

This file was downloaded from the institutional repository BI Brage - <http://brage.bibsys.no/bi>
(Open Access)

***Public expenditures, educational outcomes and grade inflation:
theory and evidence from a policy intervention in the Netherlands***

**Kristof De Witte
Maastricht University**

**Benny Geys
BI Norwegian Business School
Vrije Universiteit Brussel**

**Catharina Solondz
Technical University of Dresden**

This is the authors' accepted and refereed manuscript to the article published in

Economics of Education Review, 40(2014): 152-166

DOI: <http://dx.doi.org/10.1016/j.econedurev.2014.02.003>

The publisher, Elsevier, allows the author to retain rights to “post a revised personal version of the text of the final journal article (to reflect changes made in the peer review process) on your personal or institutional website or server for scholarly purposes, incorporating the complete citation and with a link to the Digital Object Identifier (DOI) of the article”. The final publication is available at www.elsevier.com (Publisher's policy 2013).

**Public Expenditures, Educational Outcomes and Grade Inflation:
Theory and Evidence from a Policy Intervention in the Netherlands***

Kristof De Witte ^{a, b}, Benny Geys ^{c, d} and Catharina Solondz ^e

^a Top institute for Evidence Based Education Research, Maastricht University, Kapoenstraat 2, 6200 MD Maastricht, the Netherlands, Email: k.dewitte@maastrichtuniversity.nl

^b University of Leuven (KU Leuven), Naamsestraat 69, B-3000 Leuven, Belgium, phone: 0032 16 32 66 56; Email: kristof.dewitte@econ.kuleuven.be.

^c Norwegian Business School (BI), Nydalsveien 37, N-0442 Oslo, Norway, email: Benny.Geys@bi.no

^d Vrije Universiteit Brussel (VUB), Department of Applied Economics, Pleinlaan 2, B-1050 Brussels, Belgium, email: Benny.Geys@vub.ac.be.

^e Technical University of Dresden, Chair of Economic Policy and Economic Research, D-01062 Dresden, Germany, Email: catharina.solondz@mailbox.tu-dresden.de

Abstract:

This article argues that resource expansion can fail to improve actual student performance because it might cause educators to soften grading standards (i.e., induce grade inflation). Our theoretical model shows that, depending on schools' and students' reactions to resource changes, the overall effect of resources on education outcomes is ambiguous. Schools, however, have an incentive to adjust their grading structure following resource shifts, such that grade inflation is likely to accompany resource-driven policies. Exploiting a quasi-experimental policy intervention in the Netherlands, we find that additional resources may indeed induce grade inflation, particularly when the resource increase is limited.

Keywords: Public expenditures; Grade inflation; Educational attainment; Standardized central exam.

JEL-Codes: I20, I28, H52.

* We would like to thank two anonymous referees, Sofie Cabus, Julie Cullen, Stefano Della Vigna, Dennis Epple, Wim Groot, Jaap Dronkers, Bruno Heyndels, Katharina Hilken, Daniel Horn, Alexander Kemnitz, Kai Konrad, Ronnie Schoeb, Bart Schoenmakers, John Karl Scholz, Marcel Thum, Liesbeth van Welie as well as participants of research seminars at TU Dresden, Vrije Universiteit Brussel, the "Educational Governance and Finance" Workshop (Oslo, Norway), the 2013 European Economic Association Meeting and the 2013 European Association of Labor Economists Meeting for valuable comments and discussions. The usual caveat applies. Benny Geys is grateful to FWO Vlaanderen (grant nr. G.0022.12) for financial support.

1. Introduction

The question whether or not resource-driven policies are effective in increasing educational quality and student performance remains vigorously debated (for reviews, see Hanushek, 2003; Hægeland *et al.*, 2012). Most studies in this vast literature analyze school-level exam results, rather than standardized central exit exams or SAT scores. In this article, we argue that this choice of evaluation standard is not innocuous. The reason is that schools can affect observed student performance through their choice of grading standard, which not only translate students' performance into a given grade, but also affect learning effort (Correa and Gruver, 1987; Costrell, 1994; Bonesrønning, 2004; Figlio and Lucas, 2004). This suggests that resource-driven policies may have both a *direct* effect on student performance (extensively discussed in the foregoing literature), and an *indirect* one via schools' endogenous grading structure decisions (disregarded in earlier work). The pressure on schools receiving more resources to show improved outcomes might indeed induce them to 'game' the system and 'generate' better achievements by inflating their grades.

We first set up a simple theoretical framework in which students choose their learning effort depending on grading standards, and schools use their grading policy to influence students' behavior (Costrell, 1994; Betts, 1998; Correa and Gruver, 1987; Bonesrønning, 1999). Two innovations are brought to this model. First, by explicitly incorporating both a national assessment conducted with a uniform correction model (referred to as the 'central exam') and an assessment developed and graded by each school's teachers (referred to as the 'school exam'), we assess how educational spending affects both types of evaluation standards. Second, educational spending is introduced into the objective functions of students and schools, which allows investigating how students' effort choice as well as the school's grading standard depend on education expenditures. Although existing work has studied the *direct*

effect of education expenditures, as well as the reaction of student effort to grading standards, our contribution here lies in connecting these two elements via the school's reaction to expenditure changes. This illustrates that *indirect* behavioral feedback effects of increased education expenditures on schools and students are probable, and that these may counteract the intended outcome of higher spending. Specifically, schools are shown to have an incentive to adjust their grading standard when resources change, suggesting that grade inflation (i.e., assigning higher grades than before for similar performance or similar grades for deteriorating performance) following an increase in resources is a realistic possibility.

Then, we evaluate the key implications of the model by exploiting a recent policy intervention in the Netherlands, which features two crucial characteristics. First, it created a quasi-experimental setting where 40 districts in 18 cities received substantial additional block grants from the Dutch central government totaling 250 million euro per year (while other, often quite similar, districts received no such funding). These additional funds were *earmarked* for investments in social policies such as education as of summer 2007, and the responsible minister explicitly made the improvement of educational outcomes one of the core aims of the program (Tweede Kamer, 2008-2009). Second, pupils' school-leaving test results in the Dutch education system are determined by both standardized national exit exams and school exams.¹ Since schools only have discretion over the grading standard in the school exam,² we can, like Wikström and Wikström (2005), employ the results of the central exam as a benchmark (uniformly applied to all pupils in all schools) against which to set the school exam results.

¹ The Dutch education system is not unique in its reliance on multiple performance measures. In the US as well as Sweden, both SAT scores and the student's Grade Point Average matter for college admission applications. In Italy and France, universities often organise their own entry test, but nonetheless take the school grade into account. The final grade of the German 'Abitur' (higher secondary degree) incorporates results from both state-level central exams and school-level exams collected during the last two years of schooling.

² Substantial checks and balances in the Dutch system are explicitly geared towards guaranteeing a constant central exam grading policy over time (see below).

This setting allows for a difference-in-differences (DiD) identification strategy whereby Dutch schools inside/outside the selected districts are compared over the 2004-2006 period before the intervention and the 2008-2009 period after the intervention (see also Gerritsen and Webbink, 2010; Wittebrood and Permentier, 2011). Our findings show that, on average, there is a *decline* in central exam results, but an (insignificant) relative *improvement* in school exams, in schools located in districts with additional funding. Accounting for the varying size of the investment program across districts (ranging from €1.2 million to €29.3 million, or €333 to €995 per resident annually), higher investment is found to significantly dampen the relative decline in central-level exam results, while leaving school exams unaffected. Hence, increased resources seem to have positively affected central exam results when additional funds were sufficiently elevated, but induced grade inflation when funds were limited (i.e., under approximately €1250 per resident). These findings are robust to the level of analysis (i.e. schools or districts), different specifications of the control group and the implementation of a matching estimator exploiting the purposeful assignment to the treatment.

In the next two sections, we briefly review the existing literature and provide a simple theoretical model linking public expenditures to education outcomes and incentives for grade inflation. Then, in section 4, we discuss the institutional setting and the dataset. Section 5 contains our methodological approach and empirical results. Finally, section 6 provides a concluding discussion.

2. Literature review

The results of studies analyzing whether resource-driven policies increase schooling quality and student performance are, at best, ambiguous (for a review, see Hanushek, 2003; more recent contributions include Holmlund *et al.*, 2010; Hægeland *et al.*, 2012). Hoxby (2000)

argues that this ambiguity may well derive from differing objective functions of teachers, schools or public authorities. Another reason, however, may be that exam systems differ widely. In some systems, exams and grading standards are set by schools, while in others central standards or exams are set. The latter clearly limits the opportunity for teachers and/or schools to affect the grading structure when resources are increased and policy-makers expect students' achievements to improve accordingly.³ A change in education spending may therefore have a different *observed* impact (in terms of exam results) depending on the exam system at hand.

While the role of grading standards in the resources-achievement relation has, to the best of our knowledge, not been addressed, three related literatures suggest that this may be an important oversight. The first investigates how grading standards affect students' incentives, and indicates that students adjust their learning effort to the standard imposed (Correa and Gruver, 1987; Costrell, 1994; Betts, 1998; Bonesrønning, 2004; Figlio and Lucas, 2004; DePaola and Scoppa, 2007; Babcock, 2010). A second literature considers endogenous household responses to school resources. This shows that "parents appear to reduce their effort in response to increased school resources" (Houtenville and Conway, 2008, p. 437), and that only changes in public education spending *unanticipated* by households affect test scores (Das *et al.*, 2013). Both findings suggest a "'crowding out' of school resources" (Houtenville and Conway, 2008, p. 437) due to households' re-optimization efforts. A third relevant literature investigates the presence of grade inflation in schools. It shows that, when possible, schools indeed engage in grade inflationary practices (Walsh, 1979; Wößmann, 2003). Taking these three literatures together suggests that increased resources can trigger endogenous re-optimizing responses, which may, whenever possible, take the form of schools inflating their

³ Bishop and Wößmann (2004) argue that centralized assessment standards improve grades' signaling value on the labor market because there is no option to 'inflate' grades in such a setting.

grades. Since students, in turn, might react to changing grading standards through their effort choice, grading standards may play a key role in the resources-achievement relation.

This naturally raises the question what determines grading practices (and whether resource shifts are one of these determinants). In this respect, two key factors are discussed in the existing literature. The first are accountability systems that evaluate teachers' and schools' performance via students' test scores. Apart from exerting a positive effect on students' achievements (Carnoy and Loeb 2002; Koning and van der Wiel 2012), undesired side-effects of such systems range from focusing teaching effort on pupils with achievements close to tests' thresholds (Reback, 2008; Neal and Whitmore Schanzenbach, 2010; Rockoff and Turner, 2010) to the distortion of results and cheating by teachers (i.e. grade inflation; Jacob and Levitt, 2003). The second determinant of grading practices is student 'demand'. Bonesrønning (1999), for instance, argues that rent-seeking students may press for easy grading while DePaola and Scoppa (2010) highlight diverging preferences of high- and low-ability students for precise versus noisy grading (see also Himmler and Schwager, 2013). However, the only study explicitly linking school resources to grading practices is Backes-Gellner and Veen (2008). They argue that schools have incentives to lower their grading standard if their budget depends on the number of students. Although they do not provide a formal verification, their argument suggests that grading standards might depend on financial constraints. Yet, it does not necessarily imply that this likewise holds for public education expenditures. This is the question addressed in the remainder of this article.

3. Theoretical framework

3.1 Assumptions

We consider two key actors in the educational process: students and schools (an extension including teachers is straightforward). Students' utility is assumed to depend on leisure l and exam results y : i.e., $u^{STU} = u^{STU}(y, l)$ with $u_l > 0$, $u_y > 0$, $u_{ll} < 0$, $u_{yy} < 0$ (subscripts denote partial derivatives). To obtain explicit results, we assume that the utility function is similar among students and has a Cobb-Douglas specification: $u^{STU} = y^\alpha l^{(1-\alpha)}$.⁴ Furthermore, students are endowed with one unit of time, which they can devote either to leisure l or to studying e : i.e., $l + e = 1$.

The overall exam result (y) is a function of the results in both a central (denoted by c) and a school exam (denoted by s), $y = y(c(n^c, e, x), s(n^s, e, x))$, thereby reflecting the idea that student performance is often measured via both types of exams (see note 1 for international evidence).⁵ The grading standard n^c is decided upon by a central institution and is constant across all schools, whereas the school's grading policy, n^s , is chosen locally and can differ between schools. A higher grading standard, i.e. an increase in either n^c or n^s , causes a decrease of the overall exam result y .⁶ In contrast to other models investigating the relationship between educational standards and student effort, we do not model the standard as a threshold value of points or a grade which students must obtain to successfully graduate.

⁴ While more general forms of the utility function could be imagined, the Cobb-Douglas representation captures several useful and intuitive properties also imposed in the foregoing literature. For instance, it implies positive marginal utility in both achievement and leisure (Correa and Gruver, 1987; Costrell, 1994; Bonesrønning, 1999), partial but not perfect substitutability between both goods (which appears a realistic description of human behavior), and incorporates education costs in a simple fashion (which makes further restrictive assumptions on this unnecessary).

⁵ The measuring unit of exam results is points. As the total number of points achievable in tests (and especially in final exams) is sufficiently large in most cases, c , s and y are assumed to be continuous variables.

⁶ Note that this assumption implies that the student's utility function is strictly decreasing in n^c . The national examiner could thus, in principle, make all students happier by increasing everyone's grade. In practice, however, grades awarded at the level of secondary education commonly cover the entire available spectrum (e.g., in Belgium, Italy, Norway, United Kingdom, United States or the Netherlands; exceptions occur in, for instance, France or Spain), which prevents exploitation of students' 'grade illusion'. Moreover, the model's key empirical implications remain unaffected when students do not maximize absolute achievement y , but relative achievement y/\bar{y} (where \bar{y} is the average achievement across all schools in the country; full details upon request).

Instead, we assume that graduation certificates differentiate between numerous possible grades and not only between the passing or failing of the exam. The underlying assumption is that the final grade y not only serves as a signal to employers but also contains information on a student's productivity. With the existence of central exams, the results of students from different schools are at least partly comparable, which justifies this assumption. Both exam results furthermore depend on learning effort, e , and per-pupil education expenditures available to the school, x and increase in both of these variables.

In the analysis below, we specify the exam result function as follows (though similar results are obtained with alternative specifications; see, for instance, Appendix A):

$$y = \underbrace{p_c^0 - n^c + \frac{1}{n^c} xe}_{=c(e,x,n^c)} + \underbrace{p_s^0 - n^s + \frac{1}{n^s} xe}_{=s(e,x,n^s)} \quad (1)$$

In this specification, the grades on both exams act as perfect substitutes. Nevertheless, introducing weights reflecting the relative importance of the central and school exam would not alter the results qualitatively.⁷ Both the central and school results consist of one part that is constant in student effort ($p_c^0 - n^c$ and $p_s^0 - n^s$) and another part that can be influenced by learning ($\frac{1}{n^i} xe$, with $i = s, c$). The former can be interpreted as the number of points a student achieves without any learning (i.e., the so-called specificity of a test), and measures the general difficulty of the exam (with p_i^0 reflecting the grade under the easiest exam possible).⁸ The latter part models the relation between educational expenditures and exam results as a linear function (though more general specifications do not qualitatively change the results; available

⁷ Note also that one could allow each grade to enter the determination of y non-linearly or let y reflect that a very bad grade on either exam is very damaging to students' achievement. We leave those issues aside here, and, for clarity of presentation, focus on the simplest possible formulation.

⁸ Since any $p_i^0 > 0$ merely increases the results for *all* students and thus provides no information on knowledge differences, we will assume $p_i^0 = 0$ to save on notation.

upon request) and assumes that tougher grading lowers the positive effect of an additional unit of effort on exam results.⁹ Note that although the basic model abstracts from students having different learning abilities (which influence the individual productivity of effort and thus the exam result), Appendix B shows that the model's main results are not altered if an ability distribution is introduced.

As mentioned, schools decide on their grading policy n^s , and we assume that they can enforce its implementation in all classes. Schools' utility is assumed to depend on student performance y and the difference in results on the central and the school exam: i.e. $u^{SCH} = u^{SCH}(y, (c-s)^2)$. We assume that $u_y > 0$. While there are many possible arguments to substantiate this assumption, one reason is that schools compete over exam results to attract students and parents – and often also the government – use them to evaluate schools (e.g. Wikström and Wikström, 2005; Reback, 2008; Neal and Whitmore Schanzenbach, 2010; Rockoff and Turner, 2010). Moreover, we assume that $u_{(c-s)^2} < 0$, which captures the idea that deviations between the two exam results are harmful to the school in either direction. That is, if $s < c$ parents may decide to send their kids to another school to get better overall grades, while if $s > c$ a school may lose students because teachers' requirements – and thereby students' knowledge gain – are deemed too low.¹⁰ Below, we employ a simple additive structure for the schools' utility function:

$$u^{SCH} = y - (c - s)^2.$$

⁹ While the direction of the grading policy effect on the effort-result relation would in a more general framework obviously depend on how the mapping from underlying learning to measured achievement varies across both exam types, we here implicitly assume that students can always improve their results on both exam types by increasing effort. Although this is somewhat restrictive when considering a single test (as students could in principle obtain the maximum feasible grade), it is a reasonable approximation for a set of final exams accumulated across several subjects. Still, taking a more agnostic approach and assuming that returns to effort may differ in some unknown way across exam types does not qualitatively affect our findings. We are grateful to Julie Cullen for this insight.

¹⁰ In a country without catchment areas, competition for students between schools may lead schools to also care about neighbouring schools' performance. In this case, households choose which school to attend by comparing the achievable utility, given the educational standards and expenditures of all schools (cf. Koning and van der

The timing of events is as follows: In a first step, schools choose their grading standard, knowing the per-pupil expenditures x they are (exogenously) assigned by the government. Afterwards, students observe the grading policy and choose their learning effort.¹¹

3.2 Students' decision

Solving the model backwards, the students' maximization problem is:

$$\max_e u^{STU} = y^\alpha l^{(1-\alpha)} \quad \text{s.t.} \quad 1 = l + e \quad (2)$$

The first-order-condition yields:

$$\begin{aligned} \frac{du^{STU}}{de} &= \alpha \frac{\partial y}{\partial e} l + (1-\alpha)y \frac{\partial l}{\partial e} \\ &= \alpha(1-e)x \left(\frac{1}{n^c} + \frac{1}{n^s} \right) - (1-\alpha) \left(-n^c + \frac{1}{n^c}xe - n^s + \frac{1}{n^s}xe \right) = 0. \end{aligned} \quad (3)$$

The first summand of equation (3) shows the marginal revenue of an increase in student effort: exam performance (and thus utility) rises. The second summand shows that effort decreases the amount of time devoted to leisure, which lowers utility and thus represents the marginal cost of higher effort. In equilibrium, students choose their learning effort such that marginal revenue equals marginal cost. Hence, optimal student effort as a function of expenditures and central and school grading standards equals:

$$e^* = \alpha + (1-\alpha) \frac{n^c n^s}{x}. \quad (4)$$

Wiel, 2013) - and the number of students thus may enter the school's utility function. To most clearly isolate the expenditure effects we are interested in, we abstract from such competition effects here.

¹¹ Although the government can be seen as a third actor setting both expenditures x and the central grading standard n^c , we refrain from explicitly modelling the government's optimization problem. The reason is that we are interested in schools' reaction to an (exogenous) change in expenditures, rather than the optimal choice of x and n^c . Moreover, any adjustment of n^c concomitant to a change in x would influence students' and schools' behaviour, thus distorting the effect of the expenditure change we are interested in.

From equation (4), it is easy to see that effort increases in both the central and the school's grading scheme and declines in per-student expenditures. The latter effect materializes because x directly increases exam results and thereby substitutes student effort. The intuition for the former effect is that harsher grading has a negative effect on exam results, which stimulates students to work harder in order to make up the loss (even though tougher grading diminishes the return of effort in terms of improved exam results).¹² Various empirical studies present evidence for such positive relationship between grading standards and (average) student achievement, also arguing that this effect arises from an increase in effort (Betts and Grogger, 2003; Figlio and Lucas, 2004).

3.3 Schools' decision

Anticipating students' reaction to the grading policy, the school's maximization problem and first-order condition read:

$$\max_{n^s} u^{SCH} = y(e^*) - (c(e^*) - s(e^*))^2$$

and

$$\begin{aligned} \frac{du^{SCH}}{dn^s} &= \frac{\partial y(e^*)}{\partial n^s} - 2(c(e^*) - s(e^*)) \left(\frac{\partial c(e^*)}{\partial n^s} - \frac{\partial s(e^*)}{\partial n^s} \right) \\ &= -\alpha \left(1 + \frac{x}{n^{s2}} \right) - 2 \left((2 - \alpha)(n^s - n^c) + \alpha x \left(\frac{1}{n^c} - \frac{1}{n^s} \right) \right) \left(2 - \alpha + \frac{\alpha x}{n^{s2}} \right). \end{aligned} \quad (5)$$

Equation (5) illustrates the marginal effects of a change in the school's grading scheme. First, the grading standard chosen will affect the overall exam result y :

¹² As grading standards are not modelled as threshold values that must be met, there exists no situation in which students will, in response to a standard increase, lower their effort because standards have become too demanding (Costrell, 1994; Betts, 1998).

$$\frac{\partial y}{\partial n^s} = -1 - \frac{xe}{n^{s2}} + \frac{\partial e}{\partial n^s} x \left(\frac{1}{n^c} + \frac{1}{n^s} \right). \quad (6)$$

Equation (6) shows that an increase of n^s has a direct negative effect on the school exam results, represented by the first two summands: Harsher grading lowers each student's results. Moreover, an indirect effect arises because students adjust their learning effort e . Given the positive relationship between grading standards and effort $\left(\frac{\partial e}{\partial n^s} > 0 \right)$, they will study harder to compensate for the loss of points caused by the direct effect. This reaction will positively influence the grades on both the school and the central exam. Consequently, student performance on the central exam improves following an increase in the school's grading standard $\left(\frac{dc(e^*)}{dn^s} > 0 \right)$ (as no negative direct effect exists), whereas results on the school exam may both improve or decline. At the equilibrium effort level e^* , the negative direct effect dominates both at the school level $\left(\frac{ds(e^*)}{dn^s} < 0 \right)$ and in the aggregate $\left(\frac{dy(e^*)}{dn^s} < 0 \right)$. Hence, schools can improve students' overall exam results by lowering their grading standards.

The second summand in equation (5) shows that the grading standard chosen will affect the difference between school and central exam results. As shown above, an increase in n^s raises c but lowers s at the optimal effort choice e^* . The sign of the overall effect thereby depends on the sign of the original difference $c-s$. If $c-s > 0$, an increase of the school's grading standard causes two negative effects by both lowering the overall exam result y and increasing the difference $c-s$. Thus, a corner solution arises in which the school has an incentive to decrease n^s . For an inner solution to exist, $c-s < 0$ must hold. In this case, tougher grading reduces the

difference between the two exam results. This effect provides schools with an incentive to increase its grading standard n^s , which counteracts the incentive to inflate grades discussed above.

As it is not possible to solve equation (5) for $n^{s*}(x, n^c)$ explicitly, we investigate the effect of higher education expenditures on the grading standard with the implicit function theorem

$$\frac{dn^s}{dx} = -\frac{\frac{\partial^2 u^{SCH}}{\partial n^s \partial x}}{\frac{\partial^2 u^{SCH}}{\partial n^{s2}}} \quad (7)$$

As the denominator of equation (7) is the second-order condition of the school's optimization problem in equation (5), it must be negative at the utility-maximizing standard. The sign of the overall effect in equation (7) is thereby defined by the numerator, which reads:

$$\begin{aligned} \frac{\partial^2 u^{SCH}}{\partial n^s \partial x} &= \frac{\partial^2 y}{\partial n^s \partial x} - 2 \left(\left(\frac{\partial c}{\partial x} - \frac{\partial s}{\partial x} \right) \left(\frac{\partial c}{\partial n^s} - \frac{\partial s}{\partial n^s} \right) + (c - s) \left(\frac{\partial^2 c}{\partial n^s \partial x} - \frac{\partial^2 s}{\partial n^s \partial x} \right) \right) \\ &= -\frac{\alpha}{n^{s2}} - 2 \left(\alpha \left(\frac{1}{n^c} - \frac{1}{n^s} \right) \left(2 - \alpha + \alpha \frac{1}{n^{s2}} x \right) + \left((2 - \alpha)(n^s - n^c) + \alpha x \left(\frac{1}{n^c} - \frac{1}{n^s} \right) \right) \frac{\alpha}{n^{s2}} \right). \end{aligned} \quad (8)$$

Two opposing effects can be distinguished. The first summand shows that the effect of a higher grading standard on y varies in x . At the equilibrium effort level e^* it can be shown that

$\frac{\partial^2 y}{\partial n^s \partial x} < 0$ holds. Thus, an increase in expenditures reinforces the incentive for schools to

choose an *easier* grading policy.¹³ The second term illustrates that both the difference $c-s$ as

¹³ Remember that schools' utility depends on y because students' exam results are often publicly available (e.g., in the Netherlands, the average final grade within each school becomes public information). Parents as well as governmental institutions thus are able to employ exam results to evaluate a school's performance and its use of monetary resources, which underlies schools' incentive to reduce grading standards and improve observed outcomes (see equation (5)). Moreover, as governments expect a positive (direct) effect of higher expenditures

well as the change of this difference in n^s vary in x . Inserting e^* reveals that the second summand is positive in equilibrium. Thus, higher education expenditures strengthen the decrease in the difference between school and central exam results generated by a higher n^s (see above). Hence, schools face a stronger incentive to *increase* their grading standard following an increase in expenditures.

Overall, grade inflation following an increase in educational expenditures is observed if equation (7) is negative, which requires that the cross-derivative in equation (8) is negative as well. The occurrence of this constellation depends on the original level of x as well as on the central exam grading standard n^c . The school's aim to provide its students with a high level of y causes an incentive for grade inflation, whereas the objective to minimize the difference $c-s$ counteracts this effect. Intuitively, the incentive to lower n_s grows stronger the more important exam results y become in the school's objective function compared to other goals. As such, while grade inflation following increased public education expenditures is certainly a theoretical possibility, it remains an empirical question whether or not it occurs in reality. The theoretical model thus allows us to derive the following predictions, which will be tested empirically in section 4.

Prediction 1: *An increase in educational spending changes the schools' grading behaviour.*

Prediction 2: *If n^s increases (decreases), results on the central exam improve (deteriorate), whereas results on the school-level exam deteriorate (improve).*

3.4 Effect on student attainment

on achievement, an increased incentive to engage in grade inflation arises because schools will attempt to cater to this expectation.

We can now assess the overall effect an expenditure change exerts on educational attainment (as extensively discussed in the foregoing literature). Presuming the relationships between expenditures, school grading standards and effort analyzed above, we have:

$$\frac{dy}{dx} = \left(\frac{1}{n^c} + \frac{1}{n^s} \right) \left(e + x \left(\frac{\partial e}{\partial x} + \frac{\partial e}{\partial n^s} \frac{\partial n^s}{\partial x} \right) \right) - \frac{\partial n^s}{\partial x} - \frac{1}{n^{s^2}} \frac{\partial n^s}{\partial x} x e. \quad (9)$$

Equation (9) shows that the following effects can be distinguished: First, the direct effect of an expenditure increase is unambiguously positive for both exam types. Second, students decrease their effort as a direct reaction to higher expenditures, which has a negative impact on y . Third, the school's grading standard will be altered, which directly affects the school-level exam results. Finally, student effort changes in response to the change in n^s . If an expenditure increase induces grade inflation $\left(\frac{\partial n^s}{\partial x} < 0 \right)$, the third effect will be positive (improving school exam results), but the fourth effect becomes negative, causing school exam results and effort – as well as overall results – to deteriorate. It is worth highlighting that by these various effects, our theoretical framework provides a possible explanation for the ambiguity in the empirical literature about the effects of increased educational spending. Indeed, even when *assuming* that an increase in educational spending has a positive direct influence on student achievement, adjustments in students' and schools' behavior in response to changes in available resources may create important counteracting effects (which can under certain conditions dominate the direct effect).

4. Institutional setting and data

4.1 Financing Dutch schools

Since 1917, all Dutch schools receive a fixed allowance per enrolled student from the central government budget. The size of each school's budget thus depends on the number of students, their age and the type of education (i.e. general or vocational). Higher allowances are thereby provided for children whose parents have lower education levels (as such children are assumed to have higher educational needs). Specifically, the Ministry of Education distinguishes between the situation where both parents failed to complete higher secondary education and the situation where one parent completed higher secondary education while the other one did not. Children in the former situation induce a 30% higher allowance, while children in the latter situation induce a 20% higher allowance. The total budget received by the schools is a 'lump sum' transfer covering both material (about 15% of the budget) and personnel expenses (about 85% of the budget).

4.2. The intervention: Earmarked block grants in specific districts

As in most Western countries, some neighborhoods in the Netherlands are characterized by a combination of poverty, unemployment and social instability. Shortly after its appointment on 22 February 2007, the Balkenende IV administration announced a new policy program allocating block grants to 40 such districts (labeled 'power districts', or '*krachtwijken*' in Dutch) – consisting of 83 postcode areas situated in 18 large and medium-sized Dutch cities¹⁴ – earmarked to improve their social, physical and economic environment. The total subsidy amounted to 250 million euro annually (ranging from €1.2 million to €29.3 million across districts, or €333 to €995 per inhabitant), and the selection of the districts was driven by a set of 18 indicators including the income, education and unemployment levels within the local population and the incidence of public disorder issues (Tweede Kamer, 2008-2009). The final decision to include or exclude districts was taken by the minister (i.e., Ella Vogelaar) roughly

¹⁴ The selection of postcode areas was based on a long-list with 180 additional postcode areas (which did not receive additional funding). Information on the excluded postcodes has not been made public, and is considered 'highly confidential' by the Dutch government.

one month after the new government was inaugurated, and the program was announced and implemented in July 2007. The program lasts for 10 years, and no restrictions were imposed on how the funds could be used (except being subject to agreement between the municipality, housing corporations, schools and other local stakeholders).

Although the speed and organization of the selection process precluded lobbying efforts by districts desiring to be included (thus mitigating concerns about potential self-selection), the selection process obviously was non-random since the government aimed at selecting the worst-performing districts. Fortunately, the government selected only 40 districts and thus left a substantial number of similarly ‘underperforming’ districts outside the chosen sample, which we exploit below. As a result, we are left with a quasi-experimental setting where some underperforming districts received additional funding while other underperforming districts did not.

We should also note that while the various actors involved in the policy program (i.e., schools, local government, housing corporations and the regional government) retained some leeway in setting their objectives, schooling and youth received substantial attention across the board. For instance, in 16 out of the 18 cities with power districts, investments were explicitly aimed at improving the schooling outcomes of local youth. This makes the improvement of education the most central and commonly stated ambition in the power district policy (Tweede Kamer, 2008-2009, p. 68).¹⁵

4.3. The data

¹⁵ Excluding both cities that did not explicitly mention education investments in their power districts policy program leaves our results unaffected (details upon request).

In the final year of secondary education, all students in the Netherlands have to take two exams for each course in which they received lessons (independent of the educational track). The first exam – the ‘central exam’ – is a national assessment constructed by the Central Institute for Assessments (CITO). It is, by definition, an absolute assessment with criterion-referencing. It is externally screened by professors and a prior test on a sample of students is taken to measure and monitor its difficulty, which is thereby guaranteed to remain constant over time. Correction of this central exam is based on a uniform correction model and there is a teacher from a different school acting as a second corrector. Only three small courses do not have a central exam: i.e. civics, arts and physical education.

The second exam – the ‘school exam’ – has fewer quality controls in its construction and evaluation as it is set up and corrected only by a school’s teachers. Moreover, part of the grade on the school exam is earned during the year in the form of intermediate tests and assignments. Nevertheless, the school exam is also criterion-referenced as there is a strict legal framework setting out the knowledge achievements required of students at the end of each year for each course. The student’s final overall grade consists of the arithmetic average of the central and the school exam (no additional information is incorporated).

Our central variables of interest are the results on the central and school exams, which are collected on an average level across subjects within schools on an annual basis. For ease of interpretation, we recalibrate all grades into the 0-10 band (in which 0 is the worst and 10 the best grade possible). Unfortunately, data on the individual subjects are unavailable. Yet, the average grades per school we employ in the analysis below become publicly available information, and is used by parents and Education Inspectorate to compare schools and evaluate their quality.

The dataset – originating from the Dutch Ministry of Education – includes information for 738 schools, which are well spread across the Netherlands, over the period 2004-2009 (previous years could not be included due to data inconsistencies). Descriptive statistics are presented in Table 1 for the period before (2004-06) and after (2008-09) the policy intervention, and for schools in treated/untreated districts.¹⁶ Table 1 illustrates that the average grades on the central exit exam lie below those on the school exam (a common observation in the Netherlands; see also Dronkers, 2012), and that the average difference between both types of exams increases over time. This holds for schools in both treated and untreated districts, though it appears substantially stronger in the former subsample. This increased divergence of school and central exam results is largely driven by worsening central exam results (for a similar observation, see Dronkers, 2012). Still, Table 1 hides significant heterogeneity across schools in both observations, which we exploit in the analysis below.

Table 1 about here

Table 1 also contains summary statistics for a number of control variables (likewise separated before/after treatment and by schools inside/outside treated districts). While we unfortunately lack information about, for instance, the number (or quality) of teachers and school provisions (such as the number of computer terminals and the presence/size of a school library), we do have information on the size of the student population in a subset (N=523) of schools. We also observe postcode information for each school, such that we can match each school to data on socio-demographic characteristics in its neighborhood (obtained from Statistics Netherlands).

¹⁶ We exclude the year of the intervention (i.e., 2007) even though exams for that year had already passed by the time of the intervention and thus could not possibly be influenced by it.

This provides information on the number of inhabitants, urbanization (5-point scale where 1 is urban and 5 is rural), percentage of employed residents and welfare recipients (both as share of working-age population), average income (measured as after-tax income in 1000€) and the percentage of young (under 25), old (over 65) and immigrants (each as a share of total population). One important thing to note from these statistics is that treated districts were larger, younger, poorer (which is observable in terms of income, employment and social welfare recipients) and ethnically more diverse than untreated districts. As this may induce selection bias, we will employ this information extensively in our robustness checks in section 5.3.

5. Empirical analysis

5.1. Empirical Strategy

Our analysis exploits the variation in public investment across space and time due to the July 2007 policy intervention via a difference-in-differences (DiD) approach. The existence of comparable districts without additional funding allows us to infer the counterfactual outcome and estimate the causal impact of public resources. Particularly, we compare educational outcomes in Dutch schools inside the 40 districts covered by the new legislation (the ‘treated’ group; 35 schools in 27 districts) with those not covered by the new legislation (the ‘control’ group; 703 schools in 493 districts) before/after 2007 using information covering the 2004-2009 period. Consequently, the control group in our basic specification consists of all observed schools *not* located in a power district. Still, as the similarity of the treated and the control group is critical for the validity of our inferences, we extensively test the robustness of our results to the specification of the control group in section 5.3. Note also that since DiD approaches yield inconsistent standard errors when analyzing serially correlated outcomes, we follow the suggestion of Bertrand et al. (2004) to average exam results by school over the

period before (i.e., 2004-2006) and following (i.e., 2008-2009) the intervention. This leads to the following baseline specification (with subscript i for schools and t for time):

$$SE_CE_{i,t} = \gamma_i + \beta_1 Time_t + \beta_2 PowerDistrict_{i,t} * Time_t + \sum_k \lambda_k X_{i,t} + \varepsilon_{i,t} \quad (10)$$

$SE_CE_{i,t}$ reflects the difference at time t in the mean result of school i 's pupils on the school exams (SE) and the central exams (CE). Positive numbers indicate that school exams performance exceeds that on central exams (and vice versa). We also estimate the model separately for SE and CE as this yields an indication on the progress in educational attainment. The variable $PowerDistrict_{i,t}$ is an indicator variable equal to 1 for schools in districts receiving additional block grants, and 0 otherwise. The indicator variable $Time_t$ separates the period before ($Time_t=0$; i.e., the 2004-2006 period) and after the policy intervention ($Time_t=1$; i.e., the 2008-2009 period). To control for unobserved heterogeneity across schools – also among schools *within* the power districts – Equation (10) includes school-specific (γ_i) fixed effects that capture all time-invariant differences across schools. The variable of interest is the interaction between $Time_t$ and $PowerDistrict_{i,t}$, whose coefficient β_2 estimates the causal effect of the policy intervention on $SE_CE_{i,t}$. $X_{i,t}$ stands for a vector of (k=5) control variables including the district population size, the school's student number (both in logarithmic form), and the share of immigrant, young and old residents in the district population. Their inclusion is critical to adjust for any differences in educational attainment that are a function of the population and student composition (Angrist and Pischke, 2008; Fiva, 2009) – especially when the government's selection process may have been influenced by such observable socio-demographic indicators.¹⁷

¹⁷ The information on urbanization, employment, income and the share of welfare recipients mentioned above is only available for the year 2003, and thus cannot be included in our fixed effects estimation. We do, however, use this information in our robustness checks based on a matching estimator (see section 5.3).

Still, this baseline approach ignores the variation across power districts in the level of additional resources created by the new legislation. We can exploit this information for identification purposes by expanding the model with this “explanatory variable with differing treatment intensity across localities” (Berrebi and Klor, 2008, p. 208; Angrist and Pischke, 2008). This extends our estimation equation to:

$$SE_CE_{i,t} = \gamma_i + \beta_1 Time_t + \beta_2 PowerDistrict_{i,t} * Time_t + \beta_3 Investment_{i,t} + \sum_k \lambda_k X_{i,t} + \varepsilon_{i,t} \quad (11)$$

where $Investment_{i,t}$ equals the level of annual additional public investment (in 1000€ per capita) in the district of school i at time t deriving from the new policy program. Clearly, this is 0 before the intervention, but varies across schools after the intervention (though remaining 0 in ‘untreated’ districts). Its inclusion permits disentangling the effect of receiving the status as ‘Power district’ at time t (β_2) from the effect of the public expenditures associated with this status (β_3).¹⁸

The key identifying assumption underlying equations (10) and (11) is that the trends in educational outcomes in school in treated and untreated districts would be the same except for the intervention (the parallel time trend assumption; Bertrand et al., 2004).¹⁹ This raises three issues in our setting. First, as mentioned, the government selected the worst-performing

¹⁸ Clearly, this is only one way to assess the effect of the level of investment, and we implemented a number of alternative approaches to assess the robustness of our findings. First, we estimated a model including school fixed effects, a time dummy, and the $Investment_{it}$ variable (along with its squared version to assess any non-linearity). Second, we implemented a replication of equation (10) that includes a series of three dummies effectively splitting the original DiD coefficient into three parts depending on investment size (i.e. low, medium or high). Third, we estimated equation (10) for different subsamples of schools depending on the size of the investment (i.e. low, medium or high). In all three cases, the results mirror those presented below (details upon request). We are grateful to an anonymous referee for suggesting these alternatives.

¹⁹ Obviously, treated and untreated districts should also be similar in terms of their pre-existing attributes (i.e. apart from the treatment). We deal with this assumption extensively in section 5.3.

districts non-randomly. Second, selection into the program may have triggered migration flows, which could invoke violation of the parallel time trend assumption. In this respect, it is important to observe that the list of selected districts was only made publicly available after a lengthy legal proceeding in February 2009. Consequently, any in- or outward mobility between July 2007 and (at least) February 2009 can reasonably be taken as independent of residents' district being included in the list. This is important as students in the Netherlands have free school choice (there is no catchment area). Moreover, data from Statistics Netherlands illustrate that the share of western and non-western migrants, natives, citizens under 20 or over 65 years, employed, unemployed and one-parent-families is stable over time in both treated and untreated districts (i.e., the share of these respective population groups does not change significantly over the 2004-2009 period). These statistics strongly suggests that there were no obvious changes in the underlying population in the 2004-2009 period. Third, selection into the program may have affected students' drop-out behavior. Such changes in the retention of 'marginal' students before/after the treatment could invoke violation of the parallel time trend assumption. To address this concern, we collected additional information on drop-out rates (i.e. the share of students leaving school without a diploma), and replicated our analysis using this as an alternative left-hand side variable (N=640 schools). We do not find any effect of the treatment, nor of treatment size, on drop out rates (details upon request).

To also more directly test the parallel time trend assumption, we compare the evolution of central and school exam grades across treated and untreated districts over the 2004-2007 period (given the timing of exams, this period completely precedes the July 2007 intervention). This is achieved by running a regression that mirrors equation (10) for the 2004-

2007 period.²⁰ The results, summarized in Table 2, indicate that central exam grades are significantly lower in 2006-2007 compared to 2004-2005 ($\beta = -0.203$, $p < 0.01$). Nonetheless, and crucially, we find no evidence that this downward pre-treatment trend is different for our treatment and control groups ($\beta = -0.009$, $p > 0.10$). The same observations likewise hold for school-level exam grades (Column 2). Strictly speaking, the results in Table 2 only verify the validity of equation (10). We obtain very similar results, however, when we replicate this test for different subsamples of districts depending on the size of the investment (i.e. low, medium or high). This supports the parallel time trend assumption also for schools in districts with varying investment levels, which validates equation (11).

Table 2 about here

5.2. Empirical Results

Our baseline findings, which exploit the full set of 738 available schools, are summarized in Table 3 (section 5.3 reports on a number of robustness checks with more restrictive control groups). Columns (1) through (3) provide results for the estimation of equation (10). Columns (4) through (6) also include the annual investment level due to the policy program within every district. In each case, the first column (i.e., column (1) and (4)) has as dependent variable the difference between school and central exam results (to assess Prediction 1), while the next two columns have, respectively, school and central exam results as dependent variables (to assess Prediction 2). Throughout the analysis below, all models are estimated at the school level because our key dependent variables (i.e., exam results) are recorded at this

²⁰ Note that the number of schools reduces to 636 in this time period due to missing observations. We show below that our main findings hold for both the full sample of 738 schools, and this reduced sample of 636 schools.

level. However, since the policy intervention happens at the district level, an argument could also be made for an analysis at the district level. To check whether this affects our findings, we averaged the school-level outcomes at district level, and estimated equations (10) and (11) at the district level. Our main results are unaffected by this change in the level of analysis (full results available upon request).

The results in Table 3 indicate that when looking at the policy intervention using an indicator variable (columns (1) through (3)), the evidence regarding Predictions 1 and 2 is relatively weak. The interaction effects remain statistically insignificant at conventional levels, suggesting that the policy intervention did not affect either school or central exams. Although one explanation may lie in the fact that we evaluate the policy intervention immediately after the investments started,²¹ it might also be that a simple dummy for treated/untreated districts obscures differences in the effects due to the varying treatment *intensity*. Columns (4) through (6) show that the level of the additional investment indeed plays a critical role. Particularly, while there still is no insignificant effect for school-level exam results (column (5)), we now observe a statistically significant negative effect on central exam results until the investment surpasses approximately €1250 per resident. For higher levels of investment, there is no significant impact although the marginal effect becomes positive around €2000 per resident (column (6)).

Table 3

²¹ If improvements take some time to fully develop and become visible in exam grades, this may exert downward pressure on our coefficient estimates. This possible delay should not, however, undermine our ability to detect (potential) adjustments in school-level grading practices.

To properly interpret these results, it is important to note that teachers often grade students based on (i) valued characteristics that do not readily translate into better performance on standardized exams (e.g., students' behaviour/civic attitude, or performance relative to initial ability level) or (ii) a grading curve that remains fairly stable over time. Although our DiD strategy eliminates the static component of locally-assigned grades along these dimensions, the evidence in Table 3 seems consistent with the idea that teachers continue to grade students on a similar curve regardless of their absolute levels of performance. However, keeping to the same curve despite deteriorating national exam results (i.e., the negative effect on CE) *de facto* amounts to a reduction of the grading standard. From the theoretical model, we know that such weaker school-level grading negatively affects student effort, which induces deterioration in students' performance *in the central exam* while (weakly) improving performance *in the school exam* (see section 3.3). This is exactly what we observe in Table 3. Overall, our main findings therefore suggest that the policy intervention worked to halt falling central exam results in the selected districts when additional funds were sufficiently elevated, but induced grade inflation – by schools (or teachers) failing to downgrade mean locally-assigned grades in spite of declining scores on national exams – when such funds were limited.

Reconsidering the underlying mechanism(s), one particularly interesting possibility may be that schools in districts receiving less funding might have used these resources for more basic investments (e.g., sport facilities, library expansion, computer or media rooms) that predominantly have an impact *i*) in the long run (but are unobservable within the short period analysed), *ii*) on students' behaviour/civic attitude (which is reflected rather in school than national exams), and *iii*) on topics not covered by the central test (i.e., civics, sport, art). While the first two of these effects can be categorized as specific forms of grade inflation – since

locally-assigned test results would not reflect students' real, current academic performance – the latter cannot. Unfortunately, the lack of detailed spending data prevents us from investigating this issue in further detail.

5.3. Robustness Checks

Since the power districts were purposefully chosen by the government, one evident worry is that the above results are driven by the dissimilarity of 'treatment' and 'control' districts in pre-existing attributes (see Table 1). To attenuate such concerns, we implemented a number of robustness checks. The first of these replicates all estimations in Table 3 restricting the sample to schools in districts with at least as many inhabitants, young citizens or migrants, and at most the number of older citizens than the 40 treated districts. Table 4 shows that none of these restrictions – which increase the similarity between comparison districts – changes the inferences from those reported in Table 3. The same holds also when we impose all four restrictions at the same time to obtain the most restrictive – and therefore most comparable – control group feasible.

Table 4 about here

Secondly, we implemented a matching estimator (using *psmatch2* in Stata12; Leuven and Sianesi, 2010) because this approach exploits the purposeful assignment to the treatment and allows us to incorporate additional time-invariant background characteristics of the districts (see above). To match treated districts to similar untreated districts in this analysis, we ran a probit regression using population size, percentage immigrants, the level of urbanization, the percentage of employed residents and welfare recipients (both as share of working-age

population) and average income in the district (measured as after-tax income in 1000€) as explanatory variables. We also include squared terms of the share of immigrants and income. The resulting model satisfies the balancing properties of the matching procedure as there remain no significant differences between the matched set of treated and untreated districts. Using the results predicting treatment in the matching procedure to trim the sample based on the schools' propensity scores – which provides a sample including districts receiving additional grants *and* comparable districts where no grants were awarded – leaves our results unaffected (see Table 5).

Table 5 about here

A final robustness check evaluates whether the results in Table 3 are really due to the 2007 policy intervention by implementing a placebo estimation comparing the 2004-2005 period to the 2006-2007 period. A replication of the results of Table 3 during this exercise would indicate that the findings in Table 3 are *not* specific to the 2007 policy intervention, which would cast serious doubt on our interpretation that this intervention caused grade inflation. In other words, given that no intervention had yet taken place in the placebo sample, no significant effects should arise in this exercise. This is borne out by the three left-hand side columns of Table 6. Importantly, this result is not due to the reduction in sample size (to 635 rather than 738 schools). In fact, running the original model (i.e., comparing the actual pre- and post-treatment periods) on this reduced sample produces significant effects very much in line with those reported in Table 3 (right-hand side columns of Table 6).

Table 6 about here

6. Conclusion

This paper examines, both theoretically and empirically, whether public expenditures induce improved educational attainment, grade inflation or both. By explicitly accounting for behavioral feedback effects in students' and schools' decisions following changes in the level of resources, our theoretical model shows that shifts in the grading structure chosen by schools are a real possibility when resources change. We test the model's main implications exploiting a quasi-experimental setting in the Netherlands, where some disadvantaged neighborhoods received earmarked block grants and other similar districts did not. The Dutch education system thereby allowed us to distinguish the unbiased educational attainment of students (measured by standardized national exam scores) from the potentially inflated school-level exam grade (which is at the discretion of the school).

Our results provide evidence for grade inflation following additional resource investments. More specifically, schools inflate their grades by failing to downgrade mean locally-assigned grades in spite of declining scores on national exams. Nevertheless, when the size of the additional resources is accounted for, the results are more nuanced: resources appear beneficial in terms of improving central exam results when the additional funds are sufficiently elevated, but induce grade inflation when the resources are limited. From a policy perspective, these results suggest that policy programs aimed at improving educational outcomes may easily 'fail' to reach pre-set targets when the apportioned resources are overly limited. True, rather than feigned, success requires sufficiently elevated additional funds.

References:

- Angrist, J.D. and J. Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.
- Babcock, P. (2010). Real Costs of Nominal Grade Inflation? New Evidence from Student Course Evaluations. *Economic Inquiry*, 48(4), 983-996.
- Backes-Gellner, U. and S. Veen (2008). The Consequences of Central Examinations on Educational Quality Standards and Labour Market Outcomes. *Oxford Review of Education*, 34(5): 569-588.
- Berrebi, C. and E.F. Klor (2008). Are Voters Sensitive to Terrorism? Direct Evidence from the Israeli Electorate, *American Political Science Review*, 102(3): 279-301.
- Bertrand, M., E. Duflo and S. Mullainathan (2004). How much should we trust Difference-in-Differences Estimates? *Quarterly Journal of Economics* 119(1): 249-275.
- Betts, Julian R. (1998). The Impact of Educational Standards on the Level and Distribution of Earnings. *American Economic Review*, 88(1): 266-275.
- Betts, Julian R. and Jeff Grogger (2003). The Impact of Grading Standards on Student Achievement, Educational Attainment and Entry-Level Earnings. *Economics of Education Review* 22: 343-352.
- Bishop, J.H. and L. Wößmann (2004). Institutional Effects in a Simple Model of Educational Production. *Education Economics*, 12(1): 17-38.
- Bonesrønning, H. (1999). The Variation in Teachers' Grading Practices: Causes and Consequences. *Economics of Education Review*, 18: 89-105.
- Bonesrønning, H. (2004). Do the Teachers' Grading Practices Affect Student Achievement? *Education Economics*, 12: 151-167.
- Carnoy, M. and S. Loeb (2002). Does External Accountability Affect Student Outcomes? A Cross-State Analysis. *Educational Evaluation and Policy Analysis*, 24(4): 305-331.
- Correa, H. and G.W. Gruver (1987). Teacher-Student Interaction: A Game-Theoretic Extension of the Economic Theory of Education. *Mathematical Social Sciences*, 13: 19-47.
- Costrell, Robert M. (1994). A Simple Model of Educational Standards. *American Economic Review*, 84(4): 956-971.
- Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K. and Sundararaman, V. (2013). School Inputs, Household Substitution, and Test Scores, *American Economic Journal: Applied Economics*, 5(2): 29-57.
- DePaola, M. and V. Scoppa (2007). Returns to Skills, Incentives to Study and Optimal Educational Standards. *Journal of Economics*, 92(3): 229-262.
- DePaola, M. and V. Scoppa (2010). A Signalling Model of School Grades under Different Evaluation Systems. *Journal of Economics*, 101: 199-212.
- Dronkers, J. (2012). SE-CE verschillen per vak en VWO- en HAVO-locatie. Algemene Onderwijsbond. Utrecht.
- Figlio, D.N. and M.E. Lucas (2004). Do High Grading Standards affect Student Performance? *Journal of Public Economics*, 88: 1815-1834.
- Fiva, J.H. (2009). Does Welfare Policy Affect Residential Choices? An Empirical Investigation Accounting for Policy Endogeneity, *Journal of Public Economics*, 93: 529-540.
- Gerritsen, S. and D. Webbink (2010). The Effects of Extra Funds for 40 Disadvantaged Neighborhoods. Paper presented at a Centraal Planbureau Conference at Rijksuniversiteit Groningen, March 3, 2010.

- Hanushek, E.A. (2003). The Failure of Input-Based Schooling Policies. *Economic Journal*, 113(February): F64-F98.
- Hægeland, T., O. Raaum and K.G. Salvanes (2012). Pennies from Heaven? Using Exogenous Tax Variation to Identify Effects of School Resources on Pupil Achievement. *Economics of Education Review*, 31(3): 601-614.
- Himmler, O. and R. Schwager (2013). Double Standards in Educational Standards - Are Disadvantaged Students Being Graded More Leniently? *German Economic Review*, 14(2): 166-189.
- Holmlund, H., S. McNally and M. Viarengo (2010). Does Money Matter for Schools? *Economics of Education Review*, 29(6): 1154-1164.
- Houtenville, A.J. and K.S. Conway (2008). Parental Effort, School Resources, and Student Achievement. *Journal of Human Resources*, 43(2): 437-453.
- Hoxby, C.M. (2000). The Effects of Class Size on Student Achievement: New Evidence from Population Variation. *Quarterly Journal of Economics*, 115(4): 1239-1285.
- Jacob, B.A. and S.D. Levitt (2003). Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating. *Quarterly Journal of Economics*, 118(3): 843-877.
- Koning, P. and K. van der Wiel (2012). School Responsiveness to Quality Rankings: An Empirical Analysis of Secondary Education in the Netherlands. *De Economist*, 160(4): 339-355.
- Koning, P. and K. van der Wiel (2013). Ranking the Schools: How School-Quality Information Affects School Choice in the Netherlands, *Journal of the European Economic Association*, 11(2): 466-493.
- Leuven, E. and B. Sianesi (2010). *PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing* (version 4.0.4.). <http://ideas.repec.org/c/boc/bocode/s432001.html>.
- Neal, D. and D. Whitmore Schanzenbach (2010). Left Behind by Design: Proficiency Counts and Test-Based Accountability. *Review of Economics and Statistics*, 92(2): 263-283.
- Reback, R. (2008). Teaching to the Rating: School Accountability and the Distribution of Student Achievement. *Journal of Public Economics*, 92: 1394-1415.
- Rockoff, J.E. and L.J. Turner (2010). Short Run Impacts of Accountability on School Quality. *American Economic Journal: Economic Policy*, 2(4): 119-147.
- Tweede Kamer (2008-2009). *Krachtwijken: Monitoring en verantwoording van het beleid*, Tweede Kamer der Staten-Generaal, 31.723, nr. 1-2, 's Gravenhage.
- Wikström, C. and M. Wikström (2005). Grade Inflation and School Competition: An Empirical Analysis based on the Swedish Upper Secondary Schools, *Economics of Education Review*, 24(3): 309-322.
- Wittebrood, K. and M. Permentier (2011). *Wonen, wijken en interventies: Krachtwijkenbeleid in perspectief*, Sociaal en Cultureel Planbureau.
- Wößmann, L. (2003). Schooling Resources, Educational Institutions, and Student Performance: the International Evidence. *Oxford Bulletin in Economics and Statistics*, 65(2): 117-170.
- Walsh, J. (1979). Does High School Grade Inflation Mask a More Alarming Trend? *Science* 203(4384): 982.

Table 1: Descriptive statistics (N=738 schools)

	<i>'Untreated' districts</i>			<i>'Treated' districts</i>		
	Min	Mean	Max	Min	Mean	Max
<i>Pre-intervention: 2004-06</i>						
Central exam	4.553	6.423	8.914	4.520	6.225	7.443
School exam	5.850	6.741	8.563	5.859	6.541	8.299
Diff school & central exam	-1.612	0.318	3.373	-0.857	0.316	2.843
Total population	5.589	8.841	9.948	8.056	8.938	9.459
Total number of students ^a	3.497	6.504	7.966	5.047	6.338	7.437
Income (in 1000€)	9.150	13.698	25.600	9.300	11.376	14.650
Share employed	39.000	69.268	85.000	48.333	57.734	70.000
Share welfare recipients	6.000	15.804	42.400	17.000	25.117	33.750
Urbaneness (5-point scale)	1	2.884	5	1	1.246	2.5
Immigrant population	2.173	19.178	85.786	28.480	54.563	77.691
Share younger than 25	11.852	23.028	40.771	21.850	26.163	35.484
Share older than 65	1.775	17.408	57.896	6.977	13.304	21.826
<i>Post-intervention: 2008-09</i>						
Central exam	4.372	6.249	8.936	4.236	5.983	7.142
School exam	5.919	6.736	8.700	5.824	6.560	8.258
Diff school & central exam	-1.337	0.488	3.539	-0.735	0.577	3.155
Subsidy (1000€ per capita)	0.000	0.000	0.000	0.339	1.420	3.962
Total population	5.687	8.841	9.932	8.056	8.938	9.459
Total number of students ^a	3.555	6.567	7.864	5.047	6.338	7.437
Income (in 1000€)	9.150	13.698	25.600	9.300	11.376	14.650
Share employed	39.000	69.268	85.000	48.333	57.734	70.000
Share welfare recipients	6.000	15.804	42.400	17.000	25.117	33.750
Urbaneness (5-point scale)	1	2.884	5	1	1.246	2.5
Immigrant population	2.703	19.167	86.142	28.843	54.786	78.358
Share younger than 25	12.108	23.072	41.272	21.505	26.282	34.518
Share older than 65	1.501	17.396	48.837	6.113	13.164	22.025

Note: ^a We only observe the total number of students for 523 schools.

Table 2: Pre-treatment (i.e., 2004-2007) trend in exam grades across subgroups

	Central Exam Grade	School Exam Grade
Power district (yes = 1)	-0.249 (-1.51)	-0.267 *** (-2.84)
Period 2006-2007 (yes = 1)	-0.203 *** (-4.08)	-0.054 * (-1.89)
Power district * Period 2006-2007	-0.009 (-0.04)	0.061 (0.46)
Number of schools included	636	636

Note: Dependent variables are central ($CE_{i,t}$) and school ($SE_{i,t}$) exam grades, respectively.

Entries are coefficient estimates from a random effects model comparing the 2004-2005 period to the 2006-2007 period; 'Power district' is an indicator variable separating schools in treated from those in untreated districts; t-statistics in brackets; ***, **, * significant at 1%, 5% and 10%, respectively. Adding district population size, school's student number and the share of immigrant, young and old residents as control variables leaves these findings unchanged. The same holds when estimation the model with fixed, rather than random, effects.

Table 3: Regression results using full sample (N=738 schools)

	Results of equation (10)			Results of equation (11)		
	(1) SE_CE	(2) SE	(3) CE	(4) SE_CE	(5) SE	(6) CE
Power district * Time (β_2)	0.133 (1.59)	0.042 (0.90)	-0.091 (-1.15)	0.469 *** (3.27)	0.103 (1.29)	-0.366 *** (-2.68)
Investment (1000€/cap; β_3)	-	-	-	-0.222 *** (-2.87)	-0.041 (-0.94)	0.181 ** (2.47)
Period 2 (yes = 1; β_1)	0.172 *** (9.15)	-0.015 (-1.47)	-0.187 *** (-10.52)	0.172 *** (9.123)	-0.015 (-1.47)	-0.187 *** (-10.59)
Population (log)	0.565 (1.05)	0.260 (0.87)	-0.305 (-0.60)	0.520 (0.97)	0.252 (0.84)	-0.267 (-0.53)
Students (log)	0.936 (1.51)	0.445 (1.28)	-0.491 (-0.83)	0.941 (1.53)	0.446 (1.29)	-0.496 (-0.85)
Students ² (log)	-0.072 (-1.46)	-0.030 (-1.10)	0.042 (0.89)	-0.074 (-1.51)	-0.031 (-1.11)	0.043 (0.93)
Immigrants (%)	12.227 ** (2.44)	4.531 (1.62)	-7.696 (-1.62)	11.353 ** (2.27)	4.371 (1.56)	-6.982 (-1.47)
Immigrants ² (%)	-13.909 * (-1.79)	-5.989 (-1.38)	7.979 (1.07)	-12.199 (-1.58)	-5.675 (-1.30)	6.524 (0.89)
Young (%)	-1.987 (-0.56)	-4.129 ** (-2.07)	-2.141 (-0.63)	-1.689 (-0.48)	-4.074 ** (-2.04)	-2.386 (-0.71)
Old (%)	-0.260 (-0.10)	-1.061 (-0.73)	-0.801 (-0.32)	-0.062 (-0.02)	-1.024 (-0.71)	-0.962 (-0.39)
Fixed effects (γ_i)	Yes	Yes	Yes	Yes	Yes	Yes
F (joint sign)	11.99 ***	1.64 *	14.54 ***	11.77 ***	1.56	13.83 ***

Note: Dependent variables are school ($SE_{i,t}$) and central ($CE_{i,t}$) exam grades as well as their difference ($SE_{i,t}-CE_{i,t}$). Entries are coefficient estimates from the fixed effects model given in equations (10) and (11) comparing the 2004-2006 period to the 2008-2009 period; 'Power district' is an indicator variable separating schools in treated from those in untreated districts, while 'investment' refers to the annual additional public investment (in 1000€ per capita) from the policy program; t-statistics in brackets; ***, **, * significant at 1%, 5% and 10%, respectively.

Table 4: Regression results using restricted samples based on population characteristics

	SE_CE	SE	CE	SE_CE	SE	CE	SE_CE	SE	CE	SE_CE	SE	CE
	Restricted migrant share			Restricted population size			Restricted young population			Restricted elderly population		
Power district * Time (β_2)	0.319 * (1.96)	0.110 (1.18)	-0.210 (-1.55)	0.347 ** (2.05)	0.089 (1.17)	-0.258 * (-1.89)	0.401 *** (2.65)	0.123 * (1.68)	-0.278 ** (-2.16)	0.392 ** (2.41)	0.108 (1.43)	-0.284 ** (-2.11)
Investment (1000€/cap; β_3)	-0.186 ** (-2.09)	-0.034 (-0.68)	0.151 ** (2.04)	-0.186 * (-1.94)	-0.046 (-1.07)	0.140 * (1.80)	-0.206 ** (-2.40)	-0.055 (-1.32)	0.151 ** (2.07)	-0.205 ** (-2.22)	-0.053 (-1.24)	0.152 ** (1.98)
Period 2 (yes = 1; β_1)	0.182 *** (3.82)	-0.044 (-1.62)	-0.226 *** (-5.70)	0.180 *** (7.90)	-0.003 (-0.33)	-0.183 *** (-9.97)	0.163 *** (6.95)	-0.019 * (-1.68)	-0.182 *** (-9.11)	0.164 *** (7.03)	-0.009 (-0.80)	-0.172 *** (-8.94)
N	163	163	163	674	674	674	539	539	539	612	612	612
F (joint sign)	4.75 ***	1.86 *	7.09 ***	9.47 ***	1.05	14.30 ***	8.30 ***	1.26	12.87 ***	8.15 ***	0.63	12.54 ***

Note: Dependent variables are school ($SE_{i,t}$) and central ($CE_{i,t}$) exam grades as well as their difference ($SE_{i,t} - CE_{i,t}$). Entries are coefficient estimates from the fixed effects model given in equation (11) comparing the 2004-2006 period to the 2008-2009 period; 'Power district' is an indicator variable separating schools in treated from those in untreated districts, while 'investment' refers to the annual additional public investment (in 1000€ per capita) from the policy program; Controls included as in table 3. t-statistics in brackets; ***, **, * significant at 1%, 5% and 10%, respectively.

Table 5: Regression results using restricted samples based on propensity scores

	SE_CE	SE	CE	SE_CE	SE	CE	SE_CE	SE	CE
	Excluding propensity scores <1% and >99%			Excluding propensity scores <5% and >95%			Excluding propensity scores <10% and >90%		
Power district * Time (β_2)	0.373 ** (2.27)	0.113 (1.39)	-0.259 * (-1.97)	0.433 ** (2.29)	0.094 (0.99)	-0.339 ** (-2.39)	0.467 * (2.01)	0.158 (1.38)	-0.309 * (-1.78)
Investment (1000€/cap; β_3)	-0.199 ** (-2.27)	-0.037 (-0.86)	0.162 ** (2.29)	-0.218 ** (-2.27)	-0.026 (-0.54)	0.192 ** (2.62)	-0.214 * (-1.89)	-0.017 (-0.30)	0.198 ** (2.34)
Period 2 (yes = 1; β_1)	0.164 *** (2.67)	-0.039 (-1.29)	-0.203 *** (-4.11)	0.150 * (1.79)	-0.031 (-0.72)	-0.181 *** (-2.82)	0.074 (0.69)	-0.070 (-1.34)	-0.144 * (-1.80)
N	200	200	200	138	138	138	104	104	104
F (joint sign)	3.33 ***	0.85	5.07 ***	2.65 ***	0.53	4.23 ***	1.53	0.76	2.25 **

Note: Dependent variables are school ($SE_{i,t}$) and central ($CE_{i,t}$) exam grades as well as their difference ($SE_{i,t} - CE_{i,t}$). Entries are coefficient estimates from the fixed effects model given in equation (11) comparing the 2004-2006 period to the 2008-2009 period; ‘Power district’ is an indicator variable separating schools in treated from those in untreated districts, while ‘investment’ refers to the annual additional public investment (in 1000€/per capita) from the policy program. Propensity scores obtained from a probit regression using population size and percentage immigrants (and its squared value) as well as the level of urbanization, the percentage of employed residents and welfare recipients (both as share of working-age population) and average income in the district (measured as after-tax income in 1000€, and its squared value) as explanatory variables. Controls included as in table 3 (except for Students, which is excluded here to maintain sufficient sample sizes). t-statistics in brackets; ***, **, * significant at 1%, 5% and 10%, respectively.

Table 6: Regression results of placebo estimation (2004-05, 2006-07)

	SE_CE	SE	CE	SE_CE	SE	CE
	Comparing 2004-05 to 2006-07 (placebo)			Comparing 2004-06 to 2008-09 (treatment)		
Power district * Time (β_2)	0.193 (1.35)	0.158 (1.16)	-0.035 (-0.17)	0.478 *** (3.09)	0.114 (1.31)	-0.364 ** (-2.46)
Investment (1000€/cap; β_3)	-0.106 (-1.41)	-0.068 (-0.95)	0.038 (0.36)	-0.211 ** (-2.45)	-0.048 (-1.00)	0.162 ** (1.97)
Period 2 (yes = 1; β_1)	0.178 *** (10.15)	-0.063 *** (-3.76)	-0.241 *** (-9.85)	0.173 *** (8.74)	-0.017 (-1.50)	-0.190 *** (-10.03)
N	635	635	635	636	636	636
F (joint sign)	12.26 ***	3.75 ***	11.99 ***	10.95 ***	1.67 *	12.65 ***

Note: Dependent Variables are school ($SE_{i,t}$) and central ($CE_{i,t}$) exam grades as well as their difference ($SE_{i,t} - CE_{i,t}$). Entries are coefficient estimates from the fixed effects model given in equation (11); 'Power district' is an indicator variable separating schools in treated from those in untreated districts, while 'investment' refers to the annual additional public investment (in 1000€ per capita) from the policy program; Controls included as in table 3. t-statistics in brackets; ***, **, * significant at 1%, 5% and 10%, respectively.

Appendix A: Alternative specification of the exam result function

Apart from the setting presented in the main text, we consider a second possibility to model the functional form of the function describing the overall exam results:

$$y = \underbrace{xe}_{=c(e,x)} + \underbrace{\frac{1}{n}(1+xe)}_{=s(e,x,n)}. \quad (12)$$

Again, both the central and the school exam results can be influenced by the students' effort choice. Different from before is that students now receive $\frac{1}{n}$ points on the school exam that may result from knowing the kind of questions the teacher might ask. Alternatively, it can be interpreted as reflecting the lower average difficulty of school exams (compared to central exams). The students' maximization problem in equation (2) then gives the following results:

$$e^* = \alpha - \frac{1-\alpha}{x(1+n)},$$

$$\frac{\partial e^*}{\partial n} = \frac{1-\alpha}{x(1+n)^2} > 0 \quad \text{and}$$

$$\frac{\partial e^*}{\partial x} = \frac{1-\alpha}{x^2(1+n)} > 0.$$

The main difference in results from those presented in the main text is that a change in education expenditures now has a positive effect on student effort. Apart from that, the effects in the first-order condition are qualitatively similar as above.

The school's first-order condition for the optimal grading standard and the cross-derivative with respect to x now become:

$$\begin{aligned} \frac{du^{SCH}}{dn} = & x \frac{\partial e^*}{\partial n} - \frac{1}{n^2} - \frac{1}{n^2} x e^* + \frac{1}{n} x \frac{\partial e^*}{\partial n} \\ & - 2 \left(x e^* - \frac{1}{n} - \frac{1}{n} x e^* \right) \left(x \frac{\partial e^*}{\partial n} + \frac{1}{n^2} + \frac{1}{n^2} x e^* - \frac{1}{n} x \frac{\partial e^*}{\partial n} \right) \end{aligned} \quad (13)$$

and

$$\begin{aligned} \frac{\partial^2 u^{SCH}}{\partial n \partial x} = & -\alpha \frac{1}{n^2} - 2 \left(\alpha - \frac{\alpha}{n} \right) \left(\frac{1-\alpha}{(1+n)^2} \left(1 - \frac{1}{n} \right) + \frac{1}{n^2} \left(1 + x \left(\alpha - \frac{1-\alpha}{x(1+n)} \right) \right) \right) \\ & - 2 \left(-\frac{1-\alpha}{1+n} \left(1 - \frac{1}{n} \right) + x \alpha \left(1 - \frac{1}{n} \right) - \frac{1}{n} \right). \end{aligned} \quad (14)$$

It is easy to see that the effects in equations (13) and (14) are qualitatively similar to the ones in equations (5) and (8) presented in the main text. An increase of n leads to better results in the central exam but lowers students' grades in the school exam. The difference between the results decreases in the grading standard (if $n < 1$ is assumed, which is equivalent to the assumption of $n^c > n^s$ in the main text). Higher educational spending strengthens the negative effect tougher grading has on school exam results, but also leads to a larger decrease in the results' difference. Again, the overall effect is dependent on the relative strength of both effects, but, as before, grade inflation following increased public education expenditures remains a theoretical possibility.

Appendix B: Students with differing abilities

In section 3, we abstracted from heterogeneity in student ability. We can, however, extend the model by assuming that students are equipped with different learning abilities a_j – taken from a distribution function $F(a)$ with associated density function $f(a)$. These abilities, which cannot be altered by education, positively influence the productivity of effort and thereby the overall exam result:

$$y_j = p_c^0 - n^c + \frac{1}{n^c} x e a_j + p_s^0 - n^s + \frac{1}{n^s} x e a_j. \quad (15)$$

As before, students choose their learning effort given the grading standards on both exams according to equation (3), which results in:

$$e_j^* = \alpha + (1 - \alpha) \frac{n^c n^s}{x a_j}. \quad (16)$$

Equation (16) shows that effort decreases in abilities $a_j \left(\frac{\partial e_j}{\partial a_j} < 0 \right)$, which reflects the idea that students can obtain the same grade by investing less time in learning and thus opt for an increase in leisure. In contrast, tougher grading on either the school or the central exam increases effort independent of a_j , in line with the basic model.

As exam results now depend on individuals' abilities, the different abilities and exams results are weighted with their density and thus enter the school's objective function. As a result, the school's maximization problem becomes:

$$\begin{aligned} \max_{n^s} u^{SCH} &= \bar{y}(e^*) - (\bar{c}(e^*) - \bar{s}(e^*))^2 \\ &= \int_{\underline{a}}^{\bar{a}} f(a_j) y(e_j^*) da_j - \left(\int_{\underline{a}}^{\bar{a}} f(a_j) c(e_j^*) da_j - \int_{\underline{a}}^{\bar{a}} f(a_j) s(e_j^*) da_j \right) \end{aligned}$$

where the integrals stand for the average results on the overall, central and school exams, respectively.²² The first-order condition for the optimal grading standard reads:

$$\begin{aligned} \frac{du^{SCH}}{dn^s} &= \frac{\partial \bar{y}(e^*)}{\partial n^s} - 2(\bar{c}(e^*) - \bar{s}(e^*)) \left(\frac{\partial \bar{c}(e^*)}{\partial n^s} - \frac{\partial \bar{s}(e^*)}{\partial n^s} \right) \\ &= \int_{\underline{a}}^{\bar{a}} f(a_j) \left(-\alpha \left(1 + \frac{\alpha a_j}{n^{s^2}} \right) \right) da_j - 2 \left(\int_{\underline{a}}^{\bar{a}} f(a_j) \left((2 - \alpha)(n^s - n^c) + \alpha x a_j \left(\frac{1}{n^c} - \frac{1}{n^s} \right) \right) da_j \right) \\ &\quad \cdot \left(\int_{\underline{a}}^{\bar{a}} f(a_j) \left(2 - \alpha + \frac{\alpha x a_j}{n^s} \right) da_j \right) = 0. \end{aligned} \quad (17)$$

Again, an increase of the school's grading standard decreases the average exam result \bar{y} , as shown by the first integral of equation (17). The second term is positive, which implies a decline of the difference between average results on the central and school exams across all abilities. Assuming $\bar{c} - \bar{s} < 0$ holds, the two opposing effects provide an inner solution and determine the grading standard n^s .

To investigate the existence of grade inflation, the implicit function theorem (equation (7)) is employed again. Whether an increase in education expenditures induces deteriorating grading standards depends on the sign of the cross-derivative

$$\begin{aligned} \frac{\partial^2 u^{SCH}}{\partial n^s \partial x} &= \frac{\partial^2 \bar{y}}{\partial n^s \partial x} - 2 \left(\left(\frac{\partial \bar{c}}{\partial x} - \frac{\partial \bar{s}}{\partial x} \right) \left(\frac{\partial \bar{c}}{\partial n^s} - \frac{\partial \bar{s}}{\partial n^s} \right) + (c - s) \left(\frac{\partial^2 \bar{c}}{\partial n^s \partial x} - \frac{\partial^2 \bar{s}}{\partial n^s \partial x} \right) \right) \\ &= - \int_{\underline{a}}^{\bar{a}} f(a_j) \frac{\alpha a_j}{n^{s^2}} da_j - 2 \left(\int_{\underline{a}}^{\bar{a}} f(a_j) \left(\alpha a_j \left(\frac{1}{n^c} - \frac{1}{n^s} \right) \right) da_j \right) \left(\int_{\underline{a}}^{\bar{a}} f(a_j) \left(2 - \alpha + \frac{\alpha x a_j}{n^{s^2}} \right) da_j \right) \\ &\quad - 2 \left(\int_{\underline{a}}^{\bar{a}} f(a_j) \left((2 - \alpha)(n^s - n^c) + \alpha x a_j \left(\frac{1}{n^c} - \frac{1}{n^s} \right) \right) da_j \right) \left(\int_{\underline{a}}^{\bar{a}} f(a_j) \frac{\alpha a_j}{n^{s^2}} da_j \right). \end{aligned} \quad (18)$$

²² The number of households is normalized to unity.

Similar to the basic model, equation (18) shows that an increase in expenditures on the one hand raises the school's incentive for grade inflation because the positive effects of a lower grading standard on y get stronger in x . On the other hand, the decrease of the difference $(c-s)^2$ is amplified in x as well, thereby strengthening the incentive to increase n^s . Thus, extending the model by considering students of differing abilities does not alter the model's results qualitatively. Still, grade inflation following an expenditure increase remains a possible outcome.