day to day running of the programme, interest was placed in their underlying intentions for the programmes. As will be seen in the next section, this again relied upon their competency, with many reverting to measures of participant satisfaction. This again appears linked, as Alliger and Janak (1989) had identified, to a false assumption that reactions are *linearly related to learning*. When faced with a task of evaluating a programme and its impact, the perceived impact on the individual taking the programme will appear easiest to identify, even if it is unclear whether the response is a correct interpretation.

Such linkages were considered to be part of the limitations of the evaluations applied to the programmes. Subunit members identified *structural problems* related to lack of detail (NOR2), unsophisticated (ENG1), overloaded and unwieldy (ENG2) and badly timed, either coming too early in the academic or coming too late and not giving opportunity for the members to make any changes (NOR1). Such problems were exacerbated by the fact that participants studied part-time, sometimes as distance learning students and most often in full and demanding posts within a school leadership team. The formal requirements of the systems were often inappropriate for the student body, and contributed to

further evaluation “weariness”. It was also considered problematic to *interpret* the responses coming from students. Subunit members found the reductive nature of the questions on central organisational forms as difficult to construe. In ENG1 one respondent had been attempting to update some sections on an evaluation form but was struggling to understand what the form generally was asking for. This is interesting because the format of evaluation did not suit the intentions of the team at the micro-level. As was also seen in section 8.7, the members at NOR1 had discussed this problem and were as a team left with more questions than answers. One respondent noted that in particular there was a problem with future purpose for feedback – an expression of satisfaction or contribution to programme improvement.

Such discrepancies with regard to the problematic issues of interpreting the focus of evaluation models, working with endemic structural weaknesses in the organisation and accounting for student characteristics created difficulties for the subunit members when implementing quality assurance models. Many of the problems were thought to stem from the perceived shift in focus that led to the control focus of evaluation being highlighted. Those groups offering programmes to commissioners also experienced similar issues. An issue that has already been mentioned related to control in relation to demands, concerned a shift in focus towards designs able to reveal programme effects. This issue is raised in the next sub-section.

#### **10.2.4 Cause and effect models: debates and problems**

In addition to problems and frustrations with the quality assurance systems generally, there were also additional challenges related to expectations within the field of study, school leadership. As was seen in chapter 2 there has in recent times been an increased interest amongst policy makers to ascertain the degree of impact school leaders have on pupil outcomes and ultimately to understand how training and development programmes might be assessed in terms of their effects on this relationship. In chapter 3 these policy processes were seen to be closely linked to the introduction and development of New Public Management (NPM) policies. A key part of NPM based policy was seen to be evaluative reform, focused particularly on performance management (Dahler-Larsen, 2007) leading to the introduction and focus upon results driven monitoring (Dahler-Larsen, 2005b). As was noted in chapters 2, 4 and 8, these processes have influenced both higher education policymaking as well as the field of school leadership development in particular.

Dahler-Larsen (2006b) notes that a consequence of these increased demands for evidence of outcomes from policies and programmes has caused marked debate amongst evaluators over the efficacy of models of cause and effect. The author notes that within the educational field certain pedagogical paradigms reject the concept of “effect” as a central theme of evaluation when related to teaching and learning (Dahler-Larsen, 2006b: 104). It was noted that a government commissioned report by Leithwood and Levin (2004, 2005) explored these

issues of programme impact and subsequent evaluation models thought to uncover and map such processes. The advisory work of Leithwood and Levin to the DfES<sup>186</sup> spawned a research project into developing models to ascertain the effects of school leadership programmes. While this first and foremost might be applied to the various leadership training programmes as overseen by the NCSL, there has also been suggestion that they might apply to the more developmentally based HEI postgraduate programmes. Bush (2008b) noted how the field had received this report, recognising the complexity of ascertaining transference and difficulties of attribution that the models were developed to investigate. In an age of evidence, Bush considered that such models were based more on “belief” (ibid.).

As was seen in chapter 8, respondents were asked to consider how they perceived the concepts of effects and impact related to evaluation generally, as well as to their own designs as well as the frameworks for evaluation that they were required to implement. Two major themes were highlighted, the first related to how the members understood and responded to this debate about impact and effects, and the second how they related to this any problems with such an approach. It was considered that these responses would provide further understanding of the subunit members approach to evaluation and a frame for their decision making.

#### **Debate about the notion of “impact”**

All of the respondents considered that evaluative models related to their programmes explored, to some degree, the issue of impact. In NOR1, however, this was linked more closely to the idea of reflection over practice; *self-perception of learning* and *change of practice*. If these aims were the goals then respondents felt the information could be used for such a purpose of assessing programme impact. If looking for evidence of impact on school outcomes, and pupil achievement in particular, then this was not currently relevant. This was a view shared collectively across the subunit. As was seen in chapter 8, one member at NOR1 stated, “*and what we think is that we can’t really measure [impact], but we can see it through the programme activities*”. A similar response was forthcoming from members in ENG1 and ENG2, who also noted how this notion had been built into assessment activities. This is key to understanding how evaluative activities at programme level have more widely been grounded on reflection based models informed by self-reporting.

The different subunits were, however, open to the idea developing a longitudinal post-programme research model to further explore the impact of the programme. There was a curiosity as to how participants had implemented what they had learned. But there was still a general rejection that models could be implemented to connect a programme with outcomes. This of course was also influenced by the interpretation placed on the concept of outcomes, which as Bush noted could

---

<sup>186</sup> Now the DSCF: Department for schools, children and families



be cognitive or non-cognitive (2008b). However, in Norway there was a contrast in NOR2 on this point, where the members of the subunit wished to develop a research project to investigate the relationship between programme outcomes and impact on pupils' results, considering it feasible.

#### **Further problems with models**

In section 8.8.2 five specific problems were exemplified by the respondents to designs for studying effects and impact: difficulties ascertaining cause and effect variables; level of observation possible, time frame available for investigation; model complexity and endemic organisational constraints. While I have in the paragraphs above briefly referred to the first four of these issues, the final one regarding the nature of HEIs can illuminate the connecting point between academic field of study and organisational environment.

In ENG1 discussion with one respondent went into more detail concerning reactions to developing impact models related to a national research project commissioned to discover such linkages. The respondent suggested that the results of this exercise would really only be the confirmation of something already known, that there was a linkage between a programme and school development, but that it was incredibly complex to map. But in discussing this point further an interesting reflection was made. The respondent was asked how the findings might impact the evaluation of master programmes. The respondent chose to rephrase the question, at first dropping the notion of evaluation, remarking that despite being considered to be built on knowledge already held, i.e. that there is a link between quality of school leadership and improved pupil outcomes, the findings of the research would produce some useful new data that would improve the knowledge base of HEIs running Master programmes. However, in response to the original phrasing of the question, that is, whether the findings would have any impact on the evaluation of HEI programmes, the respondent outlined the challenges related to developing evaluation activities within the bureaucratic system which such programmes were situated. This response would appear to signal the importance of organisational structures and ultimately decision making processes surrounding the evaluation frameworks. The findings from the research project were considered to be related to evaluation of impact; it was perceived that this information would affect the study programme but not the practical context surrounding it. The respondent declared, “[i]t is a university wide issue rather than one related to educational leadership programmes”. While these reflections are based on the comments of one respondent, they are also considered to succinctly relate the range of reactions across the subunits related to the organisational context, even respondents open to the idea of evaluating for impact saw endemic problems with evaluation models.

### **Framing the context**

A final point of interest in this section is related to understanding the “situational context” of an evaluation (Stake, 1990). This is considered helpful in understanding subunit responses to different demands placed upon them, but also helps to frame evaluation designs that they have developed. As was outlined in section 5.2, Stake noted that attitudes to measurement are also reliant upon the perception of context. It was seen how Stake’s model framed these attitudes as a continuum between activists who assumed the potential for programmes to impact strong change on their environments and determinists who reject this possibility and see the context as dominating. An analysis of the responses from the interviewees of their perceptions of the different groupings attitudes have been plotted into the table below. There are some interesting reflections that can be drawn from this table. On the continuum the subunits at NOR1, ENG1 and ENG2 exhibited a more deterministic perspective towards their programme evaluation. This does not imply that they saw no opportunity for their programme to make an impact on participants’ context, but rather that they highlighted the dominating nature of the context and other variables and thus the difficulties in isolating the impact of programme influence. The members of the subunit at NOR2 exhibited a more activist attitude, particularly with regard to impact models. It is from these perspectives that a number of tensions for evaluation decision making can be observed. The context for NOR1 was perceived to be more deterministic, the group notably attracting commissioners more sympathetic to their approach. As was seen earlier, the policy frameworks surrounding the field of study as well as approaches to evaluation were additionally perceived to be generally more deterministic. The subunit at NOR2 with its more activist persuasion reflected the approach of the wider institution and as a result attracted commissioning bodies of similar attitude despite the wider policy approach being more deterministic. In the English subunits the more deterministic attitude to evaluation was set against the perception of more activist attitudes both in the wider organisation and significantly within the wider task environment. It is further noted that the NCSL is included in the national policy quadrant applied to England, whilst the placement of NOR2 under “subunit” is based upon responses from members who were developing a proposal for an impact research study based on investigation of programme effects on outcome.

**Table 14: Summary of responses to understanding the situational context**

	Group: as perceived from responses				
Relationship: Programme to context	National Policy	Commissioner / Funding body	Institution	Subunit	Other external demands
Activist	ENG#	NOR2, ENG	NOR2	NOR2*	ENG
			ENG1, ENG2		
Determinist	NOR	NOR1	NOR1	NOR1, ENG1, ENG2	NOR

### 10.2.5 Summary: Defining and designing evaluations

In this section I have outlined the reflections concerning evaluation design and the attitudes of the programme providers. The purpose of this short section of the data collection has been to understand the perception of evaluation amongst the subunit members within the context of HEIs against the broad backdrop of demands as were outlined in section 10.1. Fitting with Dahler-Larsen’s reflections, also noted here are the wider societal demands related to the explicit and implicit expectation of effects developed from public sector provision (2006b: 104). This demand may be thought to supersede, or at least compete with, the value system of the programme provider. Dahler-Larsen describes an interesting conundrum, whereby many different value systems may come into play at the same time, and those objecting in any way may find their arguments considered to be “irrelevant”. While the author’s reflections were originally applied at the compulsory school level, there are impressions from the interview findings that similar findings were evident at HEI level. Such developments involve a greater challenge to the profession.

However, it was interesting to note that there was not a wholesale rejection of impact models. But, in this study the informants generally rejected the notion of discovering effects of programmes on pupil outcomes on the grounds of methodological complexity and over simplified path models rather than rejecting the concept of an outcome or an effect per se. Even amongst those open to implementing effects models, there was recognition that models currently available to them were inappropriate for such a study. Impact was, as such, related to the reflections of the participants. These responses further help explain the frame of reference of decision making with regard to demands for impact evaluation.

### **10.3 Decisions and decision makers**

This section deals with the final research sub-question, which was focused upon what decision processes take place within subunits about the choice of evaluation model. These decision processes were divided into sub points based on a framework of problem areas that might be thought to influence the evaluation process, developed from the work of Stufflebeam *et al.* (1971), as outlined in table 8. The data from the interviews was presented in chapter 9 and divided into three main areas: roles in decision making, the process of decision making and finally, responses to demands. I will use these demarcations in this section, at the same time as reiterating that they are considered to be interlinked and overlapping.

#### **10.3.1 Decision roles and structures**

In this sub-section I deal with the various roles and how they are perceived to be involved in the decision processes related to evaluation within the subunits. The data from the respondents highlighted the roles of five important actors: the academic staff, administrators linked to the programmes, the participants on the programmes, those holding roles in the wider organisation and, as in the case of Norway, commissioning bodies. The reflections from the data are observed to fit well with Hardy *et al.*'s model of three interlinked levels of decision-making in HEIs (1983: 414) presented in section 5.4. In Hardy *et al.*'s model there is recognition that many decisions are taken at the level of the individual academic based on their professional judgement. These findings can be related to the research of Hardy *et al.* (1983) and Stufflebeam *et al.* (1971), who considered that decision making was *professionally based*, building on *accumulated wisdom*. Hardy *et al.*'s model was seen to be developed from Mintzberg's view that HEIs correspond to a "professional" rather than "machine" bureaucracy, characterised by complexity of mission and activity, and relatively loosely coupled, with decentralised decision processes and specialisation at the base level.

Respondents considered their own roles in relation to evaluation decision making as well as that of their colleagues within the academic group in their subunit. Generally, across the subunits, responsibility for evaluation implementation lay with the academic and administrative staff in charge of the module or programme at hand. As was seen in chapter 9, this sometimes led to a tension between the individual's values and opinions about evaluation and the frameworks put in place by the wider organisation. There were three key issues that developed from the data. The first issue relates to how competent the various respondents felt with regard to implementing the task at hand. The second area relates to the connection to the wider organisation and perception of feedback regarding evaluation findings. The third area relates to the structure within the academic team. I will return to the second and third areas in the next section when dealing with the decision processes.

As was noted in section 9.1.1, the role of evaluator was not something that academic and administrative staff members of the subunits felt that they were prepared for by the wider organisation. Competence in the role was considered to be something learned by doing or by input from colleagues within the team. As one respondent described it, such information was picked up by “osmosis”. Each of the respondents had additional experience as working as an external evaluator and this was seen to be a source of opportunity for preparation for the role. The task of evaluator was seen to be a role expected of academics and this was also seen to be an acceptable requirement. Respondents from both England and Norway commented on the academic freedom they had as professionals, and that they were willing to take decisions based on their professional judgment. This is in line with the findings of Hardy et al. (1983) and Stufflebeam et al. (1971).

The involvement of the administrative staff provided a role of buffer (Thompson, 2003) between the organisational systems and the academic staff, controlling that evaluation processes were in accordance with organisational demands and that deadlines were met and findings reported correctly. At the same time, where opportunities arose they fed back information to the higher levels concerning the responses and criticisms of the academic members of staff. In the subunits at NOR2 and ENG1 and ENG2 this interaction took place mainly at programme unit level. But interestingly, as will be seen in the following subsections, the subunit at NOR1 had chosen to involve the administrative staff in the decision processes at team level, interacting more closely with the academic staff as equals. The strength of coupling to the wider administration and at higher levels in the organisation appeared often to be moderated by the role of the administrative staff at subunit level. This idea would need however to be explored further in another study. However, the role of administrator will be seen as important to the concept and development of *collegial construction*, outlined in the next subsections.

As was noted in chapter 5, the variance of roles related to decision making makes the allied processes more complex, challenging in particular the extent to which they can be seen as rational (Dahler-Larsen, 1998). As was seen in chapter 9, amongst the other roles given attention by respondents, it was consistently noted how programme participants had gained a more important status as customer under the introduction of quality assurance systems. This created a dichotomy for the subunit members, as feedback coming from the students was considered to be more important and yet it was becoming harder to collect it. At the same time the institutional constraints and bureaucratic frameworks surrounding these processes meant that subunit members were spending more time collecting the data. This offers an interesting challenge to the frameworks of Stufflebeam et al. (1971) and Hardy et al (1983) with regard to the further erosion of professional judgement. Even though, as was seen in section 9.1.4 an example came from NOR1 where the group had freedom to shape the content of evaluations, this was tempered by the fact that the group

had already produced a model that provided the data required for official reporting. This did not mean that the subunit members were in complete agreement with the frameworks. So, as Dahler-Larsen (1998) suggests, part of the policy change towards greater control and demands for evidence has been the erosion of professional values and judgements. The resulting question is, how have the subunit members responded?

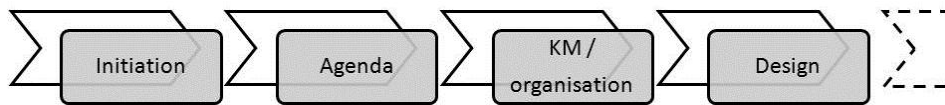
The other area shown to be of importance in Norway was related to the role of commissioning bodies. In particular focus was placed upon the perceived competence of these groups. Many had received the task of providing training and development for their school leaders but at the same time lacked expertise in this role as employer. One respondent noted that this was particularly problematic with regard to the smaller local authorities. Such findings have also been reflected in a study of how Norwegian local authorities use of funding for capacity building across the compulsory schooling system, where the larger authorities were noted to have been better prepared and more competent for the process and perceived themselves to have got more out of it than the smaller ones (Dahl et al., 2004a). But the variation in competence also strongly influenced the decision processes, and required the subunit members to adopt different roles accordingly. The competency of commissioner and the role they adopted could therefore have a strong moderating effect on evaluation decision making. As was exemplified by subunit members in section 9.1.5 this role could either be experienced as an extra tension if the commissioner was focused on implementing an outcomes based evaluation or one of release if their ambitions were closer to the subunit members' values and aims for the programme. But as the commissioners normally picked a programme that was close to their own aims, the subunit members experienced the discussions surrounding evaluation to be processes of negotiation rather than the bureaucratic demands often felt to be practised within their own organisations.

Summarising the perception of different roles and organisational structures provides a backdrop for exploring the processes of decision making and the responses to demands for evaluation. But as was seen in the section, there are an increasing number of actors involved as the processes of evaluation become more formalised. Responses in this study confirm that there is added pressure on academic staff, particularly where their role in making professional judgements is narrowed.

### **10.3.2 Elements of the evaluation decision making process**

Within the overall framework of the problematic areas to be studied it was decided to delimit the focus to specific parts of the evaluative activity. The purpose of the study has been to understand how the subunits interpret demands for evaluation and construct a design to be implemented within their own subunit. In section 5.2 I drew upon the 8 elements of the evaluation process as identified by Dahler-Larsen (2004a). Again with regard to the topic under study focus was delimited to the investigating the first four areas, outlined in the figure

below, which I consider to be the central focus of decision making about evaluation: initiation, agenda, knowledge management (KM) and organisation, and design<sup>187</sup>.



**Figure 15: The elements of evaluation decision making under investigation (after Dahler-Larsen, 2004)**

One of the important topics from across the subunits in this study was the recognition that as a result of external demands, the initiation and agenda phases of the formal evaluations were increasingly taking place centrally in the organisation, as described by Hardy et al (1983) as a result of administrative fiat. Research has shown that decentralisation of power to institutions has increased centralisation of decision-making at the institutional level in order to increase cohesion across the institution by reducing collegial governance at lower levels (Amaral et al., 2002: 289). There were, therefore, two major issues thought to be challenging the practice of professional judgement. Respondents saw that the new demands for quality assurance were challenging and limiting their role as academics and evaluators, by framing the models that they should use as well as the way data should be collected. Additionally, the time taken on implementing for these processes was limiting them from evaluating activity through models based on their own professional judgements.

The subunit members at ENG2 and NOR2, and to some extent ENG1, recognised that they spent limited time on discussing evaluation and making collective decisions about the processes. In each of these groups, respondents spoke of the overly bureaucratic way that evaluation frameworks had developed in their wider organisations. These responses are similar to the reflections of Olsen, who noted that “individuals vary in their participation in organizational decision making”, related to individual differences, organisational differences, decision differences and during what point these processes occur (Olsen, 1979: 277). But rather than merely being a question of power relationships and the right to participate, that is, issues of exclusion or invitation, Olsen suggests that participation in the processes may also be a question of pursuance or avoidance by non-leaders. The author furthers the point that in the context of the university, administrative tasks and those of reporting have often been seen as negatively competing with time spent on issues related to teaching and research (Olsen, 1979: 278). This might be further exacerbated with the increasing demand for more publishing, supervision time, and larger classes. But instead of

<sup>187</sup> The remaining elements being data collection, analysis and summary, and validation.

mere avoidance or suggestion that evaluation was a purely administrative task, the respondents focused more on the fact that evaluations not only occupy their time and resources but additionally provided little feedback that would justify their implementation to the scale demanded. It was a common complaint across the four sites that there was little feedback in the system with regard to information and data that was reported to superiors. While this was qualified by a number of comments in one institution of feeling support from the leadership at Faculty level (NOR1), there were still questions raised as to how the data was used beyond programme level. This was exacerbated by the amount of time spent on the activity, as exemplified by a member of the ENG1 subunit who commented,

*“I think there is an inverse relation between the degree to which organisations say that they are quality organisations and the quality of what they offer in the classroom... because people spend an inordinate amount of time going through bureaucratic quality processes, form filling, box ticking etc., and it impacts on the amount of time and effort devoted to the classroom”.* [ENG1g].

This point has also been recognised more widely within HEIs (Stensaker et al., 2007). The responses given implied that the subunits were supported within the organisation because they are successful but there was no actual response to the content of their evaluation reporting; the practical demands were not felt to be followed up. This furthered the idea that processes operated as a control mechanism rather than an improvement based one. Interviewees appeared to treat the systems in this way when responding to demands.

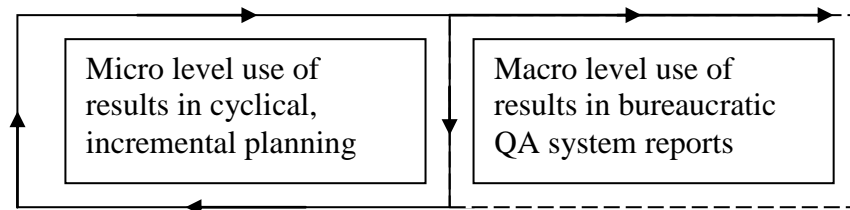
Interestingly, not one of the respondents said that evaluation, per se, was an unimportant task. Some questioned the overemphasis of my research focus on what was a fairly mundane and bureaucratic process, but at the same time reflected over the importance of their individual interaction with students and the feedback and assessment they were doing at programme level. This seemed to be separate from the more formal procedures that were meant to satisfy certain criteria and led to little or no feedback from the higher levels in the organisation.

This meant that in the main the subunit members perceived themselves to be operating at the knowledge management and organisation phase of the evaluations, implementing frameworks from higher up in the system or externally to the organisation. There was a general agreement that subunit members felt that the formal frameworks that were the main focus of evaluation for the respective organisations did not provide them with the information needed for within programme development. They still referred to their own professional judgement, but this was based on much more informal methods. For example, in ENG1 it was commented that it was difficult to gain feedback from students generally, and that having to fit the questions into the generic framework did not always provide the information needed. Where information



was provided with regard to local needs in terms of improving the quality of programme delivery, this feedback was on occasions been perceived to have been ignored. There was argued to be little feedback in the loop, questioning whether the purpose of the evaluation was anything more than symbolic.

As figure 16 below implies, the responses from the programme providers showed that the information and data ascertained from evaluations is used at programme level to enlighten their cyclical and incremental programme planning and revision, whilst the broken line indicates the lack of feedback related to information reported upwards in the internal system<sup>188</sup>. As was seen in chapter 3 the perception of evaluation utilisation can be as important as its actual utilisation. There was again agreement on this issue across the subunits; respondents had no indication of what the information was used for or how it was dealt with. Additionally, across the organisations respondents felt that this “control” format did not sufficiently serve their purpose requiring them to supplement the processes at local level. This would not normally be problematic but for the time the central evaluations were perceived to take to implement, combined with the lack of feedback from the higher levels.



**Figure 16: The perception of how evaluation data is used within the organisation**

One of the problems that Dahler-Larsen and Krogstrup raise, and that might apply even more closely to the programmes under investigation here, is that evaluation can be thought to reduce the complexity of human processing activities to the allocation of simple indicators, which are often applied in an “obscure” way (2000: 288). This happens because central goals and purposes are not always clear to evaluators, and too complex to develop evaluative criteria for. This leads to the decision about criteria for an evaluation becoming an interpretation process, which in turn is thought to aid constituency about what the programme itself is about, where goals will then appear as “retrospective constructions” (ibid.). The authors draw this perspective from Weick's notion of

<sup>188</sup> This might be thought similar to March and Olsen's (1979) concept of incomplete, or broken, learning cycles which cause ambiguity in organisations. This work is linked to understanding organisational learning and area for future research in following up the findings of this study.

sensemaking. At the same time such behaviour can help construct future practice for the field (Dahler-Larsen & Krogstrup, 2000: 290).

But despite any “macro-institutional” legitimisation of evaluation standards, it is also possible to see at the micro-level that evaluator values and behaviour can influence the level of choice of criteria; and these interpretations would seem to greatly influence the process, adding something new (Dahler-Larsen & Krogstrup, 2000: 290). As a result it is claimed that the criteria for an evaluation are not just taken for given despite the existence of central demands and control for their implementation. These are thought to be moderated by differences in the strength of demands and extent of control, while at the same time internal processes are obscured by the trend for evaluation to be seen and accepted as a legitimising organisational formula (ibid: 291).

Respondents were therefore asked to discuss their degree of perceived choice over *alternatives* in decision making. In section 5.3 it was recognised how Brunsson (1990) saw the importance of understanding how the “values, beliefs and perceptions” of decision makers inform the choices offered; the type of decision process that unfolds and the responses that ensue. In NOR1 respondents suggested that they had been able to have a higher degree of input in evaluation processes, as the organisational system had not fully been implemented when an approach was made and the competence of the provider team within this field had been recognised. Thus there was an attempt to meet internal demands within the wider organisation, despite frustration over the relative simplicity of reporting frameworks. In NOR2 there had been relatively little to no input from the teams into the frameworks. In ENG1 and ENG2 respondents spoke of the organisation putting the quality assurance frameworks in place, again with little input from the teams and little feedback to reporting procedures.

#### **10.4 Decision responses to demands: *issues of organisational coupling***

Linked to the reflections of Olsen in the subsection above, Meyer and Rowan (1992) recognised that in educational organisations there is often a separation between teaching activity and bureaucratic steering structures, i.e. that there has traditionally been little interest by the latter in the former. They maintained that this was to do the lack of “evaluative activity” as compared to, for example, private businesses. As has been noted throughout this study, the amount of external pressure to assure quality of academic delivery has increased greatly since that period of time. And yet there are suggestions from this area that despite this increased interest and demand for accountability, focus on the actual academic quality is more related to limited, and challenged, measures of quality of instruction, most notably perception of student satisfaction and throughput of students. These issues relate more to funding issues, rather than issues of study programme quality, offering only indirect interpretation. Designs developed within the organisational framework were consistently described as a “fairly

bureaucratic” response. As a result the subunits seemed to become decoupled from the institutional frameworks and as mentioned above, the participants on the programmes did not always appear clear about what the purpose of the evaluation process was, even when they were perceived to have the time to take it “seriously”.

As a result of these general impressions I turn now to look further at the internal relationships of the subunits as well as their relationship to their own organisation. I have been careful to try not to draw too strong conclusions concerning a comparison of the different subunits and the degree of cohesion that was found within them, each of the subunits was organised differently and covered a variety of different programme portfolios within their respective institutions. As a result the particular structures and organisations of these portfolios explain a great deal of the variance concerning constituency and cohesion of the different subunits.

#### **10.4.1 Impact of subunit behaviour on the organisation**

In section 9.2 a contrast was presented between the subunit at NOR1 and those at NOR2 and ENG1 and ENG2 with regard to their engagement on evaluation issues within their respective institutions. Members of all the subunits commented in some way about the existence of an academic community where discussions about the programme took place, and with that some reflection over decisions concerning evaluation. However, in NOR1 there was seen to be strong focus upon an approach which dominated the reflections of all the respondents as they spoke about these processes. In 9.2.1 focus was placed on the perceived proximity of group members within the subunit at NOR1 related to decisions made about evaluation. The various members spoke of the collective qualities of decision making as an academic group both in the formation of evaluation and when discussing the impact upon successive decision processes. These responses strongly reflected theoretical descriptions of collegial decision making models, which are often considered a normative approach to management and decision making (Bush, 2003).

As was referred to above, there are recognised challenges to collegial decision making models. Part of the issue concerns their normative nature (Bush, 2003), that they tend to argue how processes ought to be, rather than how they actually are (Burrell & Morgan, 1979). Some of the issues raised as problematic for collegial approaches were reflected in the responses of the subunit members. In particular, subunit members referred to problems concerning sustaining collegiality when faced by time restraints and bureaucratic requirements (Bush, 2003: 81-84).

The reflections from respondents in NOR1 highlighted a process that I have tentatively called here *collegial construction*, which was characterised by a high degree of internal communication and cooperation, a fair degree of within group homogeneity, a history of more formalised, yet more autonomous and locally

controlled evaluative activity prior to the introduction of the formal QA system. Interestingly this was combined with willingness by the subunit, and particularly the senior members, to seek to share experiences and influence the development of the formal QA system. This cohesion within an occupational community and its relationship within the wider organisation have been recognised to bear a positive bottom-up effect on evaluative systems. Dahler-Larsen in particular discusses how autonomous occupational communities that are characterised by homogeneity can offer a renewing role to their organisations (1998: 155). This renewing role is a positive response to demands placed at higher levels. Additionally, with its strength at the base of the organisation compared to generally weaker approach at higher levels would appear to have important institutional influence. This might also allow a greater balance of the tensions faced by those at lower levels as they attempt to justify their professional approach against wider standards, in this case which often appear to mismatch.

It was when reading through the transcriptions and analysing the descriptions of these processes of decision making as a perception of strongly collegial action that I began to code the data at a node entitled “collegial construction”. A later search revealed this term had been used in a limited way with regard to peer related dialogic interaction of teachers, with particular regard for professional development (Maggioli, 2003; S. Owen, 2005). Maggioli indirectly drew his approach from the work of Sparks (2002) who aimed to contrast collaborative forms of on-site professional development for school leaders with the individual off -site training programmes. Owen described a phenomenon that bore some of the hallmarks of a form of a community of practice underpinned by shared learning within a social setting. But in this study I have preferred to use the concept of occupational community because as Cox (2005) notes it focuses more on common work situations and structures, as opposed to directly joint practices. Collegial construction also appears somewhat similar to Westerheijden, Hulpiau, & Waeytens (2007) concept of positive learning effects<sup>189</sup>, while as will be seen below, dismissive submission might also be similar to negative learning. Westerheijden et al. noted how responses to demands for quality assurance and evaluation vary from pragmatic acceptance to rejection as “meaningless ritual”. These learning effects can influence “internal dynamics”, after developing capacity for self-evaluation become more engaged in contributing to the improvement of the institutional quality culture. Nuance within this continuum confirms the impact of “positive learning”. The field of organisational learning and culture is extensive. In this study I have delimited the focus more directly onto understanding the decision processes as a response to demands. Further understanding of collegial construction will require further investigation of wider internal learning processes but here I have focused more upon the decision response have chosen to interpret through understanding action as a sensemaking and sensegiving activity.

---

<sup>189</sup> as outlined in section 4.2

Within NOR1 group members described the “culture” within the group, much of which was built upon shared experience. Weick (1995) notes that shared experiences, more than shared meaning, drive organisations forward, shaping and being shaped by sensemaking processes. In such a view culture is seen to be “what we have done around here, not what we do around here” (1995: 189). According to Weick, meaning will not always be the same, but the collegial reflection and recollection around common experience is important for building culture. Groups act on this basis, despite interpretational differences, but “tied together by the common origins of those understandings” (ibid.). It is at that point that groups notice difference from their shared experience in the wider organisation. In terms of the responses about evaluation, the group at NOR1 demonstrated a degree of shared reflection and discussion not present in the other groups. I return to this point further below when comparing collegial construction with dismissive submission in the subunits under study.

The processes of discussion in NOR1 appear also to have enabled sensegiving within the subunit, as well as coupling the sensegiving to other groups within the wider organisation. The group members perceived that they had, on occasions, been able to influence evaluation processes within other subunits as well as influencing policy makers at higher levels. As was seen in section 5.5.4.1 these processes involve understanding (sensemaking) and the attempt to influence (sensegiving). This was recognised from the work of Gioia and Chittipeddi (1991). What was interesting in this study was that these processes began bottom up in response to a top down directive, in line with Dutton et al.’s application of *issue selling*, “the voluntary, discretionary behaviours organizational members use to influence the organizational agenda by getting those above them to pay attention to an issue” (Dutton, Ashford, Lawrence, & Miner-Rubino, 2002: 355). This process is thought to be both political and contextually embedded, recognising the “inherent intersection of micro *and* macro forces in determining change patterns in organizations” (Dutton, Ashford, O’Neill, & Lawrence, 2001: 717). While Dutton et al. (2002) investigated this behaviour in relation to individual contextual sensemaking within organisations, in this study the collective action of NOR1 also seems to exhibit similar conduct. An example of this includes email responses to Institute and Faculty leadership with regard to inadequacies with proposed and current evaluative frameworks and suggestions for replacement components for related models. Future research should explore further whether these responses can be considered as counter-episodes to organisational strategic change within recursive processes.

The findings of Dutton et al (2002) were followed up in the work of Maitlis and Lawrence (2007) identifying “triggers and enablers” of sensegiving. In applying Maitlis and Lawrence’s framework, interesting reactions can particularly be seen from the subunit at NOR1. As was seen in chapter 5, the Maitlis and Lawrence found sensegiving “triggers” at stakeholder level to be based upon “bounded responsibility”, that is, the need to take responsibility for something like a

perceived gap in leadership to offer alternative accounts, allied to perceptions of *important issues* for the organisation characterised with *incompetent leadership*. They recognised the degree of difficulty in ascertaining when issues are important enough to engage in sensegiving processes of influencing interpretation (2007: 76).

There did not appear to be any difference with regard to triggers for sensegiving across the subunits. There is a degree to which all the subunits in this study shared a conception that evaluation was an important issue within the wider organisation but not dealt with competently by higher levels, or reduced to bureaucratic rule making. It is though amongst the “enablers” that greater difference is recognised. The enablers for sensegiving also tell us something about the presence or awareness of sensemaking within organisations (Maitlis and Lawrence, 2007). Apart from NOR1 the other subunits referred mainly to the lack of feedback combined with the limited possibility to influence organisational application. These subunit members all related accounts of not being heard on the issue and lack of discussion. In each of these other subunits there was also a limited degree of internal discussion and coordinated action with regard to evaluation.

In terms of sensegiving activities as “enablers”, it was reported in the previous chapters how the subunit at NOR1 had attempted to influence the frameworks at the agenda phase of evaluation as opposed to merely implementing at the operational stage. There was a perception, following Maitlis and Lawrence’s work (2007), of the presence of *process facilitators* which provide opportunity for sensegiving. The members, both individually and collectively in their module teams had written memoranda to be discussed more widely in the group. These discussions took place at regular as well as ad hoc team meetings. Respondents also outlined how a training day was also specifically arranged for further more in depth discussion and planning for practice.

Additionally subunit member responses characterised what Maitlis and Lawrence (2007: 57) referred to as *discursive ability*. At NOR2, ENG1 and ENG2, there was clearly noted to be *issue related expertise* amongst those within the subunits, with many very experienced in the academic field of study and as evaluators. However it was only at NOR1 that respondents reported the presence of issue related legitimacy, which allowed them to “construct and articulate persuasive accounts” (ibid) and present their information to those outside of the subunit, particularly those in positions of authority. Respondents outlined how members of the subunit and the subunit collectively were observed both internally and in the wider organisation as skilled and successful, in terms of the academic field of study and also in evaluation.

Examples of enablers both within unit and with the wider system were presented at NOR1 but not at the other subunits. The subunit members at NOR1 attempted to make sense of the new demands and how they fitted to the ethos of

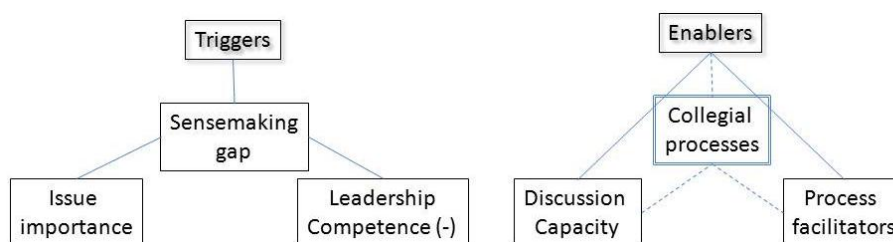
the group, not least trying to match their approach to research findings from the subject field. Based on their shared experience as educational professionals with knowledge of evaluation processes, they challenged the quality assurance framework that was to be implemented across the institution. In turn different forms of engagement were made with those higher up within the organisation. These processes are outlined in table 15 below.

**Table 15: Triggers and enablers of sensegiving in NOR1 (after Maitlis & Lawrence, 2007)**

	Triggers		Enablers		
	Sensemaking gap		Discursive ability		Process facilitators
	Important issues	Incompetent leadership	Issue related expertise	Issue related legitimacy	Sensegiving opportunities
NOR1	<b>Evaluation models and quality assurance structures</b> (incorrect focus) (reductive processes) (simplistic models) (lack of feedback)	<b>Lack of involvement</b> (little overlap) <b>Lack of feedback</b> (from upper echelons)	<b>Professional experience</b> (role of academic) (internal / external experience) (within subject field experience)	<b>Group perceived as skilled</b> (as viewed by leadership) (as viewed by other subunits) <b>Group perceived as successful</b> (as viewed by leadership)	<b>Within unit processes</b> (emails, memoranda etc.) (regular team meetings) (ad hoc discussions) (training day) <b>Within wider system</b> (presentation of data) (cross faculty meetings) (email responses)

Maitlis and Lawrence note that while sensegiving has often been presented as the “employment of linguistic devices, such as metaphor or narrative, which transcend specific organizational contexts”, they see it as a specifically grounded situated practice (2007: 79). The authors therefore proposed that enablers were relevant for particular groups, those with issue related expertise, legitimacy and opportunities to engage in sensegiving, operating as an

institutionalised practice. The findings from this study might go some way to suggesting that this behaviour is **moderated** by the decision making processes and practices of the groups, as presented in figure 17 below. In the studies here it suggests that the presence of a collegial approach leading to tighter coupling within the subunit provided the framework for further sensegiving activity. This, however, ought to be studied further with regard to another issue that Maitlis raised for further reflection, the existence of antecedents of organisational sensemaking, noting that organisations with “dynamic sets of stakeholders... always seem to contribute to the formation of accounts, irrespective of the issue” (2005: 45). In this study focus has been placed upon one particular framework of issues with the aim at improving understanding of decision processes surrounding evaluation. The subunits under study are also involved in decision making about many other issues and it would be interesting in other studies to compare decision making across such areas. In addition it might be helpful to develop on the data from this study by investigating further the linkage between decision making processes, discursive ability and process facilitators, as well as considering further variables that might mediate or moderate such behaviour.



**Figure 17: Collegial influence on enablers of sensegiving**

In section 9.2.2 focus was placed on the other subunits. These had not exhibited the same degree of internal cohesion, particularly with regard to evaluation. In most cases respondents characterised their reactions to evaluation demands as I have termed as “dismissive submission”; in many ways disagreeing with the form of the demands and frameworks of evaluation placed upon them, but needing to implement them within their organisations. This shows similarities to the way Dahler-Larsen applied Berger’s<sup>190</sup> (1964) concept of “role distance”, a socio-psychological construct where individuals implement the “organisationally defined minimum standards, but with the least possible engagement”, to the evaluation field (1998: 153). This is also similar to Westerheijden, Hulpiau, & Waeytens (2007) concept of negative learning effects, which take place when staff “learn to play the tricks” of QA without it

<sup>190</sup> From Berger, Patrick (1964). Some general observations on the problem of work in P Berger (ed). *The human shape of work*. New York: MacMillan Company.



affecting the “internal life” of the quality of teaching and research, also known as “window dressing<sup>191</sup>” (2007: 298).

In outlining the data from the interviews it was discussed how the members of the NOR2 considered themselves completely removed from the central evaluation process. As a result academic staff focused only upon the performance indicators that the wider organisation had deemed important or that the individual subunit members had interpreted as important. In this particular case it had been the students’ feedback on the quality of teaching that they had given some attention to. While the subunit members disagreed over the efficacy of such measurements they registered that the organisation assessed its staff in relation to these. Additionally these evaluations were seen as ritual acts. In the groups where the coupling appeared looser and there was less cross unit discussion, at least with regard to evaluation and resultant decision making, there was more of a sense of decoupling that Dahler-Larsen has discovered (1998:121). This reflection fits within the institutional model, where one also sees the sense of colonisation of ideas. This template offers a good explanation of the processes described. Additionally, while members of the subunit at ENG1 had expressed the importance of the academic group, they admitted to spending little time in response to the formal evaluation frameworks. The subunit at ENG2 was recognised by members to have had no discussion at all about these issues. It is not my intention to give the impression that there was no discussion about programme development and improvement, but this was not seen to be a collective activity. There was little major discussion involving the module leaders; most reflection was done at the module level. It appeared that as long as the feedback fell within the premise of the programme aims and goals, this was deemed acceptable.

In NOR2 there was noted to be a significant difference between the internal and external models which had extended the respondents’ views about the evaluation process. An interesting scenario had developed where there was greater central control over the formal quality assurance system implementing evaluations in a standard operating procedure, alongside the more “instrumental” use of evaluation data at the commissioner level. It was perceived that the providers operated their own informal evaluations of the programme implementation and development of the participants. The sub-group was able to glean some feedback from the system evaluations in their own organisation, which they considered to be ritual like; receiving no feedback from those higher up in the system. It was claimed that the internal evaluations dealt with implementation issues but had no effect on the inner life of the programme. This, as in NOR1, was felt to be taken up in the programme committees in the organisation, but did not appear to be linked to evaluations at all. At the same time, the sub-group had suggested that the evaluations run by external commissioners offered some useful data for their

---

<sup>191</sup> This term was used as a translation to a response from NOR1 as seen in section 8.7

own programme improvement. However, the tendency had been that this data was used instrumentally by the commissioning bodies as evidence that could be used in negotiation for programme change. The problem was that much of this data was drawn from participant reflections of satisfaction as a proxy for learning and development.

**Table 16: Collegial construction and dismissive submission compared**

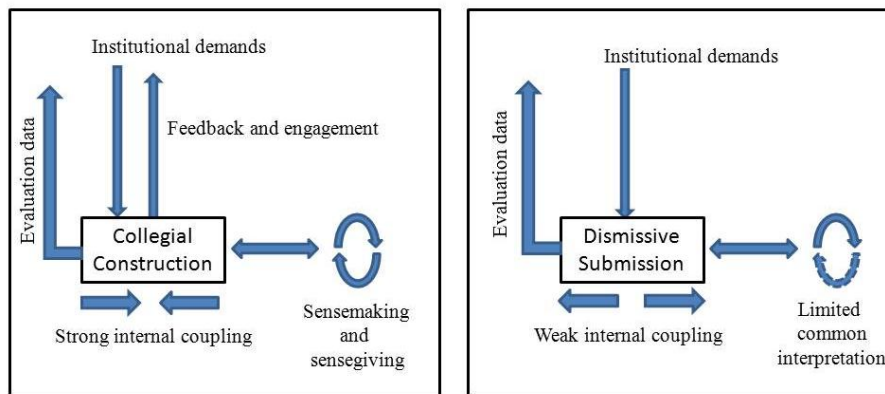
Collegial Construction		Dismissive submission
Collective action	Continuum	Collective inaction
Local initiation and power sharing		Central initiation and download
Participative and critical approach		Critical rejection
Development focused		Decoupling and disengagement
Sensemaking /sensegiving approach		Individual interpretation

The perceived impact of these two approaches on the organisation is interesting. As can be seen in figure 18 below the institutional demands were expressed to be the same in both cases and each of the subunits collected and reported the data as required. The main difference is the extent to which the NOR1 subunit developed a frame of feedback and engagement on the issue of evaluation. As was seen in section 9.2.1 there was perceived to be a strong internal coupling within the group and each of the respondents spoke at length about the collective discussion about evaluation and improvement. Additionally the group operated as a collegial unit with processes described in a similar way to Weick’s (1995) view of sensemaking and sensegiving as outlined in section 5.5.4.1. The processes were not static, described as “dynamic” and noting that as new members joined the group so they were encouraged to come with their input. The academic group was described as a “discussion group” with a “culture for change”. As one member was seen to describe it,

*I have never experienced a place where there is such a high degree of interest and willingness to change and develop systems. So we have a very development oriented group here, and as such we lie pretty much ahead (laughs a little). It’s actually true! [NOR1i].*

The group described how it had seen some evidence of its suggestions impacting the wider organisation and they continued to use the channels they had opened to feedback ideas within the system as a whole.

The other subunits were not characterised by the same degree of internal cohesion nor relationship with the wider organisation. The process of decision making about evaluation in the wider organisation were described as *dismissive submission*. Dismissive submission suggests that they performed the tasks as required but were opposed to their form. Additionally the groups had little in the way of feedback processes with other levels. Where there was any engagement it was to keep updated with the latest demands or receive limited information and training courses that were arranged. These meeting points were mostly to ensure that the groups provided the information as needed by the institution as a whole. The subunits appeared to become increasingly more decoupled from the central organisation and were also perceived to be characterised by little internal discussion with regard to evaluation development. This was despite the relative expertise and experience of the various individuals in the topic. In these groups, evaluation that would satisfy the demands of the organisational frameworks appeared to develop, as Dahler-Larsen (1998) has previously described, as a normative “ritual reflection”.

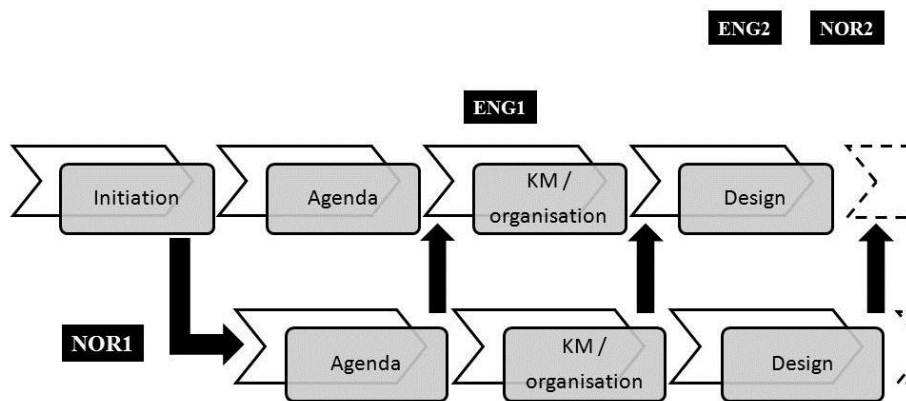


**Figure 18: Collegial construction and dismissive submission compared**

### *Elements of the decision process*

When reapplying Dahler-Larsen’s framework of evaluation elements it is possible to plot the activities of the different subunits. The subunit at ENG1 described their entry point as during the knowledge management and organisation phase of the evaluation decision making process, that is, when things need to be implemented, and their responses exhibited them to be having some limited influence over the evaluation design. The subunits at NOR2 and ENG2 were perceived to be decoupled from these centralised processes. The subunit at NOR1, however, appeared to become involved at an earlier stage in the process of development, developing their own internal processes and offering feedback at the various stages. They were thus able to influence the frameworks at the initiation stage, helping to set the agenda. This was thought

possible as processes concerning evaluation development and responses to organisational frameworks were concurrently going on within their group and thus could become applied to the wider organisational initiatives as they arose. This can be seen in figure 19 below. This does not of course suggest that every request and suggestion from NOR1 was listened to or that there was no impact from the other subunits, account is rather made of their responses based on their internal perception of decision events and influence surrounding evaluation against the organisational frameworks.

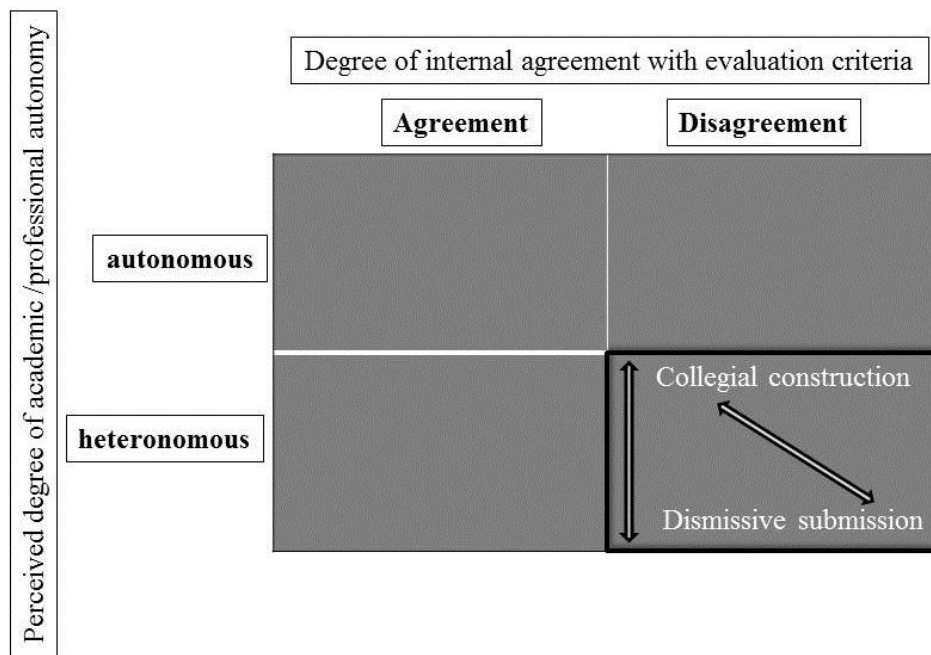


**Figure 19: The subunits within elements of the decision process**

In considering the wider impact of these characteristics on the organisation I have reapplied Dahler-Larsen's framework of occupational community reactions to evaluation demands as was seen in tables 6 and 7. This framework focused on the degree of a group's perceived autonomy within the organisation compared with the degree of the group's agreement to the evaluation criteria they must implement. Focus on the subunits in this study shows them to be generally placed in the fourth quadrant related to their lack of autonomy over decision making about evaluations within the institutional frameworks and their general disagreement with the evaluation criteria adopted by the organisation.

From the findings of this study these ideal types (Dahler-Larsen, 1998: 149) appear to operate as dimensions on a continuum, particularly within quadrant 4. As is outlined in figure 20, this means the fourth quadrant is developed to consist of collegial construction where a group is committed to development as influence; reactive and operating with a degree of autonomy but still needing to implement organisational demands. Dahler-Larsen suggested that groups in this quadrant could either live with the discrepancy to their own criteria or revise their own criteria (1998: 152). It was the former response of role distance that appears to fit most closely to that of dismissive submission. As will be seen below when discussing constitutive effects, there is also a strong possibility that the latter will take place and a group's evaluation standards will be challenged

over time as the ideas become institutionalised. Interestingly, though, the process of collegial construction appears to challenge the evaluation criteria laid down within the organisation while recognising that the group cannot strictly be described as autonomous with regard to the demands for evaluation. However, as Weick (1976; Weick & McDaniel, 1989) notes, the governance structure of universities is traditionally weak, which often allows groups to act upon their own will.



**Figure 20: Revision of Dahler-Larsen’s fourth quadrant within a continuum**

Such activity can be seen to have a positive effect on the organisation. As was seen in section 5.7.1, Dahler-Larsen outlined how organisations would often consider heteronomous groups as more preferable when developing their evaluation frameworks, as they are easier to control. This can be linked to organisational models of decision making noting that these centralised systems often react bureaucratically, setting up control structures that become institutionalised within the organisation (1998: 154). In such systems evaluation plays the role of a control mechanism rather than offering development focus. Dahler-Larsen suggests that relatively autonomous groups can be more preferable for organisations as these can “play a renewing role” (1998: 155). There are, however, some cautionary reflections. Perhaps the most significant difference is that the subunit at NOR1 had signalled their interest in involvement

before the frameworks had been finalised. Further investigation would be interesting into the background for these events and how the leadership at higher levels had received and responded to this approach. In addition, the situation with regard to commissioners in Norway is more complex, given that the definition of programme evaluation will be part of a tendering agreement and contract rather than a part of the organisational framework. There are perceived to be many factors that will influence the place on the continuum. It would be interesting to compare the subunits to other academic groups within the same institutions. It would also be helpful to investigate how groups might vary in terms of their agreement with the organisational criteria.

*Institutional mismatch: a pressure on decision making*

The tendency to consider that QA systems increase rationality and tighten coupling within organisations might well be a misnomer and lead to misdirected attention within an organisation. The current trends within quality assurance in HEIs appear to have more to do with adaptation to external demands for financial accountability and the onset of globalisation than improvement of the learning process (Westerheijden, Stensaker, et al., 2007). As such there is often an internal dissonance between perceptions of quality, focus of assessment and purpose of structures. Research from Rosa et al. (2006) recognised that it was programme coordinators who were most interested in the learning processes, whilst those higher up in the system focused more upon results and the internal structures. Such ambiguity in the system appears also to be reflected within the decision making about programme evaluation, noted here in relation to postgraduate programmes for school leadership development. Interpreting and managing such ambiguity becomes an issue for programme staff. That quality assurance systems are described as both “policy instrument and management tool” (Westerheijden, Stensaker, et al., 2007: 2) only increases the likelihood of disparate perception and response to their implementation.

Weick (2001) also noted problems of control inherent within loosely coupled, complex organisations that are characterised by segmented structures. One of the implications for this study is that management within the system is not necessarily considered to be about managing the organisation but rather about managing the “process that manages the organisation” because it is too complex to design an operating structure from the top and therefore attention is placed on decision structures (2001: 38). Operating structures are subsequently designed within the segments. This creates an interesting situation when trying to implement organisation wide evaluation structures, especially when the form and focus are developed externally to the organisation. When decisions are interpreted and implemented locally greater sources of ambiguity arise. In similar cases the “normative or formal structure” is often found to be “decoupled” from the “operational structure” allowing institutional demands to be met while retaining “some autonomy of action” (W. R. Scott, 2003: 214). This latter view of “buffering” is drawn from the work of Meyer and Rowan,

who maintained that loosely coupled systems tend to become even more loosely coupled when attempting to maintain their “ceremonial conformity” and that their formal structure differs greatly from actual work patterns (1977: 341). A consequence of this would be that “evaluation and inspection systems are subverted or rendered so vague as to provide little coordination” (1977: 343) and as evaluation and inspection otherwise might “undermine the ceremonial aspects of organizations”, they would seek to avoid or minimise the activity as much as possible (1977: 359). This was also recognised by Thompson (1967).

But, according to Dahler-Larsen (1998: 80ff) the picture has become even more greatly nuanced, especially since the development of wider societal demands related to NPM. The issue might then rather be about how to evaluate under such pressures rather than how to avoid it. Has already been suggested from the interview data noting that the groups do evaluate but not to same extent as might be required or anticipated. There was no real suggestion of avoidance, rather one of complexity and mistrust of the wider systems. Therefore this situation is more nuanced. As has been stated, Dahler-Larsen reflects that the main thesis about evaluation from a perspective of institutional theory is not that it is a technical solution to a problem, but rather a “ritual” reflection or response (1998: 79). This also resembles a sensemaking response. At one level evaluations are considered only to have a symbolic effect, loosely coupled to organisational activity. But Dahler-Larsen also goes further to suggest that the act of evaluating, even though apparently only symbolic, can itself direct behaviour and produce constitutive effects. While this sounds like an argument to do with utilisation as has already been recognised, it is difficult to perceive the evaluation systems in organisations as developing in a linear fashion and therefore the perception of previous utilisation is considered to consistently affect future decision making.

With particular regard to the leadership of HEIs, Weick noted the general difficulty in evaluating the quality of “intellectual products”, and further that “[s]egments within the university decide key issues, such as teaching and admission requirements, and the only control presidents have over these subgroups is money and final approval of personnel decisions” (2001: 39). The implications for this study apply less to the role of the top leadership, than the implementation of segmented organisation further down the system. As Weick notes, where “subsystems are loose and responsibility is delegated to groups rather than to individuals” who then begin to “act like top management” (Weick, 2001: 39). There is no question that evaluation should take place and this becomes reinforced downwards. Drawing on the work of James March, Dahler-Larsen agrees that the idea that evaluation must take place has become a “protected discourse”, going unquestioned (Dahler-Larsen, 1998: 16-17). He amplifies this by claiming that evaluations, having become institutionalised procedures and rituals, gain their own logic and can produce effects beyond that which was originally intended (Dahler-Larsen, 1998: 17). This positioning of

evaluation, as he sees it, between systemic knowledge acquisition and political action is an area not usually attributed to it, for example in the rational model.

Therefore Dahler-Larsen reiterates that while definitions of evaluation often presuppose it to be a rationalistic endeavour, it is directed by many competing factors, not always as visible as one might imagine (Dahler-Larsen, 1998: 22). In this way Dahler-Larsen brings to the forefront a major issue for evaluation generally, and specifically within the field of school leadership programmes, that an often demand for greater transparency and systematisation of evidence from evaluations should be tempered by a realisation of the complexity of processes operating at the same time. This is often a problem of whether one is looking at the purpose of an evaluation or its function, and subsequently recognising that there often is a discrepancy between them (Dahler-Larsen, 1998: 24). Indeed he later noted that when evaluations are made mandatory or obligatory, they become ritualised, or institutionalised (2006b: 35). One effect of this is that the framing of evaluation slowly influences change as it becomes institutionalised.

#### **10.4.2 Impact of the evaluation frameworks on the groups: constitutive effects**

While the previous section dealt with the impact of subunit response on the organisation, this section deals with the impact that evaluative frameworks can have over time, in this case on the subunits. In section 5.5.4, I outlined how institutional theory has begun to be applied to evaluation theory. One of the most interesting areas of study concerns that of the detection of constitutive effects. As was seen in chapter 5, this phenomenon with regard to evaluation has been described by Dahler-Larsen and Krogstrup (2000: 295), who suggested that the direction and focus of evaluation itself is another factor which affects future decision making, concerning how evaluation can “co-construct the social reality” surrounding an evaluation. The authors considered that “constructions are obscure, unintended or at the very least non-calculable”<sup>192</sup> (Dahler-Larsen & Krogstrup, 2000: 295). The complex web of demands outlined throughout this study has produced a number of different interrelated pressures on the subunits. These pressures as described by the respondents were experienced through the evaluation frameworks in the quality assurance systems, including the pressure for systematisation, the pressure for succession (completion), the pressure for satisfaction and increasingly, the pressure for subsequent impact. Each of these factors may be considered to have constitutive effects. As was seen in chapter 5, the authors see these effects within “three aspects of social reality”: the material, the time related, and in social relations and identities (ibid.). As the authors recognised, in understanding these processes one can observe the difference in approaches within and across standardised systems. The authors also argue that the way an evaluation is standardised within an organisation will affect its use,

---

<sup>192</sup> My translation from Danish.



and it would seem the perception of such use. Examples of these processes from the data collection are outlined in table 17 below.

**Table 17: Examples of constitutive effects of evaluation (after Dahler-Larsen and Krogstrup, 2000)**

Aspect of social reality	Meaning	Example from data collection
<b>Material / Content</b>	Questions raised here include how evaluation can frame interpretations, orientations and actions; - the system becomes the mental frame that work is considered through.	<i>“But in a way we’ve relied very much on kind of gut feeling and we know that what we offer is valuable and worthwhile because people tell us so... But Government Agencies want to see hard data, but it’s quite difficult to get that data, when you start to unpack the complexity of the link between a course or an intensive Master Programme over 2 or 3 years and the quality of leadership”. (ENG1)</i>
<b>Time / timing</b>	Questions raised concern the point an evaluation will take place, and how much time and resources it will take. Different timings will produce different pictures of evaluation.	<i>“We satisfy the formal demands but we use too many of our resources for my taste. This affects my capacity to work; it affects the profitability of [the organisation] and our capacity to sell the programmes”... (NOR2)</i>
<b>Social relations / identity</b>	Questions raised here include who has the right to be heard and how are roles redefined?	<i>“So another reason that we put evaluation on the agenda is because there are actors in the field who challenge us to do so” (NOR1)</i>

An additional example of *material changes*, or that which is conceived as important by the different groupings, related to different demands, includes the increasing focus upon throughput and student voice, which are required by HEIs when reporting higher up in the system and to the national authorities. While these were not considered the main purposes of evaluation by the members of various subunits, they recognised them as the areas of most importance in the quality assurance systems. In other cases, however, it was seen to be the mere

action of evaluation that was in focus rather than its subject matter. As was related from the responses of NOR1, the reflection came that *“it isn’t first and foremost what we do; it is actually more about that we undertake an evaluation that is in focus”*. And as has also been stated, in England there was a growing focus on post programme impact, particularly related to external funding. As one respondent from ENG1 noted:

*one of the conditions of their funding or partially funding of these programmes is that they want some impact data. So this is making us lecturers think much more, from the outset, about how can we collect a kind of impact data. And that’s also linked with that whole move in Britain, and indeed elsewhere, of evaluating professional development, whether it’s leadership development or teacher development, saying ok well what difference is it having. So a lot of work is being now done on impact evaluation. [ENG1g].*

As was related in chapter 7 one member of the ENG2 subunit had observed an even stronger connection between funding and evaluation, exclaiming wearily that *“alongside getting funding goes the filling in of an evaluation report...”* This was also noted to have started to influence the evaluation focus on programmes that were not under external demands to ascertain and report on impact in their thinking about how to evaluate programmes.

Secondly, with regard to *time and timing*, there was noted to be a growing expectation across the various subunits, that a positive influence would be demonstrable and forthcoming in a short space of time after programme completion. These expectations of impact from the programmes, call for a demonstration of a type of results that are not considered to strictly fit with the purpose and nature of the programmes. While demands from policy makers for impact upon on the workplace were felt to be fewer in Norway, where they had arisen from commissioners they were perceived to be at odds with the central tenets of the programmes. In NOR1, for example, members spoke of the centrality of developing leadership in social and collegial setting, while in NOR2 focus had been placed on engaging system change. Neither of these areas lends itself to quick change. In these cases the subunit members had managed to “talk around” the programme commissioners into changing their focus. A sterner problem was faced in England, where, as was related above respondents in ENG1 talked about how funding agencies were beginning to require short term evidence of impact in order to secure future funding, and on reflection this had caused the team members to focus much more on this area in their own evaluations. Dahler-Larsen (2007) also helpfully outlines how constitutive effects will establish time frames, in particular regard to when a particular outcome will be expected and accord greater attention to specific activities. He uses the pertinent example of “customer satisfaction” as an “evaluation criterion in higher education”, which leads to “the idea of instant gratification... [and tends] to redefine the role of the teachers towards the students” (2007: 29).

Members from NOR1 spoke of the tension of now balancing what students wanted on the programmes against what the academic team felt to be right for them. This begs the question, should one therefore change the programme if students struggle with it?

*Identity* challenges were also significant, with movement away from evaluation based on professional perception towards the implementation of standards based assessment. As was exemplified above, the reactions of respondents in ENG1 and ENG2 and NOR2 were classified as “dismissive submission” to the new demands, which shows similarities Dahler-Larsen’s (1998) application of Berger’s<sup>193</sup> (1964) concept of “role distance”. Respondents agreed to the importance of evaluative systems but not to the newly developed standards and demands for instantaneous impact. Responses from NOR1, however, indicated that the members of the subunit attempted as far as possible to influence the system within the institution and actively sought to offer alternatives that could be used, while strengthening competence within their own sub-group. The issue of impact assessment was also linked to the question of identity in England. The wider pressures for impact, especially since the advent of the NCSL, had seen the academics increasingly needing supporting data, even while they recognised its limitations. Already these demands were beginning to be transposed into the internal governance of the organisation. This appears to fit with the reflection, drawn from the work of Munro, that “many indicator systems emphasize how organizations internally control themselves rather than make direct examination of the practice itself” (Dahler-Larsen, 2007: 28). As a result this appears to lead to additional intentions for programmes, that in the future might become the reason for their very existence. In chapter 2 I related this to Ball’s conception of performativity. Additionally, Ball recognised that organisations within “different market positions are likely to arrive at different forms of strategic response” (Ball, 2003: 225). Ball suggests that those in a “weak ‘market’ or performance position” become an “auditable commodity” and might “submit to becoming *whatever it seems necessary to become*” for survival, where decision making is based purely upon improvements of performance. In contrast, those in a “stronger ‘market’ or performance position” might become complacent or attempt to retain “commitment to non-performative values and practices” (ibid.).

Further tensions revolve around definitions of quality at different evaluative levels: participant, institution, external bodies, wider society and not forgetting within group. These tensions involve the refinement of important variables and increasing demands on academic staff. Dahler-Larsen and Krogstrup (2000) also recognised the challenge to professions, where many based evaluation on own reflection over developments within classroom. These issues were clearly evident within this study. But, as was seen in relation to the example from

---

<sup>193</sup> From Berger, Patrick (1964). Some general observations on the problem of work in P Berger (ed). *The human shape of work*. New York: MacMillan Company.

NOR1, there was a sense in which the subunit considered that they wanted and had some degree of influence control over model development and yet this was becoming drawn in a particular direction by the criteria set out within the Quality Reform. So it would seem to be an important point that Dahler-Larsen and Krogstrup made, that the way an evaluation is standardised within an organisation will affect the use of its data and further impact on future model development and application.

### **10.5 Discussion; framing evaluation within the wider organisation**

While there are many demands placed upon the subunits it is considered that the evaluative framework of the organisation and decision processes offer an interesting framework for understanding decisions about evaluation at the micro level. In this final section I attempt to summarise the findings from the data with regard to decision making and related processes. In chapter 6 I discussed the methodological approach of using alternate templates to consider the processes of decision making. Four interlinked decision making models were outlined as parts of the templates: rational, bureaucratic, political and institutional. The purpose of templates was to have a framework against which to consider the responses from the subunit members. These templates were in operation throughout the processes of data analysis and have helped frame the areas of focus for this chapter and are outlined in figure 21.

In chapter 3 definitions of evaluation were outlined, accounting for an activity focused on establishing the value of a particular object or process under investigation. It was further argued that understanding the evaluative process requires greater study into allied decision processes. In this study each of the subunits evaluating their own programmes are required to interpret and respond to different and often competing interpretations of quality from a broad spectrum of stakeholders. Three major areas that define demands for the programme, and as such a gauge of quality, have been outlined from the data: professional judgement (deliverer reflection), participant perception and commissioner interests (which vary across the task environment). Evaluating the programmes both formally and informally in relation to the different demands was increasingly challenging for the subunits, and the design process became more complex with greater interest in effects models. A question for each of the subunits was how did they frame the evaluation processes in their work, that is, how did they decide what to evaluate and what methods and approaches should they use? How would they respond to the demands placed upon them?

In this study the emphasis has been upon the decision making surrounding programme evaluations, but within that spectrum a question arises with regard to the focus has been particularly placed upon the idea of programme quality; how it is perceived and by whom. Data findings have pointed to three main categories, or levels of focus in defining the quality of the programme: these are quality as defined by deliverer reflection, participant perception of programme

quality and commissioner interest in programme impact, mainly seen to be throughput, output and outcome. While these categories are not considered to be mutually exclusive or exhaustive, and are in addition only based upon the reflections of the programme evaluators themselves, they do appear to offer some interesting insight. In the case of the first category, programmes are self-evaluated by the academics that have designed and implemented them; this was commonly considered by them to be the primary, or most important, form of evaluation that they undertook.

Each of the subunits needed to respond to their wider organisational system for quality assurance and this was increasingly understood as the main channel for formal programme evaluation. These systems were generally perceived to be introduced, administered and maintained top down in the organisation. Responses from each of the subunits suggested that these processes were highly bureaucratic and institutionalised within their organisations having been adopted and translated (Røvik, 2007; Sahlin-Andersson, 1996) from macro models introduced following the Bologna process. While the subunits at ENG1, ENG2 and NOR2 responded that they generally considered themselves decoupled from the central processes, apart from fulfilling the data collection requirements and occasionally complaining at the reductive nature and lack of feedback, the subunit at NOR1 had been able to engage more constructively with the wider organisation. NOR1 made an interesting case as the programme group had engaged with and contributed to the development of the wider evaluation framework within the organisation.

Within regard to the decision processes, the subunit at NOR1 stood out as being both tightly coupled in the within group evaluation decision making processes as well as moderately to tightly coupled to the wider organisation. At the same time the group had noted that their internal processes of organising favouring collegial processes were becoming more time consuming and harder to operate, especially as the programme grew. Exhibiting processes of sensemaking in their decision making, the team had become more tightly coupled within a loosely coupled organisation. The collegial model, that they operated is considered to straddle the boundary between political and institutional models. The team was seen to operate in a lobbying, sensegiving fashion within the wider organisation acting for change of evaluation models. This gave them experience when negotiating contracts with commissioning bodies. At the same time they were subject to the changes made within the system.

The two English subunits and the subunit at NOR2 were observed to be more loosely coupled. In agreement with the institutional model template, evaluation within the organisation was typically seen as symbolic or ritual event. The Institutional model recognises the open system, interdependence and influence of the environment on organisational decision making. As Dahler-Larsen (1998: 121) was seen to recognise, the model consists of “loosely coupled system of values, knowledge forms, methods, organisational recipes and routines, where

imitation and the taken as given provides legitimacy, each having its own logic<sup>194</sup>. While the subunit at ENG1 was described as being loosely coupled, those at NOR2 and ENG2 appeared to be decoupled from the central system with regard to engagement over evaluation. Internally each of the groups was more loosely coupled, with little discussion concerning evaluation, which was seen to be the individual academic's responsibility. Again the units at ENG2 and NOR2 accounted for little discussion concerning evaluation, and on this topic were becoming increasingly from one another. These reflections are outlined in table 18 below and applied in figure 21, where the NOR1 subunit is seen as adopting a political lobbying role.

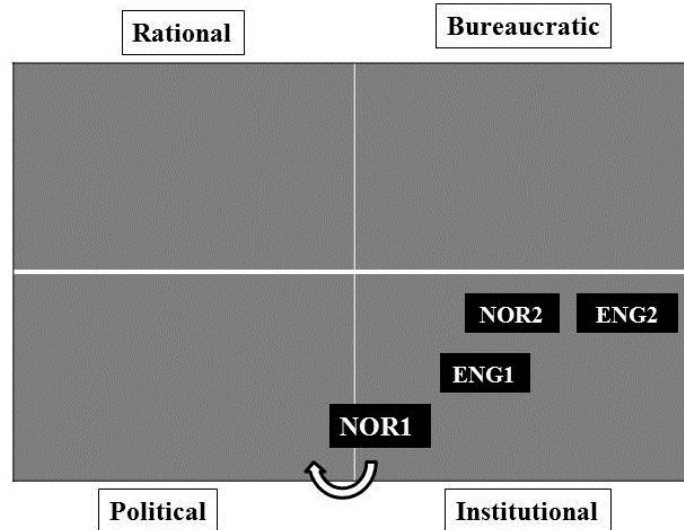
**Table 18: Placement within evaluative framework**

<b>Subunit</b>	<b>Perceived degree of coupling within unit</b>	<b>Perceived degree of coupling to organisation</b>
NOR1	tight	moderate/ tight
ENG1	loose	loose
NOR2 ENG2	loose/decoupled	decoupled

These findings are further presented in relation to decision making models within the wider organisation. The subunits at NOR2 and ENG2 are seen in figure 21 to be placed within the institutional frame but closer to the bureaucratic frame. ENG1 is placed more centrally within the institutional frame. The subunit at NOR1 is placed at the juxtaposition between the Institutional and Political frames, seen to be adopting a lobbying role but while at the same time subject to constitutive evaluation effects (Dahler-Larsen and Krogstrup, 2000).

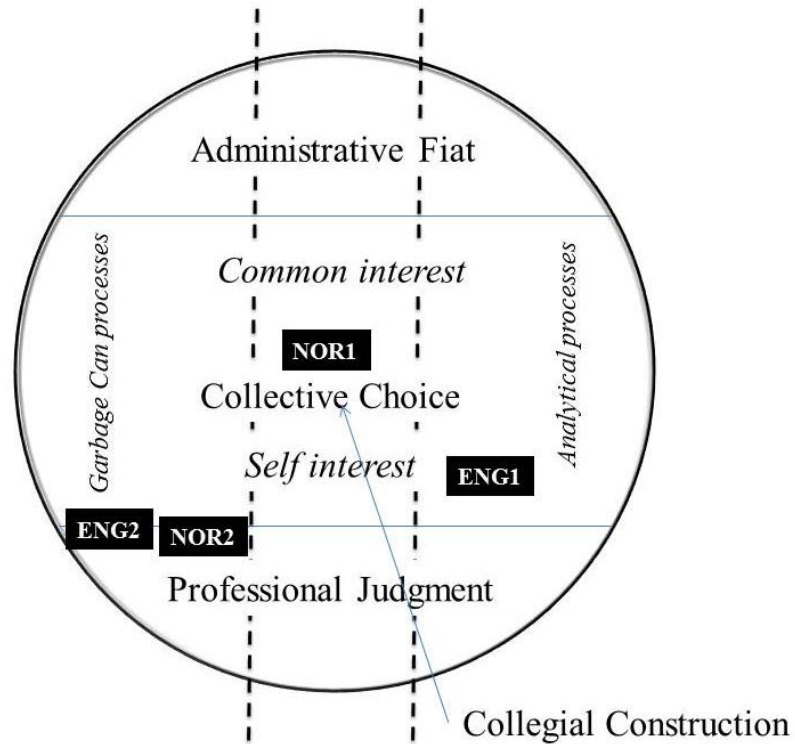
---

<sup>194</sup> My translation from Danish.



**Figure 21: Subunits placed within the template analysis**

A final point regards the perceived impact of these approaches. Once again I utilise Hardy *et al.*'s (1983) interlinked model of decision making. The processes described as collegial construction are considered to spans the divide between administrative fiat and professional judgment, allowing the subunit members to influence the development of the evaluation model within the wider organisation. The dotted lines imply an open channel leading to tighter coupling in that part of the organisation. I do not suggest that the whole HEI becomes more tightly coupled or that all suggestions are adhered to, but merely that the subunit comes closer to the wider decision making, giving opportunity to share their own expertise. Nor do I imply that the subunit drops the idea of professional judgement, but rather that the subunit is able to draw upon the information in the wider organisation. The other units are placed outside of the dotted lines due to their minimal active relationship to the central administration. These reflections are outlined in figure 22.



**Figure 22: Hardy *et al.*'s framework reappraised**

In the next and final chapter I will draw together the proposed contributions from this study as well as considering limitation and areas for future study.



## 11. Concluding remarks, limitations and implications

The purpose of this final chapter is to present an overview of the framework and findings of the study, implications for policy and practice, as well as considering limitations and suggesting areas for further investigation and research.

### 11.1 Discussion

The research problem was drawn from an observation that providers of postgraduate programmes for school leadership face increasing pressures and demands with regard to programme influence on school improvement and development. The research object was thought to be an interesting case for the study of decision making about evaluation due to the complexity of these demands. Investigation of HEI subunits offering master programmes and further and continuing education initiatives in Norway and England revealed the perceived difficulties in isolating, identifying and assessing programme impacts. Additionally, as HEI subunits, the groups faced many other pressures with regard to evaluation of their programmes, not the least from the institutional frameworks for quality assurance. With these demands a question was raised, how do groups respond to demands and pressures to evaluate their programmes? What processes of decision making take place within the group in response to these demands and pressures? As has been seen in previous chapters the respondents drawn from four different subunits taught evaluation theory and practice as part of their programmes, evaluated their own programmes and additionally were experienced as external evaluators.

These reflections framed the overall research question: *what influences the decision of how postgraduate programmes for school leadership are evaluated?* In order to answer this question the following sub-questions were developed:

1. What pressures and demands for evaluation do providers face?
2. What evaluation design frameworks are available to them?
3. What decision processes take place within subunits about the choice of evaluation model?

In order to answer these questions four areas of research were drawn together, two theoretical and two contextual. Literature review focused upon evaluation theory and its developing relationship with research into decisions. It was noted that decision making and related processes have not received enough attention from evaluation theorists, more focused on methodology, participation and utilisation (Christie, 2003; Alkin & Christie, 2005). Decision making is thought to be an integral part of the evaluation process (Dahler-Larsen, 1998). Connections between evaluation and decision theory were outlined, and problem

areas for empirical research were developed from a framework outlined by Stufflebeam et al. (1971). The problem areas were related to the elements of the evaluation process (after Dahler-Larsen, 2004a) under investigation. Decision making models were discussed and applied to add nuance to this area of research. Understanding decision processes is thought to contribute to an improvement of evaluation theory and understanding of processes within units undertaking evaluation.

To understand the context more fully, research into educational leadership and management was applied to illuminate the growing policy demands on programme providers, particularly in recent times under New Public Management (NPM) and Modernisation processes. Focus was placed upon responses from within the field. In relation to practice regarding programmes for school leadership training and development, this study provides interesting data regarding programme providers operating as subunits embedded within higher education institutions. Focus has been placed upon the complexity of internal and external demands with regard to evaluation that lead to tensions for academic and administrative staff and their subunits. Another area of interest regards observations concerning the centralized and decentralized policy structures towards school leadership training and development in Norway and England and their perceived impact upon the evaluation of programmes. The other key contextual area was research on higher education, with specific focus on the demands and pressures associated with the implementation of quality assurance frameworks. Additionally within the study of the impact and application of quality assurance (QA) within higher education data is provided here at the often understated micro-level with regard to how QA systems are experienced and operationalized by practitioners. Investigation has taken place of the perceived internal interaction as well as with the wider organisation and others from the task environment. The findings also show that continued research into the *impact* of higher education programmes is necessary.

The aim of this study has therefore been to understand what decision processes take place with regard to the evaluation of programmes, focus is placed upon decisions *about* evaluation rather than decisions emanating *from* evaluation. In this way, investigation has taken place into factors that influence decision making as well as the way decision making processes influence the evaluation process and implementation of demands. At the same time it is noted that these processes can be thought to have a recursive effect on future decision making such that experiences from one evaluation will have effects on subsequent processes and initiatives. In attempting to answer the research questions outlined above focus has been placed on the perception of subunit members, looking beyond documentation and organisational scripts in order to understand their perception of practice.

## 11.2 Summary of findings

In this section I will briefly draw together the main findings from the study. In section 5.1 I outlined research on decision making about evaluations. Focus was placed upon the work of Stufflebeam and colleagues (1971) identifying categories of problem related to evaluation decision making. Out of the 5 problematic areas identified by the authors<sup>195</sup>, a new framework was adapted with focus placed upon definition, demands, designs, decision makers, and decision making. These categories are thought to nuance Stufflebeam et al.'s framework for the area under study rather than replace them. As was noted in relation to demands, the study was delimited to focus upon the perceptions of programme providers rather than accounting for, as the original authors had suggested, the wider values of all stakeholders in the process. This framework for investigation was outlined in table 8 where the problematic areas were compared to an interpretation of the authors' original framework. In the table below I have adjusted their order to match the pattern of the data presentation in this thesis. It is these categories that I will use in summing up.

**Table 19: The problem areas of evaluation (reapplying Stufflebeam et al., 1971)**

Definition	How do subunit members understand the concept of evaluation and how does this influence the process?
Demands	What demands are placed upon the subunit and how are they interpreted?
Designs	What is the purpose of the design and what degree of agreement is there about models?
Decision makers	Who is involved in the decision making concerning evaluations?
Decision making	What responses are there to the demand and pressures for evaluation and what can application of mixed models of decision making tell us about evaluation processes?

### *Definition*

The purpose of investigating the first problematic area was to gain an understanding of how subunit members defined evaluation and how evaluation

---

<sup>195</sup> Evaluation definition, decision making, values and criteria, administrative levels, the research model

was understood across the subunit, considering this underlying understanding to be a reference point for future decision making. Drawing on the findings from the literature review, a definition of evaluation was presented, noting it to be a decision-focused activity to assess the merit/value/worth of something based on collected data. Within this generalised definition there was recognised to be great diversity in emphasis. The aim was therefore not to produce a definitive definition of evaluation but rather to recognise that the definition or perception of evaluation by evaluators can be thought to frame responses to demands placed on them. In section 3.2 it was noted that Christie's research (2003) had found little influence of evaluation theory on practice, and few evaluators indicated particular theories when outlining their work. There was little evidence of discussion about particular theory, but some respondents referred to the literature they had chosen for their programmes.

Additionally, Dahler-Larsen (1998) and Hansen (2005b) had noted how the nature of evaluation had changed within an "age of accountability". During a period where policy has increasingly been shaped by NPM and Modernisation, it was seen that Norris and Kushner had highlighted how evaluative activity has become routinized and institutionalised characterised by "internal evaluation and external auditing, inspection and monitoring arrangements and performance management systems" where the results are published (2007: 6). These developments were combined with the decline of professional autonomy, decentralisation of responsibility and demand for control of efficiency and effectiveness (2007: 7) and decisions about definitions of quality and performance standards are taken at the executive level (2007: 12). The authors were seen to agree with Dahler-Larsen that these decisions are constitutive (Dahler-Larsen, 2007) setting standards for organisations, and challenging professional autonomy (ibid). These developments were evident particularly in the English cases but more indirectly in Norway.

In this study the academic respondents could all be considered similar to Tourmen's (2009) category of experienced practitioners, having operated as both internal and external evaluators. But, as was noted in Christie's research there was generally little focus in the responses on espoused evaluation theories. The framework for evaluation had in many ways become of more critical interest for respondents as they felt that it was this that was driving the content and focus of evaluations, giving less freedom for interpretation. In that respect, respondents moved on more rapidly to speak of the different demands for evaluation rather than underlying ethos of the processes. Therefore, as was seen in chapter 7 and 8 respondents in this study spoke more of the difficulty of balancing differing perceptions of evaluation within the frameworks, feeling that the concept of evaluation had been reduced to a tool of accountability. Within the groups there was a consideration that organisational evaluations were now focused on "systems and structures" rather than balancing accountability and improvement. In chapter 5 it was noted how Stufflebeam et al. outlined 3 three broad types of definition: measurement, congruence and professional judgement

(1971: 9ff). These definitions were applied to respondents' reactions. The respondents can be described as seeing their own work based on professional judgement, whilst the quality assurance frameworks developing in their organisations were seen to be focused on what might be described as *internal congruence*, seeking data upon inputs, throughputs and outputs as a measure of programme quality. In England, and to some lesser extent in Norway through the involvement of different commissioners, frameworks are increasingly becoming more focused upon what could be described as *external congruence*, with a desire to link programme activity with impact upon practice and improved results within schooling. This more *utilitarian* framework (House, 1978) contrasts with the *pluralist* approach of the academics within the subgroups. One difference between the groups was that the subunit NOR1 appeared more focused upon *collective professional judgement* than were the 3 other subunits. The overall responses match Dahler-Larsen's reflections that evaluation is becoming redefined as an *audit approach* (2006b), characterised by an increase in result-oriented measurement and monitoring (2005b). The area concerning definition of evaluation was seen to overlap in the responses with the following section focused on demands.

### *Demands*

The second problematic area for investigation was that of demands placed upon the subunits with regard to evaluation. The literature review revealed growing pressures for evidence of impact and quality assurance, and these were seen to be part of the pervasive "evaluation wave" increasingly evident across Western Europe and wider afield and part of a growing "evaluation culture" (Dahler-Larsen, 2006b) tied particularly to NPM and thereafter Modernisation. These pervasive demands filter through to the micro-level under focus in this study. While Dahler-Larsen questions whether quality assurance is a form of evaluation, he also recognised that it is more commonly becoming synonymous with evaluation (2005b).

Responses from the subunits revealed a complex and often conflicting web of demands, where they found themselves at intersection point. The intersection of demands for evaluation is related to the different pressures within the field of study and institutional context that the subunits find themselves within. Both countries have witnessed a significant focus on improving competence of school leaders, where more generally this was seen to be the key to organisational development and subsequent improved pupil outcomes. As providers the subunits were competing within a wider context to develop programmes where there is an expectation that groups can demonstrate impact. The English subunits were subject to a centralising policy framework with regard to school leadership development and training, whilst the Norwegian subunits operate within a more fragmented, decentralised system. Demands resulting from national policy in England had seen increasing pressure from funding agencies for demonstration of impact. Although offering different types of programmes to the HEI subunits,

the frameworks for evaluation developed and implemented by the NCSL were also considered to have impacted on the evaluation perceptions within ENG1 and ENG2. Significantly the respondents had experience of evaluating and responding to tenders for evaluation from the NCSL, and where they again needed to address the issue of impact. Despite finding methodological problems with complying with these demands, there appeared to be evidence of constitutive effects (Dahler-Larsen and Krogstrup, 2000) as the demands were becoming more institutionalised within evaluation processes and the members needing to respond. At the same time, paradoxically, respondents noted that they didn't evaluate their own programmes to anything near the degree they did when operating as external evaluators on other programmes. The decentralised system in Norway meant that similar processes were perceived to be taking place much more slowly, but the policy framework had led to the developing involvement of commissioning bodies in tendering for programmes. Demands for evaluation varied greatly from mandator to mandator, dependent upon what was noted to be *commissioner competency*. This did however present the subunits with greater opportunity to influence evaluative frameworks towards their chosen position, as well as paradoxically giving them greater opportunity to investigate programme impact across the local level. This led to a very different type of process of decision making for the subunits than within their own organisations under quality assurance, but as in the case of NOR1 more room to manoeuvre in both contexts.

At the same time the subunits in this study were placed within higher education institutions meaning they were subject to quality assurance frameworks. The subunits were deliverers of master programmes and other forms of further and continuing education. Although there was not perfect parity between the programme frameworks their contexts were similar. The 4 subunits chosen (2 in Norway and 2 in England) were all subject under the follow up to the Bologna process to the implementation of quality assurance frameworks within their wider organisations. How the different HEIs apply such frameworks was seen to be of prime importance. As was seen above in reaction to perceptions of evaluation definition, the demands related to the introduction and nature of implementation of quality assurance frameworks which were often felt to conflict with programme values, exemplified by the recognition that throughput and participant satisfaction appeared to be the most central measures of quality. In the latter case, respondents from across the subunits reflected over the increasing development of participants from student to consumer.

Another area of interest that developed further through the research process is the importance of within-unit demands. These will be discussed below in relation to decision responses, but at this point it is important to note that there was discussion in each of the subunits about a common internal pressure to use evaluation to improve the quality of provision, but there was a notable variation of collective effort to achieve this goal. In the data from the subunits it was noted that definition of demands for evaluation was often generally considered

to be established outside of the subunit. However the underlying definitions and perception of the process within the group was seen as a filter to interpret pressures and demands faced. All of the subunits reacted negatively to major parts of the implementation and follow up of quality assurance frameworks in their institution. In addition this was seen to be time consuming, adding greater demands for information but paradoxically with less responsive feedback. Institutional demands were felt to become increasingly more bureaucratic.

There are many competing and complex demands upon programme providers with regard to evaluation of their programmes, and these have continued to increase over recent years. Despite the introduction of new quality assurance systems, which admittedly are under development and require further longitudinal investigation, the increase in information available in the programmes appeared loosely coupled to programme group assessment of provision and future development and subsequent programme change. This does not mean that there was a lack of evaluation at programme level, but rather that the systems did not appear to support or be supported by this flow of information. This would appear to support Hellstern's comment that "the success in institutionalization is not matched by an equal success in utilization" (1986: 279). How the decision making processes within the various subunits framed their responses to demands for evaluation, was seen to relate to the designs considered available. This is dealt with in the next problematic area.

### *Designs*

The next problematic area related to design, focused upon purpose of design and degree of agreement about models. As has already been pointed out, the literature review in chapters 2 to 4 revealed the multiple pressures for evaluation on higher education subunits in this field of study, and the various modes of evaluation design that had developed from these. When asked to consider evaluation the respondents were generally focused upon the quality assurance frameworks and their designs. While the subunits in England were increasingly expected to design and implement models that could offer impact data to external funding bodies, the quality assurance systems were centralised within the wider organisation of the HEIs. The process was similar in Norway. How the subunits responded to these demands and considered these designs will be dealt with in subsequent sections.

Descriptions of the designs can be found in chapter 8. In this section, I briefly refer to issues raised associated with designs as were seen in section 10.2. Once again a paradox was noticed by respondents. On the one hand they expressed there to be more data collected than ever before, from a greater number of stakeholders. And yet the models were considered able to reveal little more than how much programme participants thought they had achieved. Data for the organisation was considered to be focused on controlling throughput of students, as an accountability tool, but there was little or no feedback. As a design, the

QA systems did not appear to provide a framework for cross-organisation learning. The respondents generally could not see what this added to the academic self-reflection over programme delivery given that there was no response. The system, it seemed, was only geared for dealing with failing areas and these subunits had not yet reached such levels. This was further highlighted by the relative success of the programmes. As long as they did well there was no intervention. The designs did not appear to provide feedback on issues raised by the subunit members either, which reinforced their perception that processes were about control and not improvement. The lack of improvement focus in evaluation was important on two fronts. The subunit members did not reject the institutional frameworks because they saw evaluation as a purely administrative task, but rather because they saw that for them it had become a merely administrative task. There was little to no feedback from higher levels in the organisation, and therefore respondents from across the groups found it hard to interpret the activity as anything more than a task that is accountability focused with little room for micro-improvement. There was therefore a mismatch between definitions of intent for evaluation within and across institutional boundaries. This was a key problem area raised by Stufflebeam et al. (1971). The generic nature of the forms and the lack of student participation revealed conceptual difficulties with interpretation. Designs are therefore also seen to encompass the frameworks they are set in.

Responses with regard to designs for programme effects were also interesting. There was little disagreement in the importance of attempting to understand the influence and outcomes of programmes, but there was a strong reflection that current models were unsuitable for the growing demands to ascertain effects and impact on pupil outcomes. The data was noted to reflect Stake's categories, as seen in section 5.2, of *activist* or *determinist* perceptions of the relationship between impact and context, where policy makers in England generally espouse the former, believing programmes to impact context to a large degree, and the academics in this study generally the latter, considering context as dominating (Stake, 1990). Members of NOR2 and some respondents in ENG1 were, however, more open to exploring the possibility of developing designs that might give greater information on impact upon context and outcomes, but this was seen to require a research led process rather than merely be the base of evaluation focus. In section 8.8.2 it was noted that respondents felt that five key problematic areas still needed to be dealt with in this area, isolating cause and effect, agreeing the level of observation, allowing for the complexity of programme purpose, developing a time frame that was sufficiently longitudinal and dealing with endemic organisational constraints of higher education programmes.

The focus upon designs in use, then, was seen to be important for a number of reasons, it offered data regarding how the introduction of quality assurance has impacted evaluative behaviour, it helped frame how the self-perception of



expertise within the subunit related to perception in the wider organisation, and finally it was seen to be related to evaluation use within organisations.

#### *Decision makers - roles*

The fourth and fifth problematic areas are focused more specifically on the decision making processes. Within these processes, the fourth deals with the different roles played by decision makers, while the fifth deals with decision responses to demands. In order to consider these areas a model using and applying Dahler-Larsen's (2004a) elements of the evaluation process was found to be helpful. These elements are not seen as linear, rather as cyclical, recursive and complex. As this study has focused on decisions leading up to the implementation of the data collection, the first four elements were considered pertinent to this study: initiation, agenda, knowledge management and organization, and design. As I outlined earlier, evaluation researchers have often focused on the latter parts related to implementation and stakeholder participation or assessed the quality of model design and methodology. It was therefore intended to investigate something more about within unit responses, particularly given the complexity and diversity of demands. I return to these elements further below after first considering those involved in the decision processes.

An important question with regard to roles is who makes the decisions? As was outlined in section 5.4, Hardy et al.'s (1983) framework of interlinked decision making within higher education was applied to this study. The model was seen as useful as it builds upon on Mintzberg's (1979) notion of the professional bureaucracy, characterised as complex and loosely coupled. The three interlinked levels of decision making, recognises decisions to be based on professional judgement, administrative fiat, or collective choice. Interestingly this latter area was further categorised between different decision making models<sup>196</sup>, though these are weakened by the lack of institutional theory or sensemaking approaches. This updated model was used to interpret the processes described by the respondents, outlined below.

The role of the academic within the field of study of education leadership was seen to be important in relation to evaluative decision making. Respondents considered that working as an academic in higher education provided the skills, experience and competence to evaluate an academic programme, but additionally that the particular academic area that they were trained and worked in also provided further competence. This was highlighted by one respondent in NOR1 in an administrative position who considered the importance of having undergone the same kind of academic socialisation through qualification as the academic staff on the programme. The academic staff in the subunit had also

---

<sup>196</sup> Collegial, political, garbage can and rational models (Hardy et al. 1983: 412). The models are also linked to the combined frameworks as outlined by Allison (1999) and Peterson (1976)

commented on the importance of this background. Respondents across the subunits saw themselves as experienced evaluators (Tourmen, 2009), but unlike the findings in Tourmen's study their experience was not always considered to be applied creatively in decision processes about evaluation within institutional frameworks. Tourmen's findings did, however, apply to roles of external evaluation, rather than the internal perspectives in this study.

As was seen in section 5.7.1, it was important to investigate the role of occupational communities. This concept drawn from Van Maanen and Barley (1984) and applied to the evaluation field by Dahler-Larsen (1998) was important to understanding the within-unit roles and their relationship to the wider organisation as well as opinion of organisational evaluation frameworks. Dahler-Larsen's typology was adopted as a useful model for understanding the responses of the subunits to the demands placed upon them to evaluate their programmes, where the dimensions were perception of relative autonomy and degree of agreement with evaluation criteria.

#### *Decision making processes – Responses to demands*

Investigation was concentrated upon attempting to understand the collective decision processes within the subunits. Respondents discussed the degree of internal discussion over evaluation frameworks and relationship to the wider organisation. Focus was also placed on understanding how academic groups, as occupational communities, validated particular conceptualisations of evaluation and related these to the frameworks around them. While the wider organisations can be described as loosely coupled, the perception of within unit coupling with regard to evaluation was seen to be an important factor in decision processes and responses. There were noted to be two main areas of interest within the data gathered from respondents. The first area relates to the impact of the subunits upon the evaluation and decision frameworks of the wider organisation, the second area deals with the impact of evaluation frameworks upon the subunits and their evaluation and decision behaviour.

The issue of impact on the wider organisation was seen to be related to the strength of internal coupling and degree of discussion about evaluation. The subunits were seen to be relatively heteronymous with regard to evaluative demands and disagreed with institutional evaluative frameworks. There was also a general disagreement with external impact-oriented models. The findings revealed two different decision processes and responses by the subunits. As was seen in section 3.6.3 evaluators expect that those receiving their data will utilise it for the best purposes possible, likewise data can often be over-interpreted by those running programmes under investigation (Weiss, 1998a). With self-evaluation and quality assurance models applied at the micro-level there appears to be an increasing disharmony between mandators and those performing and reporting programmes. Despite as was reiterated earlier, Harvey and Newton (2007; Newton, 2000) suggesting that there need not be a tension between

accountability and improvement factors within the same evaluation framework, the authors did note tensions at the different operational levels. The tensions were outlined in figures 9 and 10, recognising the pressures of different foci between the external and internal. When key actors are, or perceive themselves to be, excluded from processes, or their findings are not considered to be used, the participants in the process become increasingly disillusioned. In this case the subunit members described how the lack of feedback further alienated those at the lower levels, highlighting their impression that evaluation only serves a control function.

Secondly, three groups in this study described themselves as increasingly decoupled from the central processes, with no apparent cross-over to their own self-evaluation. The fourth subunit, NOR1, attempted to re-couple with the central system based on collective sensemaking of evaluation processes under the new quality assurance system. In doing so the unit appeared to have worked towards re-establishing their role in collegial governance of the programmes. Evaluation was one area of considered importance to the subunit members, an area, or *sensemaking gap* (Maitlis and Lawrence, 2007) where the central systems were perceived to have become more reductive in nature and based on accountability rather than improvement. While the other groups also noted this ‘sensemaking gap’ there was little or no reported collective action for change in response. Additionally it was considered that the leadership had developed systems without feedback processes built in, reliant on external models. This led at NOR1 to a process similar to what Dutton and colleagues (2001; 2002) have referred to as *issue selling*, a process by which subordinate units in a hierarchy can attempt to gain the attention of those higher up in the system and presenting areas for decision making. The process by which the issue selling came about is interesting, as the group developed a response collegially, based upon their prior experience and professional convictions. These *triggered* (Maitlis & Lawrence, 2007) a process similar to sensegiving, *enabled* by their self-perceptions of professional experience and their relative “success” as providers of academic programmes and implementation of evaluations. These processes were described as *collegial construction*, characterised by a high degree of internal communication and cooperation, building from group homogeneity, based on evaluation experience pre-QA frameworks, and with willingness to seek to share experiences and influence the development of the formal QA system. The other subunits exhibited looser coupling, particularly with regard to evaluation. The responses were described as *dismissive submission*; disagreeing with the form of the demands and frameworks of evaluation placed upon them, but required to implement them. This bore similarities to the Dahler-Larsen (1998) application of Berger’s (1964) concept of “role distance”.

The collegial action appears to have mediated their discussion capacity and process facilitators. As a result, the subunit at NOR1 was perceived to be contributing at the *agenda* element of the evaluative decision process, where ENG1 was perceived to be only influencing at the stage of the *operationalizing*

of the evaluation and in essence the design was already in place. In the subunits of NOR2 and ENG2 the groups considered themselves completely decoupled from the central processes. These findings require further investigation.

As outlined above, the second area deals with the impact of evaluation frameworks upon the subunits and their evaluation and decision behaviour. This is related to the impact of *constitutive effects* of evaluation as discussed by Dahler-Larsen and Krogstrup (2000), outlined in section 5.5.4. These effects concern evaluation functioning as an “institutionalised phenomenon”<sup>197</sup> that “co-constructs the social reality surrounding the evaluation” where it was noted that effects can be seen within “three aspects of social reality”: the *material/content*, the *time/timing* related, and in social relations and identities (ibid: 295). The findings from this study were discussed in section 10.4.2, where it was noted that respondents were aware of pressures and demands linked to evaluation and quality assurance, including the pressure for systematisation, the pressure for succession (completion), the pressure for satisfaction and increasingly, the pressure for subsequent impact.

In agreement with Dahler-Larsen and Krogstrup, there was a perception of roles being redefined by such processes; and the sense of reduction of professional autonomy. Another area of agreement with Dahler-Larsen (1998, 2007) was that QA as a constitutive influence becomes the formal but not necessarily active image of evaluation, impacting the evaluation framework over time. There was a suggestion in all of the subunits that the demands for reporting outlined within the organisation appeared to be directing focus towards alternative ways of evaluating, and the adoption of processes that appeared contradictory to the value structures within the groups. Challenges were seen in changes in material focus of the evaluations, the time required for evaluation and the timing at which it was implemented and related to social relations and identity as groups were required to respond to the demands of stakeholders, challenging their professional autonomy further. Even though the subunit within NOR1 appeared to be collectively responding to the demands placed upon them, the frameworks were consistently directly their attention to specific ways of evaluation, and much time was spent upon responding to these frameworks. This does not mean that there was a total disagreement with the framework, but rather shows the framework’s influence on the subunit. These processes are considered to be recursive.

In relation to the discussion raised in ENG1 about complexity of impact data, there are a range of challenges here. Subunits must respond to requirements for greater understanding, contributing to better structural models for evaluation, combined with the shift from professional judgement to emphasis on measureable results, challenged in return by the perceived methodological complexity of doing so. One respondent was shown to recognise that even if

---

<sup>197</sup> My translation from Danish.

there is greater systematisation of data collection across the subunit, there still remained the problem of providing the “hard data” now required. This raises again the issue of discrepancy for actors between talk, decisions and actions (Brunsson, 2002). There appeared also to be a degree of tension between the structures of the internal evaluation system, localised understandings of the evaluation process with the academic group and pressure from the external environment in terms of resource dependency and expectations applied to that. This appears to fit the difficulties in equilibrating talk, decisions and action that Brunsson (2002) outlines with regard to balancing politics and action within a tension caused by embedded organisational hypocrisy. As Brunsson noted, the “double basis of legitimacy” requires action, based on developing integrative structures, and politics which focuses upon encouraging dissolution (2002: 33). This issue is “insoluble” and can only be handled, by decoupling politics and action.

#### *Concluding remarks*

The research in this study is therefore considered to be centred upon the fourth quadrant of Dahler-Larsen’s (1998) typology as outlined in tables 6 and 7. The focus from this research led to a reapplication of the fourth quadrant, where the quadrant is reapplied within a continuum, rooming the responses described as *collegial construction* and *dismissive submission*. All of the groups were considered to be heteronymous with regard to the demands for evaluation and it was a necessity to follow institutional frameworks for QA. The three subunits appeared to experience role distance in what was described as *dismissive submission*, whereas the fourth subunit NOR1 had implemented collective sensemaking and sensegiving processes in order to respond to the demands, being attributed here as *collegial construction*. They continued to be characterised as heteronymous and in disagreement with the main focus and utilisation purposes of the QA frameworks. However, the sensemaking response appeared to enable them to constructively engage in attempting to change these processes. These reflections were related to the different decision process models by applying alternate templates identified in chapter 5. As was seen in section 10.4.3 these processes were linked to the perceived degree of coupling within the subunit and between the subunit and wider organisation. The subunit at NOR1 was perceived to show tight internal coupling and had chosen to couple more tightly to the wider organisation in order to influence for framework change. The remaining subunits showed loose internal coupling and appeared to be oscillating between loose and decoupled relationship to the wider organisation. As was seen in figures 18 and 19, NOR2 and ENG1 and ENG2 were characterised by institutional decision processes in relation to evaluation, comprised of ritual decisions and subject to constitutive effects from the wider frameworks, which were perceived to affect their legitimacy and authority within their organisations. While there also appeared to be evidence of this in NOR1, the group was also perceived to exhibit signs of ideological bargaining, identified with the political model, where the group defended their value system

and offered alternatives to the organisational frameworks. The impact of this collegial construction was seen in relation to Hardy et al.'s (1983) model, where it was considered to span the divide between administrative fiat and professional judgment, allowing the subunit members to influence the development of the evaluation model within the organisation and maintaining the importance of their professional judgement. The other units saw a greater separation between administrative fiat and professional judgement, the latter of which additionally appearing to be weakened at the micro-level.

The responses from the subunits show the evaluative decision making processes to be embedded in a web of complex demands. Data at the subunit micro level has revealed the internal processes as well as their perceived relationships with the wider organisations and the task environment. The data shows that quality assurance systems demand greater energy and time from the subunits and the frameworks narrowed the definition of quality to throughput, satisfaction and potentially impact. Additionally the subunits receive little, if any feedback through the system, confirming their perception that the system was merely accountability based rather than focused on improvement. Within heteronomous groups disagreeing with the values of the evaluation model the possibilities have often been seen to be decoupling, ritualistic behaviour or adaptation to downloaded models. Three subunits in this study were perceived to adopt the former response, relying only on their own professional judgement for micro decision making and programme development. These groups were all characterised by weak internal coupling with regard to evaluation decision making. This was generally not seen as problematic while the programmes were considered successful. A fourth subunit, however, was much more tightly coupled and while remaining under the demands of system engaged in sensemaking and sensegiving processes in order to improve competence as well as influence the evaluative frameworks in the organisation. At the same time the direction of the evaluation system constitutively influenced the focus of evaluations as processes and models become institutionalised. These findings build particularly on the work of Dahler-Larsen (1998, 2004a; & Krogstrup, 2000). The decision process models enabled understanding of this nuance between the subunits. This is not to suggest that this is the only variable of importance, and as will be seen in the next section the units of analysis have not been claimed to be perfectly comparable. Applying decision process models is however thought to offer nuance to evaluative theory which also has implications for policy and practice. These themes will be dealt with in the next three subsections.

Finally, I summarise in table 20 below the findings from the three main areas of investigation, indicating the array of demands, types of evaluation model and design utilised and the decision processes put in place.

Table 20: Summary of main research areas

Area related to evaluation	Subunit	NOR1	NOR2	ENG1	ENG2
	Dimensions				
Demands	Perceived institutional pressure	variable	strong	strong	strong
	Perceived commissioner pressure	variable	strong	strong	strong
	Perceived policy pressure	weak	weak	strong	strong
	Perceived participant pressure	variable	variable	variable	variable
	Strength of within unit pressure	strong	weak	moderate	moderate/weak
Designs	Degree of agreement about programme goals	strong	moderate	moderate	moderate
	Degree of agreement about measurement	strong	moderate	moderate/weak	moderate/weak
	Degree of control over formal design elements	<i>moderate</i>	none	weak	weak
	Perceived limitations with formal designs	strong	strong	strong	strong
Decisions	Perception of academic identity	strong	strong	strong	strong
	Perception of internal subunit process	collaborative	individual	collaborative/ individual	individual
	Degree of cohesion	strong	weak	moderate	weak
	Group coupling to institutional framework	(moderate)	decoupled	loose	decoupled

### **11.3 Limitations of the study**

There are limitations to be discussed with regard to the study, some of which were already outlined in section 6.6. The framework for the study is complex, drawing together a number of fields of study. While this was considered to be an important step with regard to the development of evaluation theory, it also creates difficulty when attempting to combine such a broad spectrum of ideas. There is also the suggestion that important variables might be missing if the decision processes are overemphasised. Even though the subunits only exhibited the characteristics of two of the decision models it was considered important to apply all the templates to the processes. With four subunits under study it was not possible to present all the data deliberations of the alternate templates. In future research focussing on one subunit within its wider organisation would be helpful.

This is also related to the fact that the fieldwork is based upon the self-report of the subunit members. Despite using processes of member validation and checking of data it is also recognised that it would be helpful in future study to combine individual interviews with focus groups and longitudinal observation of meetings and the implementation of evaluations, following specific examples through the organisational system. This could also be combined with studies drawing on different levels of the organisation. The views of students and leaders in the organisation would also be important. However, the purpose here was to explore the processes at the level of those responsible for programme development and implementation in order to see how they focused on evaluating their own initiatives.

In terms of the practical implementation of the fieldwork there were a number of limitations. The issue was experienced on a continuum from uninteresting to extremely sensitive. This caused some difficulties for the initial sampling but also within the subunits under study there were some who declined to take part and others who did not respond. This meant that there were fewer respondents in some subunits than originally anticipated. In addition to individual anonymity, which was assured before the research process began, there were respondents who asked for assured institutional anonymity based on the issue of sensitivity within an increasingly competitive and accountability focused context. As a result I have not included specific reference to document review and presentation of data that provided the background to each interview, as well as reflection over the responses given.

### **11.4 Further implications for theory, policy and practice**

The purpose of this study was to explore the decision processes taking place within HEI subunits operating as providers of postgraduate programmes in school leadership. The implications for evaluation theory have been outlined. In particular an application (and reapplication) of problematic areas for evaluation (Stufflebeam et al., 1971) was considered important within an era of greater



accountability and renewed focus upon outcomes and impact of programmes. This also involved exploration of the evaluation process by considering the elements of evaluation as outlined by Dahler-Larsen (2004a). Research was focused on the elements leading up to evaluation implementation, bringing nuance to the understanding of processes of response to pressures and demands for evaluation. Drawing on combined models of decision process theory has allowed further investigation of micro-processes within organisations, in this case subunits within loosely coupled professional bureaucracies, namely higher education institutions.

The introduction of quality assurance systems within higher education was perceived to have both increased the amount of evaluative activity whilst decreasing the amount of evaluation. Centralisation of processes led to increasingly more formalised system for reporting of information gathered but with less feedback for actors at the lower level. This also has wider implications for decision making more generally within higher education. This research has also focused upon the response to quality assurance demands at micro level, as was seen in section 4.1, an area considered to require greater attention. For example, Amaral et al. (2002) noted that *collegial governance* at lower levels had decreased with centralisation at the institutional level aimed at increasing cohesion across the organisation. But as was related at NOR1, a subunit with strong internal coupling sought to influence the wider organisation through processes of sensemaking and sensegiving for a change of evaluation model that would reflect the subunit values. As a result this group entered and influenced the agenda phase of evaluation, gaining voice and a position by which different channels could be used to “sell” their “issue”. These findings develop on the work of Dutton et al. (2001, 2002) as well as Maitlis and Lawrence’s (2007) work on triggers and enablers of sensemaking and sensegiving, where collegial processes are tentatively considered to moderate the enablers. This requires further investigation.

There are particular implications for the wider organisations within which the subunits are embedded. In particular the subunits found the lack of feedback to be problematic. Lack of feedback was not always perceived to signal lack of support from those higher up in the system, but quality assurance was described as heavily bureaucratic and a necessary ritual. This increased the perceived mismatch between time available to the academics and time thought usefully spent. Respondents across the subunits were in relative agreement that focus had now shifted to be placed upon the satisfaction of students and their throughput. These frameworks appeared to have constitutive effects that quality indicators were being redefined. This leads to a twofold tension as different policies collide. With regard to quality assurance, it is unclear whether these processes lead to any further reflection over the quality of academic programmes, in a system perceived to be characterised by balancing between pleasing the student as a customer per se, and pleasing the student because the course was considered relevant. The problem arose that relevance could easily be equated with

excitement, even though not all students will find all things exciting. There was openness to increased voice for students, but across the subunits, however, there was a perceived problem with how data were interpreted at higher levels.

Programme group members were generally sceptical to the type of data being provided and the conclusions that could be drawn from it. They were also unaware of any utilization of this data and unaccustomed to receive any feedback within the organisational hierarchy, based upon this data. It is of course reiterated that the programmes under study were all considered to be successful, and generally unlikely to draw any negative feedback upon the criteria highlighted within the organisational evaluation models. In that way the individual attitudes to evaluation and the decision processes at programme group level became of increasing interest, as did the input over what should go into the models. It was only NOR1 that appeared to have a proactive input into the models set out. In ENG 1 and 2 it was possible to add programme specific questions to internal models, but neither followed this up consistently. NOR 2 had none. There remains an interesting question as to whether increasing feedback might resolve many of the issues. As was seen above, these complications of the context offer the backdrop to the decision making that takes place concerning evaluation. But in this case it is interesting that much has been left to the individuals to develop from their own armoury and there has been great lack of clarity over what evaluation should be about, i.e. the content; whilst the form and rationale has been set - from a QA point of view. This raises another question, has QA hindered organisations from evaluating in other ways that might reveal the quality of courses?

With regard to the focus of these programmes, school leadership development, interest is raised at different levels with regard to their impact. Although, as the subunit members recognised, at the time of data collection seemingly little to no evaluation or research for impact it is hard to ascertain what impact they might be having at school level. Thus questions are still raised as to how programme providers know that their programmes are correctly focused. Research and further discussion into these areas is required, even if there might always be an ontological, epistemological and methodological divide over any conclusions. An interesting part of this study related to programmes provided for commissioners. From an evaluation perspective this reduces the number of stakeholders to one collective that is instead of attempting to please the students, pleasing their collective employer offers a different challenge. While the involvement of commissioners gave the subunits greater opportunity to see their programmes in action at school level, there was also a problem with *commissioner competency*. These “super-stakeholders” were seen to have varied goals and purposes for buying in a service, often beyond the need to develop for further education programmes and academic improvement.

### **11.5 Further considerations for future research**

In addition to the topics raised in the subsections above there are four topics that would be particularly interesting if following up the findings of this study:

Firstly, looking to explore and further develop the decision making framework, investigating crossover with theories of learning in organisations and organisational learning.

Secondly, longitudinal investigation of evaluative decision making processes within and across one organisation, a higher education institution, related to internal and external demands and pressures for evaluation. Such a study could then be replicated in other HEIs and compared with other organisational types.

Thirdly, further research would also be helpful into investigating constitutive effects of evaluations upon higher education programmes. This should follow up indications in this current study of how programmes appear to be reshaped and will further enlighten some complexities of the decision process. This links to Dahler-Larsen's "social constructivist view" of evaluation, that adopting indicators will not automatically lead to deterministically positive or negative affects; effects are therefore "complex and depend on interpretations, relations and contexts"; the strength of these effects being differential to context (Dahler-Larsen, 2007: 30). There are indications in the data from this study to support this idea that context will play an important role in the strength of constitutive effects, but it would also, again, be important to consider different subunits behaviour within the same institutions.

Fourthly, as was outlined in the subsection above, to explore further the relationship between collegial decision making and enablers of sensegiving related to evaluation and compare this to other decision making activity within the same institution. This would also involve studying more closely the importance of the occupational community and looking at other decision processes within groups. The theoretical focus in this study was upon decisions about evaluation and it would be interesting to apply this topic to other decisions made in the subunit.

## Appendix

**Table 21: Types of Evaluation use (Weiss, 1998b; Weiss, et al., 2005)**

Type of Use	Conditions prevalent under
Instrumental	Non-controversial evaluation findings Small scale changes recommended Within a stable environment Program in 'danger', where no way out is recognisable.
Conceptual	Learning at the local level by participating evaluators, local insight gained and applied.
Legitimizing	Where evaluation to be used as 'instrument of persuasion', to make changes to weaknesses already recognised.
Enlightenment <sup>198</sup>	Influence upon external programmes, e.g. via meta-analysis
Influence	Developed from Kirkhart's integration of 'process use' into the other categories
Imposition	Programmes having 'successful' evaluations become funded and downloaded by the mandator to other providers or receivers.

**Table 22: Meta-analyses of the factors affecting evaluation utilization (after Hofstetter and Alkin, 2003)**

Authors	Empirical focus	Résumé of findings
Patton <i>et al.</i> (1977)	Interviews with federal mental health evaluators and programme personnel.	Two primary factors seen as having greatest influence on utilization – 'political consideration' and 'the personal factor'.
Alkin, Dalliak, and White (1979)	Case studies in local field settings.	8 'interrelated' categories of factors, including: pre-existing bounds; user orientation; approach; evaluator credibility; organizational factors; communication of findings; administrator style.
Alkin (1985)	Research synthesis	Three main categories of factors: human; context; and evaluation factors.
Leviton and	Research synthesis (strict focus upon utilization)	Five clusters of variables: relevance to needs and timeliness;

<sup>198</sup> Alkin and Taut consider 'enlightenment' to be a part of conceptual use.

Hughes (1981)		extent of communication; translation of findings to implications for programmes; credibility; commitment /advocacy by users.
Cousins and Leithwood (1986)	Research synthesis [note a broader definition of use: discrete decision-making; conceptual development; cognitive processing of evaluation information; and potential for use – which acts as an antecedent variable (Cousins & Leithwood, 1986: 359)].	Twelve factors influencing use, divided equally into the higher order categories of evaluation implementation and decision/policy setting, the latter applying to all potential users.
Shulha and Cousins (1997)	Research synthesis (building predominantly upon Cousins and Leithwood)	Emphasis upon understanding the context to explain evaluation use. Focus upon understanding through ‘in-depth naturalistic research’.
Preskill and Caracelli (1997)	Survey of perceptions and experience of evaluation use with members of the American Evaluation Association.	Six important factors: pre-evaluation planning for use; identifying/prioritising potential user/uses; within-budget evaluation designs; stakeholder involvement; formative communication; and a predefined communication plan.

**Table 23: Interview guide: main questions and follow ups**

**Background**

Name:  
Position:

(PERS – posit)

<p>Background and understanding of evaluation: <b>Personal</b></p> <p><b>Perception of organisation</b></p>	<ul style="list-style-type: none"> <li>➤ What is the main focus of your work?</li> <li>➤ What methods do you feel most comfortable with in your own research?</li> <li>➤ What do <i>you</i> understand the term evaluation to mean?             <ul style="list-style-type: none"> <li>○ What would characterise an <b>ideal</b> evaluation?</li> <li>○ Are there any <b>theories or texts</b> that influence your approach?</li> <li>○ How much experience of performing evaluations have you had?</li> <li>○ Was what the purpose of your most recent evaluation?</li> <li>○ What do you consider is the main purpose behind your organisation’s evaluations?</li> <li>○ How is this related to _____(org) framework for evaluation?</li> <li>○ What do you think is your organisation’s (programme of study’s) general approach to evaluation?</li> <li>○ NEW What kind of training in evaluation do you get from your organisation?</li> </ul> </li> </ul>
<p><b>Purposes / Demands</b> for evaluation</p>	<ul style="list-style-type: none"> <li>➤ What influenced your decision to evaluate the programme?</li> <li>➤ Who decides what to evaluate in relation to the programme? ♦ and how?             <ul style="list-style-type: none"> <li>○ Do your evaluations vary from programme to programme? E.g. external courses?</li> </ul> </li> <li>➤ What demands come from the institution?             <ul style="list-style-type: none"> <li>○ What demands come externally to the organisation?</li> <li>○ What kinds of demands are placed on your</li> </ul> </li> </ul>

	<p>programme in terms of expected effects / results? ♦Int /ext</p> <ul style="list-style-type: none"> <li>○ How do you attempt to meet these demands?</li> <li>○ How much freedom do you have in relation to design etc?</li> </ul>
<p><b>Designs</b> for evaluation</p>	<ul style="list-style-type: none"> <li>➤ What do you intend to evaluate in your programmes? <ul style="list-style-type: none"> <li>● How is the design <b>formed</b>?</li> <li>● What <b>role</b> do you play in the design of evaluations?</li> <li>● How do you decide what <b>methods /models</b> to use?</li> <li>● Would you like to develop / change this process? <ul style="list-style-type: none"> <li>○ What inhibits <b>change</b> if anything?</li> <li>○ What promotes <b>change</b>, if anything?</li> </ul> </li> </ul> </li> <li>➤ How is the Programme <b>rationale and design</b> reflected in the evaluation model chosen? <ul style="list-style-type: none"> <li>● To what extent do you experience that you manage to do this?</li> <li>● What role do you play in ensuring this?</li> </ul> </li> <li>➤ What is your experience of the way your evaluations are <b>followed up and used</b>? <ul style="list-style-type: none"> <li>● How do you think this influences the way you conduct evaluations?</li> </ul> </li> </ul>

Decision makers	<ul style="list-style-type: none"> <li>➤ Who makes decisions in relation to how evaluations are to be implemented? <ul style="list-style-type: none"> <li>• What role do different actors play in the process? (leadership, admin, line)</li> <li>• Who <b>is</b> responsible for the process?</li> <li>• Who <b>takes</b> responsibility for the process?</li> <li>• (Follow: <b>Reason?</b> (Demands, time, competence?))</li> </ul> </li> <li>➤ How are the findings written up and reported? <ul style="list-style-type: none"> <li>• How is the information used?</li> <li>• How do these processes influence the way you develop future evaluations?</li> </ul> </li> <li>• What kind of knowledge and skills /resources are needed, according to you, to evaluate systematically and effectively?</li> </ul>
	Who else should I speak to about these processes?



**Table 24: Initial Coding template**

Initial Template of Codes	
<b>BACKGROUND</b>	
PERS – pos	Person, position in subunit/organisation
PERS – work foc	Main focus of work
PERS – met	Main methods favoured / used in work/research
PERS – eval und	Definition of evaluation
PERS – eval ex	Evaluation experience
PERS – eval cont	Characterisation of ideal evaluation
PERS – eval theo	Evaluation theory influence over work
ORG – eval purp	Purpose of recent evaluation
ORG – loc fra	Main purpose of evaluations in organisation
ORG – inst eval	Relationship to institutional framework
ORG – inst ideol	What is organisational approach to evaluation
<b>DEMANDS</b>	
DEM – dec inf	Influence on decision to evaluate
DEM – eval foc	Who decides what to evaluate
DEM – consist	Degree of consistency/variance from programme to programme
DEM – int	Demands from institution
DEM- ext	Demands externally
DEM – eval res	Demands for effects/results
DEM – att concur	How meet demands
<b>DESIGNS</b>	
DES- purp	Intention/focus of evaluation
DES – form	How design formed
DES – invol	Involvement in process
DES – meth	Choice of methods
DES – meth qus	Type of questions asked
DES – chan proc	Suggestions for change / improvement
DES – PPG	Purpose, process and goals (PPG) of programme in evaluation
DES – imp real	Management of PPG
DES – inf use	Impact of perceived degree of utilisation
<b>DECISION MAKERS</b>	
DEC – inv	Who involved in decision process about evaluation
DEC – inv adm	Role of administration
DEC – rep	How are findings reported
DEC – use	How are findings used

DEC – use inf dec	How does perception of use affect future designs
DEC – chang	How would you change this process
<b>DECISION MAKING</b>	
DECM – skil	What skills necessary for effective evaluation

**Table 25: Final coding template**

		Tree nodes	Nodes	
<b>Demands</b>	Environmental complexity	Perceived pressure from task environment	Policy makers and agencies	Systematisation
				Regulation
				Information
				Impact focus
			Commissioner Pressure	Programme availability
				Variability
				Commissioner competence
				Memory
			Participant Expectations and Demands	Implementation
				Prior experience
	Voluntary/involuntary			
	Perceived institutional pressure	Validation		
		Systematisation		
		Operative control		
		Throughput		
Student satisfaction				
Perceived within-unit demands	Improvement focus			
	Academic anchoring	Projecting competence		
		Protecting professionalism		
<b>Designs</b>	Underlying frameworks	Attitudes to evaluation	Activist	
			Determinist	
		Reflections over influences		
	Current evaluation designs			
	Structure and approach	Basic structure		
		Formality/informality		
	Purpose	Accountability		
		Improvement		
	Focus	Participant perception		
		Deliverer reflection		
		Commissioner framework		
	Design limitations	Structural problems		
		Student characteristics		
		Interpretation problems		
	Ascertaining programme effects	Programme activity		
		Post-programme reflection		
		Investigating improved pupil results		
	Problems with	Cause and effect		

		“effects” design	Level of observation	
			Complexity of programme purpose	
			Time Frame	
			Organisational constraint (endemic)	
<b>Decision making</b>	<b>Role issues</b>	Academic staff	Level of cooperation	
			Degree of socialisation	
			Evaluative competence	
		Administrative staff	Level of cooperation	
			Degree of socialisation	
			Evaluative competence	
		Programme participants	Level of cooperation	
			Degree of motivation	
			Evaluative competence	
		Relationship to organisation	Level of leadership	
			Degree of subunit autonomy	
			Degree of feedback	
		Commissioners	Organisational constraints	
			Degree of demand	
			Level of cooperation	
		<b>Processes</b>	Perceived internal proximity	Collegial Construction
	Local initiation/power sharing			
	Critical participatory			
	Development focused			
	Sensemaking approach			
	Dismissive Submission			Collective inactivity
				Central initiation / download
				Critical rejection
				Decoupling and disengagement
Individual interpretation				
<b>Responses</b>	Internal evaluative configuration	Reductionism		
		Increased bureaucracy		
		Overload		
		Legitimatising		
		Institutionalisation		
	Student participation	Reductionism		
		Evaluation overload		
		Consumerism		
			Academic / voice conflict	

			Formalisation
		Ascertaining programme effects	Impact focus
			Formalisation
			Academic resistance
		Commissioners	Balancing tension
			Bargaining
			Buffering
			Adaptation
		Academic anchoring	Projecting competence
			Protecting professionalism

## References

- Abma, T. A., & Noordegraaf, M. (2003). Public Managers Amidst Ambiguity: Towards a Typology of Evaluative Practices in Public Management *Evaluation*, 9(3), 285-306.
- Ackroyd, S. (2004). Methodology for management and organisation studies: some implications of critical realism In S. Fleetwood & S. Ackroyd (Eds.), *Critical realist applications in organisation and management studies* (pp. XV, 366 s.). London: Routledge.
- Ackroyd, S., & Fleetwood, S. (2000). Realism in contemporary organisation and management studies. In S. Ackroyd & S. Fleetwood (Eds.), *Realist perspectives on management & organisations*: Taylor & Francis / Books.
- Adler, P. S., & Borys, B. (1996). Two Types of Bureaucracy: Enabling and Coercive *Administrative Science Quarterly*, 41(1), 61-89.
- Afsar, A., Skedsmo, G., & Sivesind, K. (2006). Evaluering og kunnskapsutvikling i ledelse av utdanning. In K. Sivesind, G. Skedsmo & G. Langfeldt (Eds.), *Utdanningsledelse* (Educational Leadership. pp. 271 s.). Oslo: Cappelen akademisk forl.
- Albaek, E. (1996). Why All This Evaluation? Theoretical Notes and Empirical Observations on the Functions and Growth of Evaluation, with Denmark as an Illustrative Case *Canadian Journal of Program Evaluation/La Revue canadienne d'évaluation de programme*, 11(2), 1-34.
- Alkin, M. C. (Ed.). (2004). *Evaluation roots : tracing theorists' views and influences*. California: Thousand Oaks.
- Alkin, M. C., & Christie, C. (2004). An evaluation theory tree. In M. Alkin (Ed.), *Evaluation roots : tracing theorists' views and influences*. California: Thousand Oaks.
- Alkin, M. C., & Christie, C. (2005). *Unraveling Theorists' Evaluation Reality*
- Alkin, M. C., & Taut, S. (2003). Unbundling Evaluation Use *Studies in Educational Evaluation*, 29(1), 1-12.
- Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's levels of training criteria: Thirty years later *Personnel Psychology*, 42(2), 331.
- Allison, G. T. (1969). Conceptual models and the Cuban missile crisis *The American Political Science Review*, 63(3), p. 689-718.
- Allison, G. T. (1971). *Essence of decision : explaining the Cuban missile crisis*. Boston: Little, Brown.
- Allison, G. T., & Zelikow, P. D. (1999). *Essence of decision : explaining the Cuban missile crisis* (2nd ed.). New York: Longman.
- Amaral, A., Jones, G. A., & Karseth, B. (2002). Governing higher education: comparing national perspectives. In A. Amaral, G. A. Jones & B. Karseth (Eds.), *Governing higher education: national perspectives on institutional governance* (pp. XXXI, 298 s.). Dordrecht: Kluwer Academic Publishers.
- Andersen, S. S. (1997). *Case-studier og generalisering : forskningsstrategi og design*. Bergen-Sandviken: Fagbokforl.
- Andersen, S. S. (2006). Aktiv informantintervjuing (Active informant interviewing). *Norsk Statsvitenskaplig Tidsskrift*, 22(3), 278-298.
- Anderson, L., & Bennett, N. (Eds.). (2003). *Developing educational leadership : using evidence for policy and practice*. London: Sage.

- Andresani, G., & Ferlie, E. (2006). Studying governance within the British public sector and without *Public Management Review*, 8(3), 415-431.
- Archer, M. S. (2000). *Being human the problem of agency*. Cambridge: Cambridge University Press.
- Askling, B., & Henkel, M. (2006). Higher Education Institutions. In M. Kogan, M. Bauer, I. Bleiklie & M. Henkel (Eds.), *Transforming higher education: a comparative study* (pp. XIV, 201 s.). Dordrecht: Springer.
- Atkinson, P., & Coffey, A. (2004). Analysing documentary realities. In D. Silverman (Ed.), *Qualitative research: theory, method and practice* (2nd ed., pp. 56 - 75). London: Sage.
- Baert, P. (2005). *Philosophy of the Social Sciences: towards Pragmatism*. Cambridge: Polity.
- Baklien, B. (1993). Evalueringsforskning i Norge [Evaluation research in Norway] *Tidsskrift for samfunnsforskning*, 34(3), 261-274.
- Baldrige, J. V. (1971). *Power and conflict in the university: research in the sociology of complex organizations*. New York: John Wiley.
- Ball, S. J. (2001). Labour, learning and the economy: a 'policy sociology' perspective. In M. Fielding (Ed.), *Taking education really seriously: four years' hard labour* (pp. xi, 268 s.). London: Routledge/Falmer.
- Ball, S. J. (2003). The teacher's soul and the terrors of performativity *Journal of Education Policy*, 18(2), 215 - 228.
- Barber, M. (2001). High expectations and standards for all, no matter what. Creating a world class education service in England. In M. Fielding (Ed.), *Taking education really seriously: four years' hard labour* (pp. xi, 268 s.). London: Routledge/Falmer.
- Barker, B. (2007). The leadership paradox: Can school leaders transform student outcomes? *School Effectiveness and School Improvement*, 18(1), 21 - 43.
- Bates, R. (2004). A critical analysis of evaluation practice: the Kirkpatrick model and the principle of beneficence *Evaluation and Program Planning*, 27(3), 341.
- Bauer, M., & Kogan, M. (2006). Higher Education Policies: historical overview. In M. Kogan, M. Bauer, I. Bleiklie & M. Henkel (Eds.), *Transforming higher education: a comparative study* (pp. XIV, 201 s.). Dordrecht: Springer.
- Beach, L. R. (1990). *Image theory : decision making in personal and organizational contexts*. Chichester: Wiley.
- Beach, L. R., & Connolly, T. (2005). *The psychology of decision making : people in organizations* (2nd ed.). Thousand Oaks: Sage.
- Becher, T., & Kogan, M. (1992). *Process and structure in higher education*. London: Routledge.
- Beck, U. (1997). *The reinvention of politics : rethinking modernity in the global social order*. Cambridge: Polity Press.
- Bell, L., Bolam, R., & Cubillo, L. (2003). *A systematic review of the impact of school leadership and management on student outcomes*. London: EPPI Centre, Social Science Research Unit, Institute of Education
- Bendor, J. B. (2003). Herbert A. Simon: Political Scientist *Annual Review of Political Science*, 6(1), 433-471.
- Bendor, J. B., & Hammond, T. H. (1992). Rethinking Allison's models *The American Political Science review*, 86(2), p. 301-322.

- Berger, P. L., & Kellner, H. (1982). *Sociology reinterpreted : an essay on method and vocation*. Harmondsworth: Penguin Books.
- Berger, P. L., & Luckmann, T. (1967). *The social construction of reality a treatise in the sociology of knowledge*. Harmondsworth: Penguin.
- Bhaskar, R. (1978). *A realist theory of science*. Hassocks: The Harvester Press.
- Bhaskar, R. (1979). *The possibility of naturalism a philosophical critique of the contemporary human sciences*. Brighton: Harvester Press.
- Bhaskar, R. (1984). *Philosophical Ideologies*. Harvester: Brighton press.
- Bhaskar, R. (1986). *Scientific realism and human emancipation*. London: Verso.
- Bhola, H. S. (2003a). Introduction: The Social and Cultural Contexts of Educational Evaluation. In T. Kellaghan & D. L. Stufflebeam (Eds.), *International Handbook of Educational Evaluation*. Dordrecht: Kluwer Academic publ.
- Bhola, H. S. (2003b). Social and Cultural Contexts of Educational Evaluation: A Global Perspective. In T. Kellaghan & D. L. Stufflebeam (Eds.), *International Handbook of Educational Evaluation* (pp. 397-416). Dordrecht: Kluwer Academic Publ.
- Blackmur, D. (2007). The public regulation of Higher Education Qualities: rationale, processes, and outcomes. In D. F. Westerheijden, B. Stensaker & M. J. Rosa (Eds.), *Quality Assurance in Higher Education: trends in regulation, translation and transformation* (Vol. 13, pp. 15 - 45).
- Blau, P. M. (1973). *The organization of academic work*. New York: Wiley.
- Bleiklie, I. (1998). Justifying the Evaluative State: New Public Management Ideals in Higher Education *European Journal of Education*, 33(3), 299-316.
- Bleiklie, I. (2004). *Institutional conditions and the responsibilities of universities, Working Paper 13 - 2004*. Bergen: Stein Rokkan Centre for Social Studies Retrieved from
- Bleiklie, I. (2006). Policy regimes and policy making. In M. Kogan, M. Bauer, I. Bleiklie & M. Henkel (Eds.), *Transforming higher education: a comparative study* (pp. XIV, 201 s.). Dordrecht: Springer.
- Bleiklie, I. (2009). Norway as Higher Education Policy Maker – from Tortoise to Eager Beaver? In C. Paradeise, I. Bleiklie, E. Ferlie & E. Reale (Eds.), *University Governance: Western European Comparative Perspectives*. Dordrecht: Springer Netherlands.
- Bleiklie, I., & Kogan, M. (2006). Comparison and theories. In M. Kogan, M. Bauer, I. Bleiklie & M. Henkel (Eds.), *Transforming higher education: a comparative study* (pp. XIV, 201 s.). Dordrecht: Springer.
- Bolam, R. (2004). Reflections on the NCSL from a Historical Perspective *Educational Management Administration Leadership*, 32(3), 251-267.
- Bolam, R., Dunning, G., & Karstedt, P. (2002). *New heads in the new Europe*. Münster: Waxmann.
- Boyne, G., Kirkpatrick, I., & Kitchener, M. (2001). Introduction To The Symposium on New Labour and The Modernization of Public Management *Public Administration*, 79(1), 1.
- Braybrooke, D., & Lindblom, C. E. (1963). *A strategy of decision policy evaluation as a social process*. New York: Free Press.
- Brennan, J. L., & Shah, T. (2000). *Managing quality in higher education: an international perspective on institutional assessment and change*. Buckingham: Open University Press.



- Broadbent, J. (2007). Performance management systems in and of Higher Education Institutions in England: professionalism, managerialism and management [Electronic Version]. *Roehampton Research Papers* 3, 1-25. Retrieved 19.01.09, from <http://rrp.roehampton.ac.uk/bsspapers/3>
- Broadfoot, P. (2001). Empowerment or performativity? Assessment policy in the late twentieth century. In R. Phillips & J. Furlong (Eds.), *Education, reform, and the state : twenty-five years of politics, policy, and practice* (pp. xii, 268 p.). London: Routledge Falmer.
- Brundrett, M. (2001). The Development of School Leadership Preparation Programmes in England and the USA: A Comparative Analysis *Educational Management & Administration*, 29(2), 17.
- Brundrett, M. (2006). Evaluating the individual and combined impact of national leadership programmes in England: perceptions and practices *School Leadership & Management*, 26(5), 473 - 488.
- Brunsson, N. (1982). The Irrationality of Action and Action of Rationality: Decisions, Ideologies and Organizational Actions *Journal of Management Studies*, 19(1), 29-44.
- Brunsson, N. (1990). Deciding for responsibility and legitimation: Alternative interpretations of organizational decision-making *Accounting, Organizations and Society*, 15(1-2), 47-59.
- Brunsson, N. (2002). *The organization of hypocrisy: talk, decisions and actions in organizations*. Oslo: Abstrakt forl.
- Brunsson, N. (2007). *The consequences of decision-making*. Oxford: Oxford University Press.
- Brunsson, N., & Jacobsson, B. (Eds.). (2000). *A world of standards*. Oxford: Oxford University Press.
- Brunsson, N., & Olsen, J. P. (1998). Organization theory: thirty years of dismantling, and then? In N. Brunsson & J. P. Olsen (Eds.), *Organizing organizations* (pp. 352 s.). Bergen-Sandviken: Fagbokforl.
- Bubb, S., & Earley, P. (2007). *Leading and managing continuing professional development*. London: Paul Chapman.
- Burgoyne, J. G. (2008). Critical Realism. In R. Holt & R. Thorpe (Eds.), *The Sage dictionary of qualitative management research* (pp. XXII, 290 s.). Los Angeles: Sage.
- Burrell, G., & Morgan, G. (1979). *Sociological paradigms and organisational analysis: elements of the sociology of corporate life*. London, : Heineman.
- Bush, T. (1998). The National Professional Qualification for Headship: the key to effective school leadership? *School Leadership & Management*, 18(3), 321 - 333.
- Bush, T. (2003). *Theories of educational leadership and management* (3rd ed.). London: Sage.
- Bush, T. (2004). *The performance of leaders: aims, impact and development*. Paper presented at the BELMAS Annual Research Conference, 2004. Stone, Staffordshire.
- Bush, T. (2005a). Control, partnership and empowerment *Management in Education*, 19(2), 34-35.

- Bush, T. (2005b). Editorial: Educational Leadership and Management: Making a Difference? *Educational Management Administration Leadership*, 33(1), 5-7.
- Bush, T. (2005c). Preparation for School Leadership in the 21st Century: International Perspectives Paper presented at the First HEAD Research Conference, Oslo, June 2005, Oslo, Norway. In CEM Centre for Education Management Research.
- Bush, T. (2006). The National College for School Leadership: A Successful English Innovation? *Phi Delta Kappan*, 87(7), 508 - 511.
- Bush, T. (2007). *School Management Training, Country Report: England - Update*. Oslo: BI Norwegian School of Management
- Bush, T. (2008a). From Management to Leadership: Semantic or Meaningful Change? *Educational Management Administration Leadership*, 36(2), 271-288.
- Bush, T. (2008b). *Leadership and management development in education*. Los Angeles ; London: SAGE.
- Bush, T. (2008c). *School Management Training, Country Report 2007: England* (No. 28). Oslo: BI Norwegian School of Management
- Bush, T., Bell, L., & Middlewood, D. (2010). Introduction: new directions in educational leadership. In T. Bush, L. Bell & D. Middlewood (Eds.), *The principles of educational leadership and management* (pp. XVI, 300 s.). Los Angeles: Sage.
- Bush, T., Briggs, A. R. J., & Middlewood, D. (2006). The Impact of School Leadership Development: Evidence from the "New Visions" Programme for Early Headship *Journal of In-service Education*, 32(2), 185-200.
- Bush, T., Glover, D., & Harris, A. (2007). *Review of School Leadership Development*. Nottingham: NCSL
- Bæck, U. D., & Ringholm, T. (2004). *Skole og utdanning på dagsordenen? : en undersøkelse av kommunale og fylkeskommunale toppledere sin oppmerksomhet om innholdet i skole og utdanning (School and education on the agenda? A survey of local and regional authority top leaders' interest in school and education)*. Tromsø: NORUT samfunnsforskning.
- Caldwell, B. (2004). *School Leadership Development: How the UK experience compares across the world*. Melbourne Educational transformations
- Caracelli, V. J. (2000). Evaluation use at the threshold of the Twenty-first century. In V. J. Caracelli & H. Preskill (Eds.), *The expanding scope of evaluation use* (Vol. Winter, pp. 99 - 111). San Francisco, Calif.: Jossey-Bass.
- Cazenave, P. (1994). Evaluation and the decision making process: possible developments. In OECD (Ed.), *Evaluation and the decision making process in higher education: French, German and Spanish experiences* (pp. 204 s.). Paris: Organisation for Economic Co-operation and Development.
- Centre for Educational Research and Innovation (2001). *New school management approaches*. Paris: Organisation for Economic Co-operation and Development.
- Chelmsky, E. (1987). What Have We Learned about the Politics of Program Evaluation? *Educational Evaluation and Policy Analysis*, 9(3), 199-213.
- Childers, M. E. (1981). What Is Political about Bureaucratic-Collegial Decision-Making? *Review of Higher Education*, 5(1), 25-45.

- Christensen, S. (1979). Decision making and socialization. In J. G. March, J. P. Olsen & S. Christensen (Eds.), *Ambiguity and choice in organizations* (2nd ed., pp. 408 s.). Bergen: Universitetsforlaget.
- Christensen, T., Lie, A., & Læg Reid, P. (2007). Still fragmented government or reassertion of the centre? In T. Christensen & P. Læg Reid (Eds.), *Transcending new public management the transformation of public sector reforms* (pp. VI, 288 s.). Aldershot: Ashgate.
- Christensen, T., & Læg Reid, P. (1999). *New public management the trade-off between political governance and administrative autonomy*. Bergen: LOS-senteret.
- Christensen, T., & Læg Reid, P. (2007). Introduction - theoretical approach and research questions. In T. Christensen & P. Læg Reid (Eds.), *Transcending new public management the transformation of public sector reforms* (pp. VI, 288 s.). Aldershot: Ashgate.
- Christie, C., A. (2003). What Guides Evaluation? A Study of How Evaluation Practice Maps onto Evaluation Theory *New Directions for Evaluation*, 2003(97), 7-36.
- Clark, B. R. (1972). The Organizational Saga in Higher Education *Administrative Science Quarterly*, 17(2), 178-184.
- Clark, T., Sweet, R., Gruber, K. H., Lourtie, P., Santiago, P., & Sohlman, Å. (2006). *Thematic review of tertiary education: country note - Norway*. Paris: OECD Retrieved from
- Cohen, M. D., March, J. G., & Olsen, J. P. (1972). A Garbage Can Model of Organizational Choice *Administrative Science Quarterly*, 17(1), 1.
- Conner, C. (2005). A response from the NCSL *Management in Education*, 19(2), 37-38.
- Coupland, C., Currie, G., & Boyett, I. (2008). New Public Management and a Modernization Agenda: Implications for School Leadership *International Journal of Public Administration*, 31(9), 1079-1094.
- Cousins, J. B. (2003). Utilization Effects of participatory evaluation. In T. Kellaghan & D. L. Stufflebeam (Eds.), *International Handbook of Educational Evaluation* (pp. 245-266). Dordrecht: Kluwer Academic Press.
- Cousins, J. B., & Earl, L. M. (1995). *Participatory evaluation in education : studies in evaluation use and organizational learning*. Washington, D.C.: Falmer Press.
- Cousins, J. B., & Leithwood, K. A. (1986). Current Empirical Research on Evaluation Utilization *Review of Educational Research*, 56(3), 331-364.
- Cowie, M., & Crawford, M. (2007). Principal preparation - still an act of faith? *School Leadership & Management*, 27(2), 129 - 146.
- Cox, A. (2005). What are communities of practice? A comparative review of four seminal works *Journal of Information Science*, 31(6), 527-540.
- Creswell, J. W. (2003). *Research design : qualitative, quantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, Calif.: Sage Publications.
- Cruickshank, J. (2003). *Realism and sociology: anti-foundationalism, ontology, and social research*. London: Routledge.
- Cummings, R. (2002). Rethinking Evaluation Use. Paper presented at the Australasian Evaluation Society International Conference, Wollongong, Australia. In

- Dahl, T., Klewe, L., & Skov, P. (2004a). *En skole i bevægelse: evaluering af satsning på kvalitetsudvikling i den norskegrundskole (A school in motion: evaluation of the concentration on quality development in Norwegian basic education)*. København: Danmarks Pædagogiske Universitets Forlag.
- Dahl, T., Klewe, L., & Skov, P. (2004b). *Norsk skole i udvikling, men i ujevn takt: noen momenter fra evalueringen av kvalitetsutvikling i grunnskolen 2000-2003*. København: Danmarks Pædagogiske Universitets Forlag.
- Dahler-Larsen, P. (1998). *Den rituelle refleksion : om evaluering i organisationer (The ritual reflection: evaluation in organisations)*. Odense: Odense universitetforlag.
- Dahler-Larsen, P. (2000). Surviving the Routinization of Evaluation *Administration & Society*, 32(1), 70.
- Dahler-Larsen, P. (2001). Organisationalisering af evaluering. In P. Dahler-Larsen & H. K. Krogstrup (Eds.), *Tendenser i evaluering* (pp. 270 s.). Odense: Odense Universitetsforlag.
- Dahler-Larsen, P. (2004a). *Evaluering og magt : magtutredningen (Evaluation and power: Power report)*. Aarhus: Aarhus universitetsforlag.
- Dahler-Larsen, P. (2004b). Evalueringens mange anvendelsesformer - kontrol, læring, oplysning og konstitution af ny skolevirklighed. In I. Vinje, H. Bjørnsrud, T. E. Hauge, Utdanningsdirektoratet, Nasjonalt nettverk for skolebasert vurdering & Høgskolen i Vestfold . Avdeling for lærerutdanning (Eds.), *Kvalitet, tall og tillit : landskonferansen i skolebasert vurdering 2003, 17.-18. november 2003* (pp. 80 s.). Oslo: Utdanningsdirektoratet.
- Dahler-Larsen, P. (2005a). Evaluation and public management. In E. Ferlie, L. E. Lynn & C. Pollitt (Eds.), *The Oxford handbook of public management* (pp. XIII, 789 s.). Oxford: Oxford University Press.
- Dahler-Larsen, P. (2005b). New Public Management og evaluering. In T. Busch, E. Johnsen, J. O. Vanebo & K. Klaudi Klausen (Eds.), *Modernisering av offentlig sektor : utfordringer, metoder og dilemmaer* (2. utg. ed.). Oslo: Universitetsforl.
- Dahler-Larsen, P. (2006a). Evaluation after disenchantment? Five issues shaping the role of evaluation in society. In I. F. Shaw, M. M. Mark & J. C. Greene (Eds.), *Handbook of evaluation : policies, programs and practices* (pp. XXI, 608 s.). London: Sage Publ.
- Dahler-Larsen, P. (2006b). *Evalueringkultur: et begreb bliver til ('Evaluationculture': a concept is formed)*. Odense: Syddansk Universitetsforlag.
- Dahler-Larsen, P. (2007). Constitutive effects of performance indicator systems. In S. Kushner & N. J. Norris (Eds.), *Dilemmas of engagement: evaluation and the new public management* (pp. X, 157 s.). Amsterdam: Elsevier JAI.
- Dahler-Larsen, P. (2008). *Kvalitetens beskaffenhed (The characteristics of quality)*. Odense: Syddansk Universitetsforlag.
- Dahler-Larsen, P., & Krogstrup, H. K. (2000). Evaluering og institutionelle standarder: nyinstitutionelle betragtninger af evaluering som vor tids ritual (Evaluation and institutional standards: New-institutional considerations of evaluation as the ritual of our time). *Politica: tidsskrift for politisk videnskab*, 32(3), 283-299.

- Danø, T., & Stensaker, B. (2007). Still Balancing Improvement and Accountability? Developments in External Quality Assurance in the Nordic Countries 1996-2006 *Quality in Higher Education*, 13(1), 81 - 93.
- Das, T. K., & Teng, B.-S. (1999). Cognitive Biases and Strategic Decision Processes: An Integrative Perspective *Journal of Management Studies*, 36(6), 757-778.
- Davies, P., Newcomer, K., & Soydan, H. (2006). Government as structural context for evaluation. In I. F. Shaw, M. M. Mark & J. C. Greene (Eds.), *Handbook of evaluation : policies, programs and practices* (pp. XXI, 608 s.). London: Sage Publ.
- Day, C., Sammons, P., Hopkins, D., Leithwood, K., & Kington, A. (2008). Research into the impact of school leadership on pupil outcomes: policy and research contexts *School Leadership & Management*, 28(1), 5 - 25.
- de la luz Fernández-Alles, M., & Valle-Cabrera, R. (2006). Reconciling institutional theory with organizational theories: How neoinstitutionalism resolves five paradoxes *Journal of Organizational Change Management*, 19(4), 503-517.
- Denham, A. (2003). Public services. In P. Dunleavy (Ed.), *Developments in British politics 7* (pp. XVIII, 398 s.). Basingstoke: Palgrave.
- Denzin, N. K., & Lincoln, Y. S. (Eds.). (2000). *Handbook of qualitative research* (2nd ed.). Thousand Oaks, Calif.: Sage.
- Desforges, C. (2003). Evidence Informed Policy and Practice in teaching and learning. In L. Anderson & N. Bennett (Eds.), *Developing educational leadership : using evidence for policy and practice* (pp. XX, 188 s.). London: Sage.
- DeYoung, D. J., & Conner, R. F. (1982). Evaluator Preconceptions about Organizational Decision Making: Rational versus Incremental Perspectives *Evaluation Review*, 6(3), 431-440.
- DfEE. (2000). *National College for School Leadership remit letter, 25th September 2000*. Retrieved from.
- DfES, & NCSL. (2004). *School leadership: end to end review of school leadership policy and delivery*. Retrieved 12.11.2007. from <http://www.ncsl.org.uk/mediastore/image2/etoe-review-final-report.pdf>.
- DiMaggio, P. J., & Powell, W. W. (1982). *The iron cage revisited : conformity and diversity in organizational fields*. New Haven, Conn.: Institution for Social and Policy Studies Yale University.
- DiMaggio, P. J., & Powell, W. W. (1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields *American Sociological Review*, 48(2), 147-160.
- Djupedal, Ø. K. (2007). *Læring og ledelse i Kunnskapsløftet (Learning and leadership in the Knowledge Promotion reform), Minister of Education and Research's presentation at the University of Oslo "School Leadership Days" 27.07.2007* Retrieved 02.07.2007 from <http://utdanningsledelse.uio.no/skoldag/2007/bidragstere/presentasjoner/files/Djupedal.htm>.
- Dooley, D. (1990). *Social research methods* (2nd ed.). Englewood Cliffs, N.J.: Prentice Hall.
- Dornbusch, S. M., & Scott, W. R. (1975). *Evaluation and the exercise of authority*. San Francisco: Jossey-Bass.

- Drake, D. (1921). The approach to critical realism. In D. Drake, A. O. Lovejoy, J. B. Pratt, A. K. Rogers, G. Santayana, R. W. Sellars & C. A. Strong (Eds.), *Essays in critical realism a co-operative study of the problem of knowledge* (pp. IX, 244 s.). London: Macmillan.
- Drake, D., Lovejoy, A. O., Pratt, J. B., Rogers, A. K., Santayana, G., Sellars, R. W., et al. (Eds.). (1921). *Essays in critical realism a co-operative study of the problem of knowledge*. London: Macmillan.
- Dutton, J. E., Ashford, S. J., Lawrence, K. A., & Miner-Rubino, K. (2002). Red Light, Green Light: Making Sense of the Organizational Context for Issue Selling *Organization Science*, 13(4), 355-369.
- Dutton, J. E., Ashford, S. J., O'Neill, R. M., & Lawrence, K. A. (2001). MOVES THAT MATTER: ISSUE SELLING AND ORGANIZATIONAL CHANGE *Academy of Management Journal*, 44(4), 716-736.
- Earley, P. (2005). Assessing Impact - why evaluation is hard *Management in Education*, 19(2), 37.
- Earley, P., & Evans, J. (2004). Making a Difference?: Leadership Development for Headteachers and Deputies--Ascertaining the Impact of the National College for School Leadership *Educational Management Administration Leadership*, 32(3), 325-338.
- Easterby-Smith, M. (1994). *Evaluating management development, training, and education* (2nd ed. ed.). Aldershot: Gower.
- Eliassen, K. A., & Sitter, N. (2007). *Understanding public management and administration*: Sage.
- Elliott, J., & Doherty, P. (2001). Restructuring educational research for the Third Way. In M. Fielding (Ed.), *Taking education really seriously: four years' hard labour* (pp. xi, 268 s.). London: Routledge/Falmer.
- Ellstrom, P.-E. (1983). Four Faces of Educational Organizations *Higher Education*, 12(2), 231-241.
- ENQA. (2005). *Standards and guidelines for quality assurance in the European Higher Education Area*. Helsinki: European Association for Quality Assurance in Higher Education Retrieved 26.03.2007 from <http://www.enqa.eu/files/ENQA%20Bergen%20Report.pdf>.
- Etzioni, A. (1967). Mixed-Scanning: A "Third" Approach to Decision-Making *Public Administration Review*, 27(5), 385.
- Etzioni, A. (1986). Mixed Scanning Revisited *Public Administration Review*, 46(1), 8.
- Etzioni, A. (1989). Humble Decision Making *Harvard Business Review*, 67(4), 122.
- European Universities Association. (2001). *Salamanca Convention 2001: the Bologna Process and the European Higher Education Area*. Brussels: EUA Retrieved 03.02.09 from [www.EUA.be](http://www.EUA.be).
- European Universities Association. (2007). *EUA policy position on Quality*. Brussels: EUA Retrieved 03.02.09 from [www.EUA.be](http://www.EUA.be).
- Evan, W. M. (1993). *Organization theory: research and design*. New York: Macmillan.
- Fairclough, N. (2005). Peripheral Vision: Discourse Analysis in Organization Studies: The Case for Critical Realism *Organization Studies*, 26(6), 915-939.

- Feldman, M. S., & March, J. G. (1981). Information in Organizations as Signal and Symbol *Administrative Science Quarterly*, 26(2), 171.
- Ferlie, E., Musselin, C., & Andresani, G. (2008). The steering of higher education systems: a public management perspective *Higher Education*, 56(3), 325-348.
- Ferlie, E., Musselin, C., & Andresani, G. (2009). The steering of Higher Education: a public management perspective. In C. Paradeise, E. Reale, I. Bleiklie & E. Ferlie (Eds.), *University Governance: Western European Comparative Perspectives*. Dordrecht: Springer.
- Fielding, M. (2001). Taking education really seriously: four years' hard labour. In M. Fielding (Ed.), *Taking education really seriously: four years' hard labour* (pp. xi, 268 s.). London: Routledge/Falmer.
- Finstad, N., & Kvåle, G. (2004). Skoleutvikling og ledelse: Skoleledelse mellom kommunale, statlige og profesjoenlle interesser (School development and leadership: School leadership caught between the local authority, state and professional interests). In G. Imsen (Ed.), *Det ustyrilige klasserommet* (The ungovernable classroom. pp. s. 165-196). Oslo: Universitetsforlaget.
- Fitz-Gibbon, C. T. (2002). Evaluation in an Age of Indicators: Challenges for Public Sector Management *Evaluation*, 8(1), 140-148.
- Flecknoe, M. (2000). Can Continuing Professional Development for Teachers Be Shown To Raise Pupils' Achievement? *Journal of In-service Education*, 26(3), 437-457.
- Flecknoe, M. (2002). Measuring the Impact of Teacher Professional Development: Can It Be Done? *European Journal of Teacher Education*, 25(2-3), 119-134.
- Fleetwood, S. (2005). Ontology in Organization and Management Studies: A Critical Realist Perspective *Organization*, 12(2), 197-222.
- Flynn, N. (2007). *Public sector management* (5th ed.). London: Sage.
- Foss, O., & Mønnesland, J. (Eds.). (2005). *Evaluering av offentlig virksomhet : metoder og vurderinger* (Evaluation of Public Activity: methods and valuations. 2. utg. ed.). Oslo: NIBR
- Furusten, S. (2000). The knowledge base of standards. In N. Brunsson & B. Jacobsson (Eds.), *A world of standards* (pp. IX, 188 s.). Oxford: Oxford University Press.
- Gamoran, A., & Dreeben, R. (1986). Coupling and Control in Educational Organizations *Administrative Science Quarterly*, 31(4), 612-632.
- Gibbs, G. R. (2002). *Qualitative data analysis: explorations with NVivo*. Buckingham: Open University Press.
- Gibbs, G. R. (2007). *Analyzing qualitative data*. London: Sage.
- Gigerenzer, G., & Selten, R. (2001). Rethinking rationality. In G. Gigerenzer & R. Selten (Eds.), *Bounded Rationality: the adaptive toolbox*. Cambridge, Mass.: The MIT Press.
- Gioia, D. A., & Chittipeddi, K. (1991). SENSEMAKING AND SENSEGIVING IN STRATEGIC CHANGE INITIATION *Strategic Management Journal*, 12(6), 433-448.
- Glatter, R. (2004). Leadership and leadership development in education. In J. Storey (Ed.), *Leadership in organizations: current issues and key trends* (pp. XIII, 349 s.). London: Routledge.

- Glatter, R., Mulford, B., & Shuttleworth, D. (2003). Governance, Management and Leadership. In Centre for Educational Research and Innovation (Ed.), *Networks of innovation: towards new models for managing schools and systems* (pp. 182 s.). Paris: Organisation for economic co-operation and development.
- Goldstein, I. L., & Ford, J. K. (2002). *Training in organizations : needs assessment, development, and evaluation* (4th ed.). Australia: Wadsworth/Thomson Learning.
- Gomm, R., Foster, P., & Hammersley, M. (2000). *Case study method : key issues, key texts*. London: Sage.
- González, M., Glasman, N. S., & Glasman, L. D. (2002). Daring to Link Principal Preparation Programs to Student Achievement in Schools *Leadership and Policy in Schools, 1*(3), 265 - 283.
- Gornitzka, Å., Maassen, P. A. M., Olsen, J. P., & Stensaker, B. (2007). "Europe of knowledge:" Search for a new pact. In P. A. M. Maassen & J. P. Olsen (Eds.), *University dynamics and European integration* (pp. XII, 245 s.). Dordrecht: Springer.
- Gornitzka, Å., & Olsen, J. P. (2006). *Europeiske endringsprosesser og høyere utdanningsinstitusjoner (European change processes and higher education institutions)*. Oslo: Arena.
- Grant, W. (2003). Economic policy. In P. Dunleavy (Ed.), *Developments in British politics 7* (pp. XVIII, 398 s.). Basingstoke: Palgrave.
- Gray, A., & Jenkins, B. (2002). Policy and Program Evaluation in the United Kingdom: A Reflective State? In R. C. Rist, R. Sandahl & J.-E. Furubo (Eds.), *International atlas of evaluation*. New Brunswick, N.J.: Transaction Publ.
- Greene, J. C. (1988). Stakeholder Participation and Utilization in Program Evaluation *Evaluation Review, 12*(2), 91-116.
- Greene, J. C. (2002). *Towards Evaluation as a 'Public Craft' and Evaluators as Stewards of the Public Good or On Listening Well*. Paper presented at the 2002 Australasian Evaluation Society International Conference. Wollongong, Australia. Retrieved 28.04.08, from [http://www.evaluationcanada.ca/distribution/20021030\\_greene\\_jennifer.pdf](http://www.evaluationcanada.ca/distribution/20021030_greene_jennifer.pdf)
- Greene, J. C., & McClintock, C. (1985). Triangulation in Evaluation: Design and Analysis Issues *Eval Rev, 9*(5), 523-545.
- Greene, J. C., & Walker, K. (2001). Review of evaluation: an integrated framework for understanding, guiding, and improving policies and programs: Melvin M. Mark, Gary T. Henry, and George Julnes. Jossey-Bass, San Francisco, 2000 *Evaluation and Program Planning, 24*(4), 368.
- Greve, A. (1995). *Organisasjonsteori: nyere perspektiver*. Oslo: Universitetsforl.
- Greve, C. (2006). Ikke mange offentlige organisationer i Andeby! New Public Management og offentlig organisationsteori (There are not many public organisations in Duckburg! New Public Management and public organisation theory). In H. F. Hansen (Ed.), *Den organiserede forvaltning: politik, viden og værdier i samspil* (The organised administration: politics, knowledge and values in interaction.). Copenhagen: Forlaget Politiske Studier.



- Gronn, P. (2003). Leadership's place in a community of practice. In M. Brundrett, N. Burton & R. Smith (Eds.), *Leadership in education* (pp. xx, 202 s.). London: Sage.
- Gu, Q., Sammons, P., & Mehta, P. (2008). Leadership characteristics and practices in schools with different effectiveness and improvement profiles *School Leadership & Management*, 28(1), 43 - 63.
- Gulbrandsen, M., & Stensaker, B. (2003). Evaluering av forskning og utdanning - bidrag til en helhetlig kvalitetsutvikling? (Evaluation of research and education - contributing to an holistic quality development?). In I. M. Larsen & B. Stensaker (Eds.), *Tradisjon og tilpasning: organisering og styring av universitetene* (Tradition and adaptation - organising and governing the Universities. pp. 196 s.). Oslo: Cappelen.
- Gunter, H. (1999). Contracting Headteachers as Leaders: an analysis of the NPQH *Cambridge Journal of Education*, 29(2), 251 - 264.
- Gunter, H. (2001). Modernising headteachers as leaders. An analysis of the NPQH. In M. Fielding (Ed.), *Taking education really seriously: four years' hard labour* (pp. xi, 268 s.). London: Routledge/Falmer.
- Gunter, H. (2005). Putting Education Back into Leadership *FORUM: for promoting 3-19 comprehensive education*, 47(2), 181-188.
- Guskey, T. R. (2000). *Evaluating professional development*. Thousand Oaks ; London: Corwin Press.
- Guskey, T. R. (2002). Does It Make a Difference? Evaluating Professional Development *Educational Leadership*, 59(6), 45-51.
- Guskey, T. R. (2003). *The Characteristics of Effective Professional Development: A Synthesis of Lists*. Access ERIC: FullText (143 Reports--Research; 150 Speeches/Meeting Papers). Kentucky
- Gvaramadze, I. (2008). From Quality Assurance to Quality Enhancement in the European Higher Education Area *European Journal of Education*, 43(4), 443-455.
- Hagen, A., Nyen, T., & Hertzberg, D. (2006). *Evaluering av "Kompetanse for utvikling. Strategi for kompetanseutvikling i grunnsopplæringen 2005-2008": Delrapport 1 (Evaluation of "Competence for development. The strategy for competence development in compulsory education 2005 - 2008": Part 1)*. Oslo: FAFO Retrieved from
- Hall, R. H., & Tolbert, P. S. (2005). *Organizations: structures, processes, and outcomes*. Upper Saddle River, N.J.: Prentice Hall.
- Hallinger, P. (Ed.). (2003). *Reshaping the landscape of school leadership development : a global perspective*. Lisse: Swets & Zeitlinger Publ.
- Hallinger, P., & Heck, R. H. (1998). Exploring the Principal's Contribution to School Effectiveness: 1980-1995 *School Effectiveness and School Improvement*, 9(2), 157-191.
- Hallinger, P., & Heck, R. H. (2003). Understanding the contribution of leadership to school improvement. In M. Wallace & L. Poulson (Eds.), *Learning to read critically in educational leadership and management* (pp. viii, 247 s.). London: Sage.
- Hamblin, A. C. (1974). *Evaluation and control of training*. London: McGraw-Hill.
- Hansen, H. F. (2001). Evaluering: teori og praksis i designprosessen *Nordiske Organisasjonsstudier*, 3(3), 43 - 62.

- Hansen, H. F. (2005a). Choosing Evaluation Models: A Discussion on Evaluation Design *Evaluation*, 11(4), 447-462.
- Hansen, H. F. (2005b). Kvalitet i evalueringspraksis (Quality in evaluation practice). *Nordisk Administrativ Tidsskrift*, 86(1), 27 - 43.
- Hansen, H. F. (2009). Educational Evaluation in Scandinavian Countries: Converging or Diverging Practices? *Scandinavian Journal of Educational Research*, 53(1), 71 - 87.
- Hansen, H. F. (Ed.). (2006). *Den organiserede forvaltning: politik, viden og værdier i samspil* (The organised administration: politics, knowledge and values in interaction.). Copenhagen: Forlaget Politiske Studier.
- Hansen, H. F., & Borum, F. (1995). Beneath the smooth surface of evaluation rhetorics. Paper presented at the 13th Nordic Conference on Business Studies, Copenhagen. In/
- Hardy, C. (1990a). Putting power into university governance. In J. C. Smart (Ed.), *Higher Education: handbook of theory and research* (Vol. VI, pp. viii, 528 s.). New York: Agathon Press.
- Hardy, C. (1990b). Strategy and Context: Retrenchment in Canadian Universities *Organization Studies* (Walter de Gruyter GmbH & Co. KG.), 11(2), 207-237.
- Hardy, C. (1991). Pluralism, power and collegiality in universities *Financial Accountability & Management*, 7(3), 127.
- Hardy, C., Langley, A., Mintzberg, H., & Rose, J. (1983). Strategy Formation in the University Setting *Review of Higher Education*, 6(4), 407-433.
- Hargreaves, D. H. (2001). Revitalising educational research. Past lessons and future prospects. In M. Fielding (Ed.), *Taking education really seriously: four years' hard labour* (pp. xi, 268 s.). London: Routledge/Falmer.
- Harris, A. (2003). *Effective leadership for school improvement*. London: RoutledgeFalmer.
- Harvey, L. (2004-8). Analytic Quality Glossary [Electronic Version]. *Quality Research International*. Retrieved 06.08.08, from <http://www.qualityresearchinternational.com/glossary/>
- Harvey, L., & Newton, J. (2007). Transforming quality evaluation: moving on. In D. F. Westerheijden, B. Stensaker & M. J. Rosa (Eds.), *Quality Assurance in Higher Education: trends in regulation, translation and transformation* (pp. 295 - 312).
- Harvey, L., & Stensaker, B. (2008). Quality Culture: Understandings, Boundaries and Linkages *European Journal of Education*, 43(4), 427-442.
- Haug, P., & Monsen, L. (2002). *Skolebasert vurdering : erfaringer og utfordringer* (School based assessment: experiences and challenges). Oslo: Abstrakt forl.
- Hegtun, A. (2007). "Improving school leadership": profesjonell skoleledelse - noe vi alle vil - trenger vi OECD for å få det til? ("Improving school leadership": professional school leadership - something we all want - do we need OECD to accomplish it?). Unpublished Master thesis, University of Oslo, Oslo.
- Hellstern, G.-M. (1986). Assessing evaluation research. In V. Ostrom, G. Majone, F.-X. Kaufmann & The Bielefeld Interdisciplinary Project (Eds.), *Guidance, control, and evaluation in the public sector : the Bielefeld*

- Interdisciplinary Project* (Vol. De Gruyter studies in organization ; 4, pp. XIII, 830 s.). Berlin: De Gruyter.
- Henkel, M. (2000). *Academic identities and policy change in higher education*. London: Jessica Kingsley.
- Henkel, M. (2002). Emerging concepts of academic leadership and their implications for intra-institutional roles and relationships in higher education *European Journal of Education*, 37(1), 29 - 41.
- Henkel, M. (2004). *The impacts of evaluation upon academic identities and the links with managerialism*. Paper presented at the UNESCO Forum: First regional research seminar for Europe and North America: Managerialism and evaluation in Higher Education. Paris, France 4-5 November, 2004. Retrieved 14.05.08, from <http://unesdoc.unesco.org/images/0014/001467/146741E.pdf>
- Henry, G. T. (2001). Evaluation models - Coeditor's notes *New directions for evaluation*, 89 (April 2001), 106 p.
- Henry, G. T., & Mark, M. M. (2003). Beyond Use: Understanding Evaluation's Influence on Attitudes and Actions *The American Journal of Evaluation*, 24(3), 293.
- Herling, R. W. (2003). Bounded Rationality and the Implications for HRD *Advances in Developing Human Resources*, 5(4), 393-407.
- Hesketh, A., & Fleetwood, S. (2006). Beyond Measuring the Human Resources Management - Organizational Performance Link: Applying Critical Realist Meta-Theory *Organization*, 13(5), 677-699.
- Higham, R., Hopkins, D., & Ahtaridou, E. (2007). *Improving School Leadership: Country background report for England*. London: LCEL, Institute for Education, University of London.
- Hofstetter, C. H., & Alkin, M. (2003). Evaluation Use Revisited. In T. Kellaghan & D. L. Stufflebeam (Eds.), *International Handbook of Educational Evaluation* (pp. 197-222). Dordrecht: Kluwer Academic Publishers.
- Holton III, E. F. (1996). The Flawed Four-Level Evaluation Model. *Human Resource Development Quarterly*, 7(1), 5 - 21.
- Holton III, E. F. (2005). Holton's Evaluation Model: New Evidence and Construct Elaborations *Advances in Developing Human Resources*, 7(1), 37.
- Holton III, E. F., & Naquin, S. (2005). A critical analysis of HRD evaluation models from a decision-making perspective *Human Resource Development Quarterly*, 16(2), 257.
- Holton, J. A. (2007). The coding process and its challenges. In A. Bryant & K. Charmaz (Eds.), *The SAGE handbook of grounded theory* (pp. XXX, 623 s.). Los Angeles, Calif.: Sage.
- Homburg, V., Pollitt, C., & van Thiel, S. (2007). Introduction. In C. Pollitt, S. van Thiel & V. Homburg (Eds.), *New public management in Europe: adaptations and alternatives* (pp. XII, 229 s.). Houndmills, Basingstoke: Palgrave Macmillan.
- Hood, C. (1991). A public management for all seasons? *Public Administration*, 69(1), 3.
- Hood, C. (2005). Public Management: the word, the movement, the science. In C. Pollitt, E. Ferlie & L. E. Lynn (Eds.), *The Oxford handbook of public management* (pp. XIII, 789 s.). Oxford: Oxford University Press.

- Hood, C., & Peters, G. (2004). The Middle Aging of New Public Management: Into the Age of Paradox? *Journal of Public Administration Research and Theory*, 14(3), 267-282.
- House, E. R. (1978). Assumptions underlying evaluation models *Educational Researcher*, 7(3), 4-12.
- House, E. R. (1993). *Professional evaluation : social impact and political consequences*. Newbury Park, Calif.: Sage.
- House, E. R. (2006). Democracy and Evaluation *Evaluation*, 12(1), 119-127.
- House, E. R., & Howe, K. R. (1999). *Values in evaluation and social research*. Thousand Oaks, Calif.: Sage.
- Huber, G. P., & McDaniel, R. R. (1986). The Decision-making paradigm of organizational design *Management Science*, 32(5), 572-589.
- Huber, S. G. (2004). School leadership and leadership development: Adjusting leadership theories and development programs to values and the core purpose of school *Journal of Educational Administration*, 42(6), 669-684.
- Hyyryläinen, E., & Viinamäki, O.-P. (2008). The Implications of the Rationality of Decision-Makers on the Utilization of Evaluation Findings *International Journal of Public Administration*, 31(10/11), 1223-1240.
- Immergut, E. M. (1998). The theoretical core of the new institutionalism *Politics & Society*, 26(1), 5-34.
- Imsen, G. (2003). Democratic evaluation and evaluation for learning - a reply to Ove Karlsson. In P. Haug, T. A. Schwandt & N. forskningsråd (Eds.), *Evaluating educational reforms : Scandinavian perspectives* (pp. 149 - 168). Greenwich, Conn.: Information Age Publ.
- Jennings, P. D., & Greenwood, R. (2003). Constructing the iron cage: institutional theory and enactment. In R. I. Westwood & S. Clegg (Eds.), *Debating organization: point-counterpoint in organization studies* (pp. 195 - 207). Malden, Mass.: Blackwell Publishing.
- Johansen, K., & Tjeldvoll, A. (1989). Milieu and Leadership in the School - a Norwegian Inset-programme without Goal Attainment. In J. Wilson (Ed.), *The effectiveness of in-service education and training of teachers and school leaders: report of the Fifth All-European Conference of Directors of Educational Research Institutions, Triesenberg (Liechtenstein) 11-14 October 1988* (pp. 161 s.). Amsterdam: Swets & Zeitlinger.
- Johnsen, Å., Nørreklit, H., & Vakkuri, J. (2006). Introducing a Nordic Perspective on Public Sector Performance Measurement *Financial Accountability & Management*, 22(3), 207-212.
- Johnson, R. B. (1998). Toward a theoretical model of evaluation utilization *Evaluation and Program Planning*, 21(1), 93.
- Kalberg, S. (1980). Max Weber's Types of Rationality: Cornerstones for the Analysis of Rationalization Processes in History *American Journal of Sociology*, 85(5), 1145.
- Karlsen, G. E. (1993). *Desentralisering - løsning eller oppløsning : søkelys på norsk skoleutvikling og utdanningspolitikk (Decentralisation - solution or dissolution: spotlight upon Norwegian school development and education policy)*. [Oslo]: Ad Notam Gyldendal.
- Karlsen, G. E. (2003). Enhetskolen, styring og marked (The comprehensive school, governance and market). In P. Østerud & J. Johnsen (Eds.), *Leve skolen!* :

- enhetsskolen i et kulturkritisk perspektiv* (Long live school!: the comprehensive school in a culture-critical perspective. pp. 218-230). Vallset: Oplandske bokforl.
- Karlsen, G. E. (2004). INNHOGG: St. meld. nr. 30 (2003-2004) Kultur for læring - et utdanningspolitisk perspektiv (Government White Paper number 30 (2003-2004) Culture for Learning - an educational policy perspective). *Norsk pedagogisk tidsskrift*, 2004(4), 331 - 335.
- Karlsen, G. E. (2006). *Utdanning, styring og marked: norsk utdanningspolitikk i et internasjonalt perspektiv* (Education, governance and market: Norwegian education policy in an international perspective) (2nd ed.). Oslo: Universitetsforl.
- Karlsson, O. (2003a). Democratic evaluation: an arena for dialogue and learning? In P. Haug, T. A. Schwandt & N. forskningsråd (Eds.), *Evaluating educational reforms : Scandinavian perspectives* (pp. 133-147). Greenwich, Conn.: Information Age Publ.
- Karlsson, O. (2003b). Evaluation politics in Europe: trends and tendencies *Studies in Educational Policy and Educational Philosophy: E-tidsskrift*, 2003: 1.
- Karlsson, O. (2003c). Program Evaluation in Europe: Between Democratic and New Public Management Evaluation. In T. Kellaghan & D. L. Stufflebeam (Eds.), *International Handbook of Educational Evaluation* (pp. 429-440). Dordrecht: Kluwer Academic Publ.
- Kemm, J. (2006). The limitations of 'evidence-based' public health *Journal of Evaluation in Clinical Practice*, 12(3), 319-324.
- King, N. (1998). Template Analysis. In G. Symon & C. Cassell (Eds.), *Qualitative methods and analysis in organizational research : a practical guide* (pp. XII, 255 s.). London: Sage.
- King, N. (2004). Using templates in a thematic analysis of text. In G. Symon & C. Cassell (Eds.), *Essential guide to qualitative methods in organizational research* (pp. XVIII, 388 s.). London: Sage Publications.
- King, N. (2008a). Developing your template. *Template Analysis*. Retrieved 16.07.2008, from [http://www.hud.ac.uk/hhs/research/template\\_analysis/technique/dtemplate.htm](http://www.hud.ac.uk/hhs/research/template_analysis/technique/dtemplate.htm)
- King, N. (2008b). Frequently asked questions. *Template Analysis*. Retrieved 16.07.2008, from [http://www.hud.ac.uk/hhs/research/template\\_analysis/faq.htm](http://www.hud.ac.uk/hhs/research/template_analysis/faq.htm)
- King, N. (2008c). Interpretation. *Template Analysis*. Retrieved 16.07.2008, from [http://www.hud.ac.uk/hhs/research/template\\_analysis/technique/interpretation.htm](http://www.hud.ac.uk/hhs/research/template_analysis/technique/interpretation.htm)
- King, N. (2008d). Producing your initial template. *Template Analysis*. Retrieved 16.07.2008, from [http://www.hud.ac.uk/hhs/research/template\\_analysis/technique/itemplate.htm](http://www.hud.ac.uk/hhs/research/template_analysis/technique/itemplate.htm)
- King, N. (2008e). Quality checks. *Template Analysis*. Retrieved 16.07.2008, from [http://www.hud.ac.uk/hhs/research/template\\_analysis/technique/qualityreflexivity.htm](http://www.hud.ac.uk/hhs/research/template_analysis/technique/qualityreflexivity.htm)

- King, N. (2008f). The technique. *Template Analysis*. Retrieved 16.07.2008, from [http://www.hud.ac.uk/hhs/research/template\\_analysis/technique/technique.htm](http://www.hud.ac.uk/hhs/research/template_analysis/technique/technique.htm)
- King, N. (2008g). Themes and codes. *Template Analysis*. Retrieved 16.07.2008, from [http://www.hud.ac.uk/hhs/research/template\\_analysis/technique/themesandcodes.htm](http://www.hud.ac.uk/hhs/research/template_analysis/technique/themesandcodes.htm)
- King, N. (2008h). Transcription. *Template Analysis*. Retrieved 16.07.2008, from [http://www.hud.ac.uk/hhs/research/template\\_analysis/technique/transcription.htm](http://www.hud.ac.uk/hhs/research/template_analysis/technique/transcription.htm)
- King, N. (2008i). What is template analysis? *Template Analysis*. Retrieved 16.07.2008, from [http://www.hud.ac.uk/hhs/research/template\\_analysis/whatis.htm](http://www.hud.ac.uk/hhs/research/template_analysis/whatis.htm)
- King, N. (2008j). Writing up your study. *Template Analysis*. Retrieved 16.07.2008, from [http://www.hud.ac.uk/hhs/research/template\\_analysis/technique/writingup.htm](http://www.hud.ac.uk/hhs/research/template_analysis/technique/writingup.htm)
- Kirke- utdannings-og forskningsdepartementet (E. a. r. The Ministry of Church). (2001). *Gjør din plikt - krev din rett: kvalitetsreform av høyere utdanning : tilråding fra Kirke-, utdannings- og forskningsdepartementet av 9. mars 2001, godkjent i statsråd samme dag (Do your duty - demand your rights: the Quality Reform of Higher Education)*. Oslo: Departementet.
- Kirkhart, K. E. (2000). Reconceptualizing Evaluation Use: an integrated theory of influence. In V. J. Caracelli & H. Preskill (Eds.), *New Directions for Evaluation* (Vol. Winter, pp. 5 - 23). San Francisco, Calif.: Jossey-Bass.
- Kirkpatrick, D. L. (1971). *A practical guide for supervisory training and development*: Reading, Mass., etc.: Addison-Wesley Publishing Co.
- Kirkpatrick, D. L. (1998). *Evaluating training programs : the four levels* (2nd ed.). San Francisco, Calif.: Berrett-Koehler.
- Kogan, M. (2004). *Framework paper: the issues*. Paper presented at the UNESCO Forum: First regional research seminar for Europe and North America: Managerialism and evaluation in Higher Education. Paris, France 4-5 November, 2004. Retrieved 14.05.08, from <http://unesdoc.unesco.org/images/0014/001467/146741E.pdf>
- Kogan, M., Bauer, M., Bleiklie, I., & Henkel, M. (Eds.). (2000). *Transforming higher education: a comparative study*. London: Jessica Kingsley.
- Kraiger, K. (2002). Decision-based evaluation. In K. Kraiger (Ed.), *Creating, implementing, and managing effective training and development : state-of-the-art lessons for practice* (pp. xxxviii, 415 p.). San Francisco: Jossey-Bass ; [Chichester : Wiley [distributor]].
- Krogstrup, H. K. (2006). *Evalueringsmodeller* (2. udg. ed.). Århus: Academica.
- FOR 2005-09-08 nr 1040: Forskrift om akkreditering, evaluering og godkjenning etter lov om universiteter og høyskoler, (2005).
- Kunnskapsdepartementet (Ministry of Education and Research). (2008a). Innføring av skolelederutdanning for rektorer (Introduction of school leader education for head teachers). *KD-@ktuelt, 2008: 7* [www.regjeringen.no/nb/dep/kd/tidsskrift\\_nyhetsbrev](http://www.regjeringen.no/nb/dep/kd/tidsskrift_nyhetsbrev) accessed: 13.10.2008.

- Kunnskapsdepartementet (Ministry of Education and Research). (2008b). *Kvalitet i skolen (Quality in school)*. Oslo: Departementet.
- Kunnskapsdepartementet, & Utdannings- og forskningsdepartementet. (2005). *Rundskriv: Forskrift om akkreditering, evaluering og godkjenning etter lov om universiteter og høyskoler (Directive on the Regulations on accreditation, evaluation and approval in relation to the Universities' and University Colleges' Act)*. Retrieved from.
- Kushner, S. (2001). Culture, standards, and program qualities. In C. Lloyd, A. Benson & D. M. Hinn (Eds.), *Visions of quality : how evaluators define, understand and represent program quality* (Vol. 7, pp. 121-134). Greenwich, Conn.: JAI.
- Kuzel, A. J. (1999). Sampling in qualitative inquiry. In B. F. Crabtree & W. L. Miller (Eds.), *Doing qualitative research* (pp. XVII, 406 s.). Thousand Oaks, Calif.: Sage.
- Langdridge, D. (2004). *Introduction to research methods and data analysis in psychology*. Harlow: Pearson.
- Langdridge, D. (2007). *Phenomenological Psychology: Theory, Research and Method*. Harlow: Prentice-Hall.
- Langfeldt, L., & Hovdhaugen, E. (2006). *Kriterie- og metodegrunnlag for (re)akkreditering av høyere utdanning*. Oslo: NIFU STEP.
- Langley, A. (1999). Strategies for theorizing from process data *Academy of Management Review*, 24(4), 691-710.
- Langley, A., Mintzberg, H., Pitcher, P., Posada, E., & Saint-Macary, J. (1995). Opening up Decision Making: The View from the Black Stool *Organization Science*, 6(3), 260-279.
- Lave, J., & Wenger, E. (1991). *Situated learning : legitimate peripheral participation*. Cambridge: Cambridge University Press.
- Leithwood, K. A., Anderson, S. E., Mascall, B., & Strauss, T. (2010). School leaders' influences on student learning: the four paths. In T. Bush, L. Bell & D. Middlewood (Eds.), *The principles of educational leadership and management* (pp. XVI, 300 s.). Los Angeles: Sage.
- Leithwood, K. A., Day, C., Sammons, P., Harris, A., & Hopkins, D. (2006a). *Seven strong claims about successful school leadership*. London: Department for Education and Skills
- Leithwood, K. A., Day, C., Sammons, P., Harris, A., & Hopkins, D. (2006b). *Successful school leadership : what it is and how it influences pupil learning*. Nottingham: Department for Education and Skills.
- Leithwood, K. A., & Levin, B. (2004). *Approaches To The Evaluation Of Leadership Programs And Leadership Effects* London: DfES
- Leithwood, K. A., & Levin, B. (2005). *Assessing school leader and leadership programme effects on pupil learning : conceptual and methodological challenges*. Nottingham: DfES Publications.
- Leithwood, K. A., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004). *How leadership influences student learning*: University of Minnesota and University of Toronto
- Levačić, R. (2005). Educational Leadership as a Causal Factor: Methodological Issues in Research on Leadership 'Effects' *Educational Management Administration Leadership*, 33(2), 197-210.

- Levačić, R., & Glatter, R. (2003). Developing Evidence-Informed Policy and Practice in Educational Leadership and Management: a way forward. In N. Bennett & L. Anderson (Eds.), *Developing educational leadership : using evidence for policy and practice* (pp. XX, 188 s.). London: Sage.
- Lewis, M. W., & Grimes, A. J. (1999). Metatriangulation: Building theory from multiple paradigms *Academy of Management Review*, 24(4), 672-690.
- Lindblom, C. E. (1988). The science of muddling through. In C. E. Lindblom (Ed.), *Democracy and Market System*.
- Lindblom, C. E., & Cohen, D. K. (1979). *Usable knowledge : social science and social problem solving*. New Haven, Conn. : Yale University Press.
- Lipsett, A. (2007). Half of secondary schools must do better. *Guardian*. Retrieved 17.10.2007, from <http://www.guardian.co.uk/education/2007/oct/17/schools.uk2>
- Lipshitz, R., Klein, G., Orasanu, J., & Salas, E. (2001). Taking stock of naturalistic decision making *Journal of Behavioral Decision Making*, 14(5), 331-352.
- Love, A. J. (1991). *Internal evaluation : building organizations from within*. Newbury Park, Calif.: Sage.
- Love, A. J. (1998). Internal evaluation: integrating evaluation and social work practice *Scandinavian Journal of Social Welfare*, 7(2), 145-151.
- Lutz, F. W. (1982). Tightening up Loose Coupling in Organizations of Higher Education *Administrative Science Quarterly*, 27(4), 653-669.
- Lycke, K. H. (2004). Perspectives on quality assurance in higher education in Norway *Quality in Higher Education*, 10(3), 219-229.
- Læg Reid, P., Roness, P. G., & Rubecksen, K. (2006). Performance Management in Practice: the Norwegian Way *Financial Accountability & Management*, 22(3), 251-270.
- Læg Reid, P., Roness, P. G., Røvik, K. A., & Christensen, T. (2004). *Organisasjonsteori for offentlig sektor : instrument, kultur, myte [Organisation Theory for the public sector: Instrument, culture and myth]*. Oslo: Universitetsforl.
- MacBeath, J. (1998). *Effective school leadership: responding to change*. London: Sage Publ.
- MacDonald, B. (1976). Evaluation and the control of education. In D. Tawney (Ed.), *Curriculum evaluation today: trends and implications*. London: Macmillan.
- Madaus, G. F., & Kellaghan, T. (2000). Models, metaphors and definitions in evaluation. In G. F. Madaus, T. Kellaghan & D. L. Stufflebeam (Eds.), *Evaluation models : viewpoints on educational and human services evaluation* (2nd ed., pp. X, 509 s.). Boston: Kluwer Academic.
- Madill, A., & Gough, B. (2008). Qualitative research and its place in psychological science *Psychological Methods*, 13(3), 254-271.
- Madill, A., Jordan, A., & Shirley, C. (2000). Objectivity and reliability in qualitative analysis: Realist, contextualist and radical *British Journal of Psychology*, 91(1), 1.
- Maggioli, G. D. (2003). Fulfilling the promise of professional development [Electronic Version]. *IATEFL Voices* Oct/Nov 2003 number 175. Retrieved 18.08.08, from <http://www.iatefl.org/content/newsletter/174.php>



- Maitlis, S., & Lawrence, T. B. (2007). TRIGGERS AND ENABLERS OF SENSEGIVING IN ORGANIZATIONS *Academy of Management Journal*, 50(1), 57-84.
- Maitlis, S., & Ozcelik, H. (2004). Toxic Decision Processes: A Study of Emotion and Organizational Decision Making *Organization Science: A Journal of the Institute of Management Sciences*, 15(4), 373.
- March, J. G., Guetzkow, H., & Simon, H. A. (1958). *Organizations*. New York: Wiley.
- March, J. G., Guetzkow, H., & Simon, H. A. (1993). *Organizations* (2nd ed.). Cambridge, Mass.: Blackwell Business.
- March, J. G., & Heath, C. (1994). *A primer on decision making : how decisions happen*. New York: Free Press.
- March, J. G., Olsen, J. P., & Christensen, S. (1976). *Ambiguity and choice in organizations*. Bergen: Universitetsforlaget.
- March, J. G., Olsen, J. P., & Christensen, S. (Eds.). (1979). *Ambiguity and choice in organizations* (2nd ed.). Bergen: Universitetsforlaget.
- Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation : an integrated framework for understanding, guiding, and improving public and nonprofit policies and programs*. San Francisco: Jossey-Bass ; [Chichester : Wiley [distributor]].
- Marquand, D. (1997). *The new reckoning capitalism, states and citizens*. Cambridge: Polity Press.
- Marshall, C., & Rossman, G. B. (2006). *Designing qualitative research* (4th ed.). Thousand Oaks, Calif.: Sage Publications.
- Marshall, G. (1998). *A Dictionary of sociology* (2nd ed.). Oxford: Oxford University Press.
- Mathieu, J. E., Tannenbaum, S. I., & Salas, E. (1992). Influences of individual and situational characteristics on measures of training effectiveness *Academy of Management Journal*, 35(4), 828.
- Maxcy, S. J. (2003). Pragmatic threads in mixed methods research in the social sciences: the search for multiple modes of inquiry and the end of the philosophy of formalism. In C. Teddlie & A. Tashakkori (Eds.), *Handbook of mixed methods in social & behavioral research*. Thousand Oaks, California: Sage.
- Maxwell, J. A. (1996). *Qualitative research design : an interactive approach*. Thousand Oaks, Calif.: Sage Publications.
- Maxwell, J. A. (2002). Understanding and validity in qualitative research. In A. M. Huberman & M. B. Miles (Eds.), *The Qualitative researcher's companion* (pp. 37 - 64). Thousand Oaks, Calif.: Sage.
- Maynard-Moody, S., & McClintock, C. C. (1981). Square pegs in round holes: Program evaluation and organizational uncertainty *Policy Studies Journal*, 9(5), 644-666.
- McLaughlin, K., & Osborne, S. P. (2002). Current trends and future prospects of public management: a guide. In K. McLaughlin, S. P. Osborne & E. Ferlie (Eds.), *New public management: current trends and future prospects* (pp. XI, 355 s.). London: Routledge.

- McLaughlin, M. W. (1985). Implementation realities and evaluation design. In M. M. Mark & R. L. Shotland (Eds.), *Social Science and Social Policy*. Beverly Hills: Sage.
- Meyer, J. W. (1996). Otherhood: the promulgation and transmission of ideas in the modern organizational environment. In B. Czarniawska & G. Sevón (Eds.), *Translating organizational change* (pp. XII, 284 s.). Berlin: Walter de Gruyter.
- Meyer, J. W., & Rowan, B. (1977). Institutional organizations: formal structures as myth and ceremony *The American Journal of Sociology*, 83(2), 340-363.
- Meyer, J. W., & Rowan, B. (1992). The structure of educational organizations. In J. W. Meyer & W. R. Scott (Eds.), *Organizational environments: ritual and rationality* (pp. 302 s.). Newbury Park, Calif.: Sage.
- Meyer, J. W., & Scott, W. R. (1983). *Organizational environments: ritual and rationality*. Beverly Hills, Calif.: Sage.
- Meyer, J. W., & Scott, W. R. (1992). *Organizational environments: ritual and rationality*. Newbury Park, Calif.: Sage.
- Michelsen, S., & Aamodt, P. O. (2007). *Evaluering av Kvalitetsreformen: sluttrapport*. Oslo: Norges forskningsråd.
- Miles, M. B. (1979). Qualitative Data as an Attractive Nuisance: The Problem of Analysis *Administrative Science Quarterly*, 24(4), 590-601.
- Miller, G. (1997). Contextualizing texts: studying organizational texts. In G. Miller & R. Dingwall (Eds.), *Context and method in qualitative research* (pp. VI, 226 s.). London: Sage Publications.
- Mintzberg, H. (1979). *The structuring of organizations: a synthesis of the research*. Englewood Cliffs, N.J.: Prentice-Hall.
- Monsen, L., & Haug, P. (2004). *Evaluering som styringsverktøy (Evaluation as a governing tool)*.
- Muijs, D., Harris, A., Lumby, J., Morrison, M., & Sood, K. (2006). Leadership and leadership development in highly effective further education providers. Is there a relationship? *Journal of Further and Higher Education*, 30(1), 87 - 106.
- Mulford, B., & Silins, H. (2003). Leadership for Organisational Learning and Improved Student Outcomes--What Do We Know? *Cambridge Journal of Education*, 33(2), 175-195.
- Mulford, B., Silins, H., & Leithwood, K. (2004). *Educational leadership for organisational learning and improved student outcomes*. Dordrecht: Kluwer akademik publishers.
- Munro, E. (2004). The Impact of Audit on Social Work Practice *Br J Soc Work*, 34(8), 1075-1095.
- Musselin, C. (2002). Editorial *European Journal of Education*, 37(1), 1 - 6.
- Møller, J. (1997). Some moral dilemmas in educational management. In L. Kydd, M. Crawford & C. Riches (Eds.), *Professional development for educational management* (pp. VIII, 195 s.). Buckingham: Open University Press.
- Møller, J. (2004). *Lederidentiteter i skolen : posisjonering, forhandlinger og tilhørighet*. Oslo: Universitetsforl.
- Møller, J. (2006a). Det norske kasuset i lys av andre land (The Norwegian cases related to other countries). In J. Møller & O. L. Fuglestad (Eds.), *Ledelse i*

- anerkjente skoler* (Leadership in successful schools. pp. 296 s.). Oslo: Universitetsforl.
- Møller, J. (2006b). Hvilke svar gir forskning om god skoleledelse? (What answers does research offer about good school leadership?). In J. Møller & O. L. Fuglestad (Eds.), *Ledelse i anerkjente skoler* (Leadership in successful schools. pp. 296 s.). Oslo: Universitetsforl.
- Møller, J. (2006c). Nye utfordringer og nye kompetansekrav i utdanning av skoleledere (New challenges and demands for competence in the education of school leaders). In K. Sivesind, G. Skedsmo & G. Langfeldt (Eds.), *Utdanningsledelse* (Educational Leadership. pp. 271 s.). Oslo: Cappelen akademisk forl.
- Møller, J. (2008). *Scandinavian approaches to school leadership*. Paper presented at the Improving School Leadership OECD Conference. Copenhagen, 14 -15 April 2008.
- Maassen, P. A. M., & Olsen, J. P. (Eds.). (2007). *University dynamics and European integration*. Dordrecht: Springer.
- Nachmias, D. (Ed.). (1980). *The Practice of policy evaluation*. New York: St. Martin's Press.
- Nachmias, D., & Henry, G. T. (1980). The Utilization of evaluation Research: problems and prospects. In D. Nachmias (Ed.), *The Practice of policy evaluation*. New York: St. Martin's Press.
- Nasser, F., & Fresko, B. (2002). Faculty Views of Student Evaluation of College Teaching *Assessment & Evaluation in Higher Education*, 27(2), 187 - 198.
- NCSL, & DFES. (2004). *School leadership: end to end review of school leadership policy and delivery*. Retrieved 20.11.2004. from <http://www.ncsl.org.uk/mediastore/image2/etoe-review-final-report.pdf>.
- Neave, G. (1998). The Evaluative State Reconsidered *European Journal of Education*, 33(3), 265-284.
- Neave, G. (2004). *The Bologna Process and the evaluative state: a viticultural perspective*. Paper presented at the UNESCO Forum: First regional research seminar for Europe and North America: Managerialism and evaluation in Higher Education. Paris, France 4-5 November, 2004. Retrieved 14.05.08, from <http://unesdoc.unesco.org/images/0014/001467/146741E.pdf>
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, N.J.: Prentice-Hall.
- Newman, I., Ridenour, C. S., Newman, C., & DeMarco Jr, G. M. P. (2003). A typology of research purposes and its relationship to mixed methods. In C. Teddlie & A. Tashakkori (Eds.), *Handbook of mixed methods in social & behavioral research*. Thousand Oaks, California: Sage.
- Newton, J. (2000). Feeding the Beast or Improving Quality?: academics' perceptions of quality assurance and quality monitoring *Quality in Higher Education*, 6(2), 153-163.
- NOKUT. (2003). *Kriterier for evaluering av universiteters og høgskolers kvalitetssikringssystem for utdanningsvirksomheten* (Criteria for evaluating Universities' and University Colleges' quality assurance systems for educational activities ). Oslo: NOKUT Retrieved 22.03.07 from

[http://www.nokut.no/graphics/NOKUT/Artikkelbibliotek/Norsk\\_utdanning/Forskrifter\\_Kriterier\\_mm/kriterier\\_evaluering\\_revidert\\_250106.pdf](http://www.nokut.no/graphics/NOKUT/Artikkelbibliotek/Norsk_utdanning/Forskrifter_Kriterier_mm/kriterier_evaluering_revidert_250106.pdf).

- Norris, N. J., & Kushner, S. (2007). The New Public Management and evaluation. In S. Kushner & N. J. Norris (Eds.), *Dilemmas of engagement: evaluation and the new public management* (pp. X, 157 s.). Amsterdam: Elsevier JAI.
- OECD. (2005a). *Improving school leadership: rationale and plans for the activity*. Paris: OECD Retrieved 05.06.07 from <http://www.oecd.org/dataoecd/46/33/37126110.pdf>.
- OECD (2005b). *Teachers matter: attracting, developing and retaining effective teachers*. Paris: OECD.
- OECD. (2006). *Improving school leadership: guidelines for country participation. Design and Implementation plan for the activity*. Paris: OECD Retrieved 05.06.07 from <http://www.oecd.org/dataoecd/46/31/37126083.pdf>.
- OECD (2008). *Measuring improvements in learning outcomes: best practices to assess the value-added school*. Paris: OECD.
- OECD (Ed.). (1994). *Evaluation and the decision making process in higher education: French, German and Spanish experiences*. Paris: Organisation for Economic Co-operation and Development.
- Office of Public Services Reforms. (2002). *Better government services: executive agencies in the 21st century*. Retrieved 12.11.2007. from <http://www.pm.gov.uk/files/pdf/opsr-agenciesm.pdf>.
- OfSTED. (2004). *Making a difference: the impact of award-bearing in-service training on school improvement*. Retrieved 17.05.2006. from <http://www.ofsted.gov.uk/Ofsted-home/Publications-and-research/Browse-all-by/Education/Teachers-and-teacher-training/Continuing-professional-development/Making-a-difference-the-impact-of-award-bearing-in-service-training-on-school-improvement>.
- Oldroyd, D. (2003). Educational Leadership for results or for learning? Contrasting directions in times of transition. *Managing Global Transitions*, 1(1), 49-67.
- Oliver, C. (1992). The Antecedents of Deinstitutionalization *Organization Studies (Walter de Gruyter GmbH & Co. KG.)*, 13(4), 563.
- Olsen, J. P. (1979). University governance: non-participation as exclusion or choice. In J. G. March, J. P. Olsen & S. Christensen (Eds.), *Ambiguity and choice in organizations*. Bergen: Universitetsforlaget AS.
- Olsen, J. P. (1992). Analyzing institutional dynamics *Staatswissenschaften und Staatspraxis*, 2, 247-271.
- Olsen, J. P. (1996). Norway: Slow Learner - or Another Triumph of the Tortoise? In B. G. Peters & J. P. Olsen (Eds.), *Lessons from experience : experiential learning in administrative reforms in eight democracies*. Oslo: Scandinavian University Press.
- Olsen, J. P. (2003). *Citizens, public administration and the search for theoretical foundations*. Oslo: ARENA.
- Olsen, J. P., & Maassen, P. A. M. (2007). European debates on the knowledge institution: the modernization of the university at the European level. In P. A. M. Maassen & J. P. Olsen (Eds.), *University dynamics and European integration* (pp. XII, 245 s.). Dordrecht: Springer.

- Orasanu, J., & Connolly, T. (1993). The reinvention of decision making. In G. A. Klein, J. Orasanu, R. Calderwood & C. E. Zsombok (Eds.), *Decision making in action: models and methods*. Norwood, N.J.: Ablex Pub.
- Osborne, S. P. (2006). The New Public Governance? *Public Management Review*, 377-387.
- Osborne, S. P., & McLaughlin, K. (2002). NPM in Context. In K. McLaughlin, S. P. Osborne & E. Ferlie (Eds.), *New public management : current trends and future prospects* (pp. 7-14). London: Routledge.
- Ouston, J. (1998). Introduction *School Leadership & Management*, 18(3), 317 - 320.
- Ovrelid, R., & Bastoe, P.-Ø. (2002). Norway: Toward a results-oriented Government Administration. In R. C. Rist, R. Sandahl & J.-E. Furubo (Eds.), *International atlas of evaluation*. New Brunswick, N.J.: Transaction Publ.
- Owen, J. M. (2004). Evaluation forms: toward an inclusive framework for evaluation practice. In M. C. Alkin (Ed.), *Evaluation roots : tracing theorists' views and influences*. California: Thousand Oaks.
- Owen, J. M., & Rogers, P. J. (1999). *Program evaluation : forms and approaches* (International ed.). London: Sage.
- Owen, S. (2005). *Career stage Professional Development and the National Teacher Quality Agenda*. Paper presented at the AARE 2005. Parramatta, Australia. Retrieved 18.08.08, from <http://www.aare.edu.au/05pap/owe05069.pdf>
- Palumbo, D. J. (Ed.). (1987). *The Politics of program evaluation*. Newbury Park, Calif.: Sage.
- Palumbo, D. J., & Nachmias, D. (1983). The Preconditions for Successful Evaluation - Is There an Ideal Paradigm *Policy Sciences*, 16(1), 67-79.
- Paradeise, C., Reale, E., Goastellec, G., & Bleiklie, I. (2009). Universities steering between stories and history. In C. Paradeise, E. Reale, I. Bleiklie & E. Ferlie (Eds.), *University Governance: Western European Comparative Perspectives*. Dordrecht: Springer.
- Patterson, G. (2001). The Applicability of Institutional Goals to the University Organisation *Journal of Higher Education Policy & Management*, 23(2), 159-169.
- Patton, M. Q. (1997). *Utilization-focused evaluation : the new century text* (3rd ed.). Thousand Oaks, Calif.: Sage.
- Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). Thousand Oaks, Calif.: Sage Publications.
- Patton, M. Q. (2003). Utilization- Focused Evaluation. In T. Kellaghan & D. L. Stufflebeam (Eds.), *International Handbook of Educational Evaluation* (pp. 223-244). Dordrecht: Kluwer Academic Publishers.
- Performance and Innovation Unit. (2003). *Strengthening leadership in the public sector: A research study by the PIU*. Retrieved 12.11.2007. from <http://www.cabinetoffice.gov.uk/upload/assets/www.cabinetoffice.gov.uk/strategy/piu-leadership.pdf>.
- Peters, B. G. (2000). Institutional theory: problems and prospects [Electronic Version]. *Political Science Series*, 69 69. Retrieved 18.08.2008,
- Peterson, P. E. (1976). *School politics, Chicago style*. Chicago: University of Chicago Press.

- Pfeffer, J. (1981a). Management as Symbolic Action. In L. L. Cummings & B. M. Staw (Eds.), *Research in organizational behavior: an annual series of analytical essays and critical reviews* (Vol. 3, pp. 1 - 52). Greenwich, Conn.: JAI Press.
- Pfeffer, J. (1981b). *Power in organizations*. Marshfield, Mass.: Pitman.
- Pfeffer, J., & Salancik, G. R. (1974). Organizational Decision Making as a Political Process: The Case of a University Budget *Administrative Science Quarterly*, 19(2), 135.
- Pfeffer, J., & Sutton, R. I. (2006). *Hard facts, dangerous half-truths, and total nonsense : profiting from evidence-based management*. Boston, Mass.: Harvard Business School Press.
- Poland, B. D. (2003). Transcription quality. In J. F. Gubrium & J. A. Holstein (Eds.), *Inside interviewing new lenses, new concerns* (pp. 557 s.). Thousand Oaks, Calif.: Sage.
- Pollitt, C., & Bouckaert, G. (2000). *Public management reform: a comparative analysis*. Oxford: Oxford University Press.
- Pont, B., Nusche, D., & Moorman, H. (2008). *Improving School Leadership* (Vol. 1: Policy and practice). Paris: OECD.
- Potter, R., B., & Beach, L., Roy. (1994). Imperfect Information in Pre-choice Screening of Options *Organizational Behavior and Human Decision Processes*, 59(2), 313-329.
- Power, M. (1997). *The audit society : rituals of verification*. Oxford: Oxford University Press.
- Power, M. (2005). The theory of the audit explosion. In C. Pollitt, E. Ferlie & L. E. Lynn (Eds.), *The Oxford handbook of public management* (pp. XIII, 789 s.). Oxford: Oxford University Press.
- Preskill, H., Zuckerman, B., & Matthews, B. (2003). An Exploratory Study of Process Use: Findings and Implications for Future Research *The American Journal of Evaluation*, 24(4), 423.
- Prior, L. (2004). Doing things with documents. In D. Silverman (Ed.), *Qualitative research: theory, method and practice* (2nd ed., pp. 76 - 94). London: Sage.
- Ragin, C. C. (1994). *Constructing social research : the unity and diversity of method*. Thousand Oaks, Calif.: Pine Forge Press.
- Reichert, S. (2007). Looking back - looking forward: Quality Assurance and the Bologna Process. Paper presented at the 2nd European Quality Assurance Forum, Sapienza Universita di Roma, Italy. In A. Beso, L. Bollaert, B. Curvale, H. Toft-Jensen, L. Harvey, E. Helle, B. Maguire, A. Mikkola & A. Sursock (Eds.). (2008). *Implementing and using quality assurance: strategy and practice* European University Association.
- Rennie, D. L. (2009). Toward a meta-methodology of qualitative research. Unpublished Expansion of a keynote address presented at the annual meeting of the British Psychological Society, March 22, 2007. Toronto: York University.
- Ribbins, P., Bates, R., & Gunter, H. (2003). Reviewing research in education in Australia and the UK: evaluating the evaluations *Journal of Educational Administration*, 41(4), 423 - 444.

- Ribbins, P., & Gunter, H. (2003). Leadership studies in education: maps for EPPI reviews? In L. Anderson & N. Bennett (Eds.), *Developing Educational Leadership: using evidence for policy and practice*. London: Sage.
- Richards, L. (2004). *Validity and reliability? Yes! Doing it in Software*. Paper presented at the Strategies Conference. University of Durham, England.
- Richards, L. (2005). *Handling qualitative data : a practical guide*. London: Sage.
- Riley, D., & Mulford, B. (2007). England's National College for School Leadership: a model for leadership education? *Journal of Educational Administration*, 45(1), 80-98.
- Rist, R. C., Sandahl, R., & Furubo, J.-E. (2002). *International atlas of evaluation*. New Brunswick, N.J.: Transaction Publ.
- Robinson, V. M. J., Lloyd, C. A., & Rowe, K. J. (2008). The Impact of Leadership on Student Outcomes: An Analysis of the Differential Effects of Leadership Types *Educational Administration Quarterly*, 44(5), 635-674.
- Rogers, P. J., & Hough, G. (1995). Improving the effectiveness of evaluations: Making the link to organizational theory *Evaluation and Program Planning*, 18(4), 321.
- Rosa, M. J., Tavares, D., & Amaral, A. (2006). Institutional Consequences of Quality Assessment *Quality in Higher Education*, 12(2), 145-159.
- Rosati, J. A. (1981). Developing a systematic decision-making framework - Bureaucratic politics in perspective *World Politics*, 33(2), 234-252.
- Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation : a systematic approach* (7th ed.). Thousand Oaks, Calif.: Sage.
- Rowan, B., & Miskel, C. G. (1999). Institutional theory and the study of educational organizations. In J. Murphy & K. S. Louis (Eds.), *Handbook of research on educational administration: a project of the American Educational Research Association* (pp. XXVII, 548 s.). San Francisco: Jossey-Bass Publishers.
- Rowe, C. (1989). Analysing management decision-making: further thoughts after the Bradford Studies *Journal of Management Studies*, 26(1), 29.
- Rubin, H. J., & Rubin, I. (1995). *Qualitative interviewing : the art of hearing data*. Thousand Oaks, Calif.: Sage Publications.
- Rubin, H. J., & Rubin, I. (2005). *Qualitative interviewing : the art of hearing data* (2nd ed.). Thousand Oaks, Calif.: Sage Publications.
- Russ-Eft, D. F., Atwood, R., & Eggherman, T. (2002). Use and Non-use of Evaluation Results: Case Study of Environmental Influences in the Private Sector *American Journal of Evaluation*, 23(1), 19-31.
- Russ-Eft, D. F., & Preskill, H. S. (2001). *Evaluation in organizations : a systematic approach to enhancing learning, performance, and change*. Cambridge, Mass. : Perseus Pub.
- Røvik, K. A. (1998). *Moderne organisasjoner: trender i organisasjonstenkningen ved tusenårsskiftet*. Bergen-Sandviken: Fagbokforl.
- Røvik, K. A. (2002). 'The secrets of the winners: Management ideas that flow'. In K. Sahlin-Andersson & L. Engwall (Eds.), *The Expansion of management knowledge: carriers, flows, and sources* (pp. XVI, 373 s.). Stanford, Calif.: Stanford Business Books.

- Røvik, K. A. (2007). *Trender og translasjoner: ideer som former det 21. århundrets organisasjon (Trends and translations: ideas that formed the organisation in the 21st century)*. Oslo: Universitetsforl.
- Sahlin-Andersson, K. (1996). Imitating by editing success: the construction of organization fields. In B. Czarniawska & G. Sevón (Eds.), *Translating organizational change* (pp. 69 - 92). Berlin: Walter de Gruyter.
- Sanders, J. R. (2002). Presidential Address: On Mainstreaming Evaluation *American Journal of Evaluation*, 23(3), 253-259.
- Sayer, A. (2004). Foreword: why critical realism? In S. Fleetwood & S. Ackroyd (Eds.), *Critical realist applications in organisation and management studies* (pp. XV, 366 s.). London: Routledge.
- Schwartz, R., & Mayne, J. (2005). Assuring the quality of evaluative information: theory and practice *Evaluation and Program Planning*, 28(1), 1-14.
- Schwartz, R., & Rosen, B. (2004). The Politics of Evidence-Based Health Policy-Making *Public Money & Management*, 24(2), 121-127.
- Schwarz, S., & Westerheijden, D. F. (Eds.). (2004). *Accreditation and evaluation in the European higher education area*. Dordrecht ; London: Kluwer Academic.
- Scott, D. (2007). Resolving the quantitative - qualitative dilemma: a critical realist approach *International Journal of Research & Method in Education*, 30(1), 3 - 17.
- Scott, W. R. (1977). The effectiveness of organizational effectiveness studies. In P. S. Goodman & J. M. Pennings (Eds.), *New perspectives on organizational effectiveness* (pp. xvii,275 s.). San Francisco: Jossey-Bass.
- Scott, W. R. (1995). Accomplishments, continuing controversies, new directions. In W. R. Scott (Ed.), *Institutions and Organizations*. Thousand Oaks, CA.: Sage.
- Scott, W. R. (1995). *Institutions and organizations*. Thousand Oaks, Calif.: Sage.
- Scott, W. R. (2001). *Institutions and organizations* (2nd ed.). Thousand Oaks, Calif.: Sage Publications.
- Scott, W. R. (2003). *Organizations : rational, natural, and open systems* (5th ed.). Upper Saddle River, N.J.: Prentice Hall : Pearson Education International.
- Scott, W. R. (2006). Ad astra per aspera: A journey from the periphery *Organization Studies*, 27(6), 877-897.
- Scott, W. R., & Davis, G. F. (2007). *Organizations and organizing: rational, natural, and open system perspectives*. Upper Saddle River, N.J.: Pearson Education.
- Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park, Calif.: Sage.
- Scriven, M. (2003). Evaluation theory and metatheory. In T. Kellaghan & D. L. Stufflebeam (Eds.), *International handbook of educational evaluation*. Dordrecht: Kluwer Academic Publ.
- Sebba, J. (2003). A Government strategy for research and development in Education. In L. Anderson & N. Bennett (Eds.), *Developing educational leadership : using evidence for policy and practice* (pp. XX, 188 s.). London: Sage.
- Sergiovanni, T. (1979). Rational, Bureaucratic, Collegial, and Political Views of the Principal's Role *Theory Into Practice*, 18(1), 12.



- Shadish, W. R., & Epstein, R. (1987). Patterns of Program Evaluation Practice Among Members of the Evaluation Research Society and Evaluation Network *Eval Rev*, 11(5), 555-590.
- Shadish, W. R., & Leviton, L. C. (2001). Descriptive values and social justice. In A. Benson, D. M. Hinn & C. Lloyd (Eds.), *Visions of quality : how evaluators define, understand and represent program quality. Advances in program evaluation* (Vol. 7, pp. 181-200). Greenwich, Conn.: JAI.
- Shapiro, J. Z. (1984). Conceptualizing evaluation use: Implications of alternative models of organizational decision-making. In R. F. Conner, D. G. Altman & C. Jackson (Eds.), *Evaluation studies review annual* (Vol. 9, pp. 633-645). Beverly Hills, Calif., Sage Publications.
- Shaw, I. (1999). *Qualitative evaluation*. London: Sage.
- Simkins, T. (2005a). Leadership in Education: 'What Works' or 'What Makes Sense'? *Educational Management Administration Leadership*, 33(1), 9-26.
- Simkins, T. (2005b). Matching methods to purpose *Management in Education*, 19(2), 35-36.
- Simkins, T., Coldwell, M., Close, P., & Morgan, A. (2007). Outcomes of In-school Leadership Development Work: A Study of Three NCSL Programmes. Sheffield: Sheffield Hallam University.
- Simkins, T., Coldwell, M., Close, P., & Morgan, A. (2009). Outcomes of In-school Leadership Development Work: A Study of Three NCSL Programmes *Educational Management Administration Leadership*, 37(1), 29-50.
- Simon, H. A. (1964). On the Concept of Organizational Goal *Administrative Science Quarterly*, 9(1), 1.
- Simon, H. A. (1976). *Administrative behavior a study of decision-making processes in administrative organization* (3rd ed.). New York: Free Press.
- Simon, H. A. (1993). Decision Making: Rational, Nonrational, and Irrational *Educational Administration Quarterly*, 29(3), 392-411.
- Simon, H. A. (1997). *Administrative behavior : a study of decision-making processes in administrative organization* (4th ed.). New York: Simon & Schuster : Free Press.
- Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment *Organizational Behavior and Human Performance*, 6(6), 649-744.
- Smedstad, T. (2007). Nasjonal skoleleder utdanning? (National training for school leaders?). *Skolelederen (The school leader)*, 7(6), 6-7.
- Smith, J. A., & Osborn, M. (2008). Interpretive Phenomenological Analysis. In J. A. Smith (Ed.), *Qualitative psychology: a practical guide to research methods* (pp. XI, 276 s.). London: Sage.
- Sonnichsen, R. C. (2000). *High impact internal evaluation : a practitioner's guide to evaluating and consulting inside organizations*. Thousand Oaks, Calif.: Sage Publications.
- Southworth, G. (2004). A Response from the National College for School Leadership *Educational Management Administration Leadership*, 32(3), 339-354.
- Sparks, D. (2002). *Designing Powerful Professional Development for Teachers and Principals. Access ERIC: FullText* (055 Guides--Nonclassroom). Ohio: National Staff Development Council, Oxford, OH.

- Stake, R. E. (1990). Situational context as influence on evaluation design and use *Studies In Educational Evaluation*, 16(2), 231-246.
- Stame, N. (2006). Governance, Democracy and Evaluation *Evaluation*, 12(1), 7-16.
- Stensaker, B. (2000). *Høyere utdanning i endring: dokumentasjon og drøfting av kvalitetsutviklingstiltak ved seks norske universiteter og høyskoler 1989-1999*. Oslo: Norsk institutt for studier av forskning og utdanning.
- Stensaker, B. (2003). Trance, Transparency and Transformation: the impact of external quality monitoring on higher education *Quality in Higher Education*, 9(2), 151-159.
- Stensaker, B. (2004). The blurring boundaries between accreditation and audit: the case of Norway. In S. Schwarz & D. F. Westerheijden (Eds.), *Accreditation and evaluation in the European higher education area* (Vol. Higher education dynamics ; v. 5, pp. 347 - 369). Dordrecht ; London: Kluwer Academic.
- Stensaker, B. (2006a). *Evaluering av kvalitetsreformen Delrapport 2: Institusjonelle kvalitetssystemer i høyere utdanning - vil de bidra til bedre kvalitet? (Evaluation of the Quality Reform Sub-report 2: Institutional quality assurance systems in Higher Education - will they contribute to improved quality?)*. Oslo: NIFU STEP.
- Stensaker, B. (2006b). *Evaluering av kvalitetsreformen Delrapport 5: Kvalitet som forhandling: NOKUT i norsk høyere utdanning 2003-2006 (Evaluation of the Quality Reform sub-report 5: Quality as bargaining: NOKUT in Norwegian Higher Education 2003-2006)*. Oslo: NIFU STEP.
- Stensaker, B., Rosa, M. J., & Westerheijden, D. F. (2007). Conclusions and further challenges. In D. F. Westerheijden, B. Stensaker & M. J. Rosa (Eds.), *Quality Assurance in Higher Education: trends in regulation, translation and transformation* (Vol. 13, pp. 295 - 312).
- Stockdill, S. H., Baizerman, M., & Compton, D. W. (2002). Toward a definition of the ECB process: A conversation with the ECB literature *New Directions for Evaluation*, 93 (April 2002), 7-26.
- Strauss, A. L., & Corbin, J. M. (1998). *Basics of qualitative research: techniques and procedures for developing grounded theory*. Thousand Oaks, Calif.: Sage Publications.
- Stufflebeam, D. L. (2001). Evaluation models *New directions for evaluation*, 89 (April 2001), 106 p.
- Stufflebeam, D. L., Foley, W. J., Gephart, W. J., Guba, E. E., Hammond, R. L., Merriman, H. O., et al. (1971). *Educational evaluation and decision making*. Itasca, Ill.: Peacock.
- Stufflebeam, D. L., & Webster, W. J. (1983). An analysis of alternative approaches to evaluation. In G. F. Madaus, M. Scriven & D. L. Stufflebeam (Eds.), *Evaluation models : viewpoints on educational and human services evaluation* (pp. xiii, 423 s.). Boston: Kluwer-Nijhoff.
- Sverdrup, S. (2002). *Evaluering : faser, design og gjennomføring (Evaluation: phases, design and implementation)*. Bergen: Fagbokforl.
- Søgnen, A. (2003). *NOU 2003:16. I første rekke: forsterket kvalitet i en grunnopplæring for alle (In the front row: strengthening quality within an education for all)*. Retrieved 31.05.07. from

<http://www.regjeringen.no/Rpub/NOU/20032003/016/PDFS/NOU200320030016000DDDPDFS.pdf>

- Sørensen, A. B. (1994). *Evaluering som middel til Organisationsindlæring [Evaluation as a means to Organisational Learning]*. Paper presented at the Evaluering 90 årenes mirakelkur? [Evaluation, the 1990s miracle cure?]. Oslo.
- Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology : combining qualitative and quantitative approaches* (Vol. 46). Thousand Oaks, California: Sage.
- Taylor, D. (2005). Governing Through Evidence: Participation and Power in Policy Evaluation *Journal of Social Policy*, 34(4), 601-618.
- Taylor, F. W. (1911). *The Principles of scientific management*. New York: Harper & Brothers.
- TDA. (2007). *Impact evaluation of CPD*. Retrieved 17.03.2009. from [http://www.tda.gov.uk/teachers/continuingprofessionaldevelopment/cpdleadership/evaluate\\_impact/how\\_to.aspx](http://www.tda.gov.uk/teachers/continuingprofessionaldevelopment/cpdleadership/evaluate_impact/how_to.aspx).
- Teddlie, C., & Tashakkori, A. (2003). *Handbook of mixed methods in social & behavioral research*. Thousand Oaks, Calif.: SAGE.
- Telhaug, A. O., Mediås, O. A., & Aasen, P. (2006). The Nordic model in education: education as part of the political system in the last 50 years *Scandinavian Journal of Educational Research*, 50(3), 245 - 283.
- The Quality Assurance Agency for Higher Education. (2003). *A brief guide to quality assurance in UK higher education*. Retrieved 15.09.2007. from [www.qaa.ac.uk](http://www.qaa.ac.uk).
- Thompson, J. D. (1967). *Organizations in action: social science bases of administrative theory*. New York: McGraw-Hill.
- Thompson, J. D. (2003). *Organizations in action: social science bases of administrative theory*. New Brunswick, N.J.: Transaction Publishers.
- Thompson, J. D., & Tuden, A. (1987). Strategies, structures, and processes of organizational decision. In J. D. Thompson & J. Woodward (Eds.), *Comparative studies in administration* (pp. 195 - 216). New York: Garland.
- Tiplic, D. (2008). *Managing organizational change during institutional upheaval: Bosnia-Herzegovina's higher education in transition*, Handelshøyskolen BI, Oslo.
- Tjeldvoll, A., Wales, C. D., & Welle-Strand, A. (2005). School Leadership Training under Globalisation: Comparisons of the UK, the US and Norway. *Managing Global Transitions*, 3(1), 23-49.
- Tourmen, C. (2009). Evaluators' Decision Making: The Relationship Between Theory, Practice, and Experience *American Journal of Evaluation*, 30(1), 7-30.
- Turner, S. P. (2006). The Philosophy of the Social Sciences in Organizational Studies. In S. R. Clegg, C. Hardy, T. B. Lawrence & W. R. Nord (Eds.), *The SAGE handbook of organization studies* (pp. XXII, 895 s.). London: SAGE.
- UK Treasury (2003). *Every child matters*. London: TSO.
- Urwick, L., & Gulick, L. (1947). *Papers on the science of administration* (2nd ed.). New York: Institute of Public Administration.
- Utdannings og forskningsdepartementet (2004a). *Kompetanse for utvikling : strategi for kompetanseutvikling i grunnsopplæringa 2005-2008 (Competence for*

- development: a strategy for capacity building in compulsory education 2005-2008* ([Nynorskutg.] ed.). [Oslo]: Utdannings- og forskningsdepartementet.
- Utdannings og forskningsdepartementet (Ministry of Education and Research). (2004b). *Kultur for læring (Culture for Learning)*. [Oslo]: Departementet.
- Utdanningsdirektoratet (Norwegian Directorate for Education and Training). (2007). *Improving School Leadership: Country background report for Norway*. Retrieved 05.06.07. from [http://udir.no/templates/udir/TM\\_Artikkel.aspx?id=2552](http://udir.no/templates/udir/TM_Artikkel.aspx?id=2552).
- Utdanningsdirektoratet (Norwegian Directorate for Education and Training). (2008a). *Kompetanse for en rektor - forventninger og krav. Grunnlagsdokument og utgangspunkt for anbudsutlysning (Competence for a head teacher - expectations and demands. The basis for a tendered bid.)*. Retrieved 19.03.2009. from [http://www.utdanningsdirektoratet.no/upload/skoleutvikling/kompetansekrav\\_rektorer.pdf](http://www.utdanningsdirektoratet.no/upload/skoleutvikling/kompetansekrav_rektorer.pdf).
- Utdanningsdirektoratet (Norwegian Directorate for Education and Training). (2008b). *KONKURRANSEGRUNNLAG: "Konkurranse om nasjonalt utdanningstilbud for rektorer" (Competition basis: Competition for national educational initiative for head teachers)*. Retrieved 19.03.2009. from [http://www.utdanningsdirektoratet.no/upload/skoleutvikling/konkurranse\\_rektorskolen.pdf](http://www.utdanningsdirektoratet.no/upload/skoleutvikling/konkurranse_rektorskolen.pdf).
- Valovirta, V. (2002). Evaluation Utilization as Argumentation *Evaluation*, 8(1), 60-80.
- Van de Ven, A. H., & Astley, W. G. (1981). Mapping the field to create a dynamic perspective on organization design and behavior. In A. H. Van de Ven & W. F. Joyce (Eds.), *Perspectives on organization design and behavior* (pp. xxiii, 486). New York : Wiley.
- Van der Meer, F.-B. (2007). New Public Management and Evaluation. In C. Pollitt, S. van Thiel & V. Homburg (Eds.), *New public management in Europe: adaptations and alternatives* (pp. XII, 229 s.). Houndmills, Basingstoke: Palgrave Macmillan.
- Van Maanen, J. (1979a). The Fact of Fiction in Organizational Ethnography *Administrative Science Quarterly*, 24(4), 539-550.
- Van Maanen, J. (1979b). Reclaiming Qualitative Methods for Organizational Research: A Preface *Administrative Science Quarterly*, 24(4), 520-526.
- Van Maanen, J., & Barley, S. R. (1984). Occupational communities: culture and control in organizations. In B. M. Staw & L. L. Cummings (Eds.), *Research in organizational behavior: an annual series of analytical essays and critical reviews* (Vol. 6, pp. 287-365). Greenwich, Conn.: JAI Press.
- Van Maanen, J., Sørensen, J. B., & Mitchell, T. R. (2007). The interplay between theory and method *Academy of Management Review*, 32(4), 1145-1154.
- van Thiel, S., & Leeuw, F. L. (2002). The performance paradox in the public sector *Public Performance & Management Review*, 25(3), 267 - 281.
- van Vught, F. A., & Westerheijden, D. F. (1993). *Quality measurement and quality assurance in European higher education*. Enschede : University of Twente,

- Center for Higher Education Policy Studies.: Universiteit Twente.; Center for Higher Education Policy Studies.
- Vedung, E. (1994). *Utvärdering i offentliga sektorn : problem och lösningar (Evaluation in the public sector: the problem and solutions)*. Stockholm: Fritzes.
- Vedung, E. (1997). *Public policy and program evaluation*. New Brunswick, N.J.: Transaction Publishers.
- Vedung, E. (1998). *Utvärdering i politik och förvaltning (Evaluation of politics and administration)* (2. uppl. ed.). Lund: Studentlitteratur.
- Vedung, E. (2000). *Utvärdering, demokrati och reflexiv modernitet (Evaluation, democracy and reflexive modernity)*. Gävle: Uppsala universitet, Institutet för bostadsforskning.
- Vedung, E. (2003). Models of evaluation. In P. Haug, T. A. Schwandt & Norges forskningsråd (Eds.), *Evaluating educational reforms : Scandinavian perspectives* (pp. 41-71). Greenwich, Conn.: Information Age Publ.
- Vedung, E. (2006a). Evaluation research. In B. G. Peters & J. Pierre (Eds.), *Handbook of public policy* (pp. XI, 512 s.). London: Sage.
- Vedung, E. (2006b). Utvärdering som megatrend, gigatrend och fyre böljor (Evaluation as mega trend, giga trend and four waves). In H. F. Hansen (Ed.), *Den organiserede forvaltning: politik, viden og værdier i samspil (The organised administration: politics, knowledge and values in interaction.)*. Copenhagen: Forlaget Politiske Studier.
- Vroeijenstijn, A. I. (1995). *Improvement and accountability navigating between Scylla and Charybdis : guide for external quality assessment in higher education*. London: Jessica Kingsley.
- Wales, C. D. (2004). *Same problems, different solutions? : developments in school leadership training in Wales and Norway*. Oslo: C.D. Wales.
- Wales, C. D., & Welle-Strand, A. (2005). *School Management Training, Country Report: Norway* (No. 16). Oslo: BI Norwegian School of Management
- Wales, C. D., & Welle-Strand, A. (2008). *School Management Training, Country Report 2007: Norway* (No. 29). Oslo: BI Norwegian School of Management
- Wallace, M. (2001). 'Really Good Ideas?': Developing Evidence-Informed Policy and Practice in Educational Leadership and Management: A 'Rejoinder' to Rosalind Levacic and Ron Glatter *Educational Management Administration Leadership*, 29(1), 27-34.
- Weber, M. (1947). *The theory of social and economic organization*. New York: The Free Press of Glencoe.
- Weick, K. E. (1976). Educational Organizations as Loosely Coupled Systems *Administrative Science Quarterly*, 21(1), 1.
- Weick, K. E. (1989). Loose coupling: beyond the metaphor *Current Contents (citation classic)*, 21(12), 14.
- Weick, K. E. (1995). *Sensemaking in organizations*. Thousand Oaks, Calif.: Sage.
- Weick, K. E. (2001). Sources of order in underorganized systems: themes in recent organizational theory. In K. E. Weick (Ed.), *Making sense of the organization* (pp. XII, 483 s.). Oxford: Blackwell.
- Weick, K. E. (2003). Enacting an environment: the infrastructure of organizing. In R. I. Westwood & S. Clegg (Eds.), *Debating organization: point-*

- counterpoint in organization studies* (pp. 184 - 194). Malden, Mass.: Blackwell Publishing.
- Weick, K. E., & McDaniel, R. R. (1989). How professional organizations work: implications for school organization and management. In T. J. Sergiovanni & J. H. Moore (Eds.), *Schooling for tomorrow: directing reforms to issues that count* (pp. xxii, 403 s.). Boston: Allyn and Bacon.
- Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the Process of Sensemaking *Organization Science*, 16(4), 409-421.
- Weiss, C. H. (1972). Utilization of Evaluation: Toward Comparative Study. In C. H. Weiss (Ed.), *Evaluating action programs : readings in social action and education*. Boston: Allyn and Bacon.
- Weiss, C. H. (1979). The many meanings of research utilization *Public Administration Review*, 39(5).
- Weiss, C. H. (1982). Policy Research in the Context of Diffuse Decision Making *Journal of Higher Education*, 53(6), 619-639.
- Weiss, C. H. (1987). Where Politics and Evaluation research meet. In D. J. Palumbo (Ed.), *The Politics of program evaluation*. Newbury Park, Calif.: Sage.
- Weiss, C. H. (1998a). *Evaluation : methods for studying programs and policies* (2nd ed.). Upper Saddle River, N.J.: Prentice Hall.
- Weiss, C. H. (1998b). Have we learned anything new about the use of evaluation? *The American Journal of Evaluation*, 19(1), 21.
- Weiss, C. H. (2004). Making a difference *Contexts: Understanding People in Their Social Worlds* 3(2), 70-71.
- Weiss, C. H., Murphy-Graham, E., & Birkeland, S. (2005). An Alternate Route to Policy Influence: How Evaluations Affect D.A.R.E *American Journal of Evaluation*, 26(1), 12-30.
- Welch, C., & Piekkari, R. (2006). Crossing language boundaries: Qualitative interviewing in international business *Management International Review*, 46(4), 417 - 437.
- Westerheijden, D. F. (2007). States and Europe and quality of higher education. In D. F. Westerheijden, B. Stensaker & M. J. Rosa (Eds.), *Quality assurance in higher education: trends in regulation, translation and transformation* (pp. xii, 262 s.). Dordrecht: Springer.
- Westerheijden, D. F., Hulpiau, V., & Waeytens, K. (2007). From Design and Implementation to Impact of Quality Assurance: An Overview of Some Studies into what Impacts Improvement *Tertiary Education and Management*, 13(4), 295 - 312.
- Westerheijden, D. F., Stensaker, B., & Rosa, M. J. (2007). Introduction. In D. F. Westerheijden, B. Stensaker & M. J. Rosa (Eds.), *Quality assurance in higher education: trends in regulation, translation and transformation* (pp. xii, 262 s.). Dordrecht: Springer.
- Willig, C. (2001). *Introducing qualitative research in psychology adventures in theory and method*. Buckingham: Open University Press.
- Witt, L. A., Andrews, M. C., & Kacmar, K. M. (2000). The role of participation in decision-making in the organizational politics-job satisfaction relationship *Human Relations*, 53(3), 341-358.

- Witziers, B., Bosker, R. J., & Kruger, M. L. (2003). Educational Leadership and Student Achievement: The Elusive Search for an Association *Educational Administration Quarterly*, 39(3), 398-425.
- Wolcott, H. F. (2001). *Writing up qualitative research*. Thousand Oaks: Sage Publications.
- Wollmann, H. (2003). *Evaluation in public-sector reform: concepts and practice in international perspective*. Cheltenham: Edward Elgar.
- Wright, N., & Colquhoun, D. (2005). Making evaluation count *Management in Education*, 19(2), 33-37.
- Yardley, L. (2008). Demonstrating validity in qualitative psychology. In J. A. Smith (Ed.), *Qualitative psychology: a practical guide to research methods* (pp. XI, 276 s.). London: Sage.
- Zucker, L. G. (1991). Institutionalization and cultural persistence- postscript: Microfoundations of Institutional thought. In P. J. DiMaggio & W. W. Powell (Eds.), *The New institutionalism in organizational analysis* (pp. VII, 478 s.). Chicago: University of Chicago Press.
- Øvretveit, J. (2005). Public service quality improvement. In C. Pollitt, E. Ferlie & L. E. Lynn (Eds.), *The Oxford handbook of public management* (pp. XIII, 789 s.). Oxford: Oxford University Press.
- Aarebrot, F. (2006). *Skole, politikk og ledelse (School, politics and leadership)*, Presentation to Sognal Union of Education Norway, 20th December 2006. Retrieved 27.02.2009 from <http://www.utdanningsforbundet.no/upload/Fylkeslag/SognogFjordane/Lokallag/Sogndal/Sogndal%20-%20lærere%20-%20desember%20-%202006.ppt>