The Measurement and Communication of Effect Sizes in Management Research

Carl F. Fey
Aalto University School of Business
Carl.Fey@aalto.fi

Tianyou Hu
School of Business, Maynooth University
Tianyou.Hu@mu.ie

Andrew Delios
National University of Singapore
andrew@nus.edu.sg

January 2, 2022

Forthcoming, *Management and Organization Review*

The Measurement and Communication of Effect Sizes in Management Research

ABSTRACT

The measurement and communication of the effect size of an independent variable on a dependent variable is critical to effective statistical analysis in the Social Sciences. We develop ideas about how to extend traditional methods of evaluating relationships in multivariate models to explain and illustrate the statistical power of a focal independent variable. Even with a growing acceptance of the need to report effect sizes, scholars in the management community have few well-established protocols or guidelines for reporting effect sizes. In this editorial essay we: 1) review the necessity of reporting effect sizes; 2) discuss commonly used measures of effect size and accepted cut-offs for large, medium, and small effect sizes; 3) recommend standards for reporting effect sizes via verbal descriptions and graphical presentations; and 4) present best practice examples of reporting and discussing effect size. In summary, we provide guidance for authors on how to report and interpret effect sizes, advocating for rigor and completeness in statistical analysis.

Keywords: effect size, statistical analysis, statistical reporting, confidence interval, data visualization

INTRODUCTION

Quantitative empirical research in the management and organization literature emphasizes identification of directional associations between independent and dependent variables. Researchers work to establish evidence supportive of claims of causality between two variables in a population. This focus directs researchers to give attention to interpreting associations based on the level of certainty that a proposed relationship exists as reflected in variances around point estimates in a coefficient. This variance is often interpreted as a level of statistical significance or a 'p value'. Reporting standards are changing from the need to show whether p values are lower than accepted levels of so-called statistical significance (e.g., 0.05, 0.01) towards a norm of reporting precise p values and discussing the implications, without resorting to hard cut-offs.

Aside from levels of certainty in the precision of a point estimate, another important aspect to consider is the magnitude of the association between an independent and dependent variable in an estimated relationship. This aspect of statistical reporting is commonly known as the effect size and sometimes also referred to as the economic significance or substantive importance of a relationship. Olejnik and Algina (2003: 434) define effect size as an index that "estimates a parameter that is independent of sample size and quantifies the magnitude of the difference between populations or the relationship between explanatory and response variables." Interestingly, even given a growing awareness of its importance, effect size reporting is rare in management research. Part of the reasons for the slow uptake of effect size reporting is that management researchers still lack clear guidance on how to report and discuss effect size. It is our goal to provide this form of guidance.

In contemporary management research, it has become more common to use large samples as such datasets are increasingly available. With the greater power that comes from larger sample

sizes, it is more likely, or some could say easier, to identify associations between variables of interest with a small value of *p*. However, this increased power does not equate to an indication that the independent variable has a larger substantive impact on the dependent variable. One advantage of effect size is that it is not influenced by sample size. Instead, there are substantial differences in the implications for theory and practice between a 1% and a 20% change in the dependent variable, and both can be a consequence of the same level of change in an independent variable. In other words, it is increasingly important to identify the extent a change in an independent variable leads to a change in the dependent variable. If researchers only focus on reporting significance analysis, they lose the opportunity to inform scholars and practitioners about level and importance of the influence of focal variables, and thus fail to demonstrate the usefulness of a study.

As management theory and research continue to mature, we strongly contend that the research community needs to move to research that can translate into results that have more substantive meaning for the practice of management. As such, we are confident to make the statement that management researchers must give due consideration to effect size, as reflected in the level of influence that focal independent variables have on the dependent variable(s).

Connected to this statement, increasingly management journals are requesting authors to report effect size. Yet, clear guidance on how to develop and implement effect size analysis is lacking. Moreover, effect size reporting is still not commonly practiced in most present-day publications in the management literature. As such, in this editorial essay, we develop and present best practice approaches to help guide emergent standards for reporting effect sizes.

In our discussion of effect size calculation and reporting, we arm researchers with a variety of techniques for determining effect sizes. We believe that researchers should not only

make textual reports on the magnitude of the effect size in numerical terms, but also use data visualization techniques such as graphs and diagrams to provide intuitive and informative depictions of effect sizes.

In the remainder of this editorial essay, we accomplish three objectives. First, we walk through the steps required to develop good estimates of effect size. By deliberate design we avoid discussion of deep technicalities in the measures, as our intent is to foster the adoption of reporting effect size in management research. Second, we identify the ways in which effect sizes can be depicted to maximize their clarity of impact using textual and graphical descriptions. Third, we identify several articles, which we use as examples of best practice in terms of presenting and discussing effect size. By accomplishing these three objectives, we present our core recommendation that accurate and sufficient reporting of effect size should be standard in reporting statistical analysis in management research.

FROM NULL HYPOTHESIS TESTING TO ESTIMATING EFFECT SIZE

Prevailing research traditions in management scholarship have led to a focus on exploring the nature of the association between independent and dependent variables. As part of this focus, and albeit it is less common in current research, scholars often conflate p values as a measure of both the existence and the strength of a relationship. However, p values do not yield information on the strength of a relationship.

Scholars face limitations when using p values as the sole indicator of associations between variables. First, p values are dependent on the sample size. One general property is that p values tend to decrease with increases in sample size. Within very large samples, finding coefficient estimates with low p values is often of little surprise (Sullivan & Feinn, 2012).

Second, a p value does not measure the strength of a relationship. If X_1 and X_2 are independent variables and they have p values of 0.06 and 0.03 respectively in their relationship with the dependent variable Y, it does not mean that X_1 has a larger effect size or a more substantive impact on Y than X_2. Third, an independent variable with a low p value might not have a substantive influence on the dependent variable. In other words, we can have confidence that the observed association exists due to the demonstrated low p value, but any movement on the independent variable might only lead to trivial changes on the dependent variable. Effect size analysis helps identify the level of change or what we can call the practical significance.

Perhaps most troubling from the focus on p is the emergence of the notorious practice of p-hacking. With a greater availability of large data sets and increasing ease of running increasing sophisticated models, p-hacking has become more common (Baum and Bromiley, 2019; Broder, Cook, and Heyes, 2020). Understandably journal editors and researchers are concerned about the problem as p-hacking leads to practices such as opportunistic modifications to models, or limited reporting of models. As an example, a researcher can play with various combinations of control variables, to identify the specification that moves the independent variable of interest from being above $p=0.05$ to being below $p=0.05$.

Increasingly p-hacking has been identified as an important and common problem in management research (Baum and Bromiley, 2019). It can be damaging to the stability and replicability of findings (Lewin et al., 2016). Unfortunately, it is difficult to know whether p-hacking has occurred in a specific paper. One advantage of effect size measurements is that they are not prone to change with increases or decreases with sample size. By promoting both reporting of statistical significance and the practical significance (effect size) of research, we can reduce the stress that authors face to find that elusive $p<0.05$, as even that p value might not have

a large effect size. As a result, reporting effect size might help decrease the prevalence of *p*-hacking.

The recent management literature has identified well these aforementioned limitations in the use and interpretation of *p* values (Meyer, van Witteloostuijn & Beugelsdijk, 2017). Alongside this development has been a push to gain stronger causal inference in research design. An increasing number of management scholars are now implementing research designs that use experimental methods such as quasi-experiments or randomized controlled trials. In addition, there has been increasing attention given to matched pairs or matched samples analyses, in which the variable(s) of interest can be isolated more effectively and then explored using difference-in-difference analytical techniques, for example. These techniques align strongly with the dictates of establishing evidence in support of causality, more so than can be obtained by using lagged variables in multivariate analysis. Yet, even though we have seen these changes in research designs and methods, we still see a substantial body of researchers using cross-sectional data either obtained via primary sources such as in-person or paper (online) mediated surveys, or from secondary source such as archival databases.

The advancements in research methods to more acutely tease-out causal relationships has permitted researchers to design and report research more effectively. Clearly, *p* values still carry important information for good statistical inference, however, we contend that the next stage of improvement in reporting on one's econometric analyses should be to calculate and communicate effect sizes. Effect size reporting should ideally become an essential part of the communication about a study's empirical results.

Importantly, we note that this consideration of effect sizes moves beyond the treatment effect considerations we discussed briefly above, where researchers attempt to show causality.

By this we mean that even within the consideration that management research has been making advances in terms of the methods by which to develop and show consistent and reliable empirical evidence that A causes B, it has been doing so within the confines of ruling out, so to speak, alternative explanations, such that C or D or E cannot be advanced as reasonable alternative causes of B. As much has been written about better causal inference, we focus our attention on the next steps of improvements in the research process.

Put another way, the starting point in our discussion is that we assume that the researcher has utilized up-to-date, accepted techniques for establishing reasonable causal inference for A to B. Given that the causal inference is acceptable, we then move to the importance of how to determine, report, describe, and depict the effect size of a relationship as an aid to understanding if the relationship has a large enough substantive impact to make a useful contribution to management theory and practice. Effect size is a useful complement to causal inference as it helps to show that an issue of interest has a large enough influence to be of interest to scholars and practitioners.

In terms of the background for some of our claims, the fields of psychology and education were early movers in the social sciences for promoting reporting effect sizes when presenting empirical results. *Educational and Psychological Measurement* made its first call for reporting effect size in 1994 (Thompson et al., 1994). Further, the American Psychological Association's task force on statistical inference followed up its initial 1994 recommendations in 1999 by stating that researchers should "always present effect sizes for primary outcomes….interval estimates should be given for any effect sizes involving principal outcomes." (Wilkinson, 1999: 599). The APA has continued to stress the importance of reporting effect size (APA, 2020). The American Educational Research Association (2006) has also

advocated the reporting of effect size. It is, however, worth noting that while calls for reporting effect size in psychology and educational research have increased substantially, effect sizes have not been as consistently reported and well interpreted even in psychology and education articles as desirable. Meanwhile, journals such as the *Journal of Applied Psychology, Educational and Psychological Measurement,* and *Psychological Science* have been early movers in requiring reporting effect size (Kelly and Preacher, 2012).

Management journals have not been opaque to issues around effect size reporting when interpreting the managerial meaning of results. For example, Bettis, Ethiraj, Gambardella, Helfat & Mitchell (2016) in the *Strategic Management Journal* and Ellis (2010) in the *Journal of International Business Studies*, the guide for authors for *Organizational Behavior and Human Decision Processes* (OBHDP Editors, 2021), and Lewin et al. (2016) in *Management and Organization Review* have advocated reporting effect size. However, a clear guide with sufficient direction on when and especially how to report effect size has been lacking in management research. This has led to the situation where management research has fallen behind other Social Science disciplines in terms of informing audiences about the substantive meaning of the results within its empirical research. Critically, without a sufficient analysis or consistent reporting of effect size, management research has not been able to identify fully the practical implications of its research on individuals and on organizations. As such, in the management field there is a lack of a clear understanding of where and how advice can be derived for practitioners. Studies fail to point to where practitioners could focus their efforts to improve the performance of objectives of interest as predicated upon an understanding of the nature of causality and the strength of a relationship. Focussing on effectively identifying and reporting effect sizes can help to rectify this persistent shortcoming in management research.

A recent anecdote helps illustrate how potential is currently being realized in internal research departments in large organizations. In the late 2010s, at the American Economics Association Meetings, Amazon was one of the most prolific recruiters of new PhD economists. The reason for Amazon's recruiting drive was that PhD economists have developed empirical skills by which they were not only able to develop models with strong causal inference, but they were also able to develop predictions about how strongly an independent variable X could affect a dependent variable Y. With good causal inference, and with good methods for identifying and reporting effect sizes, measurements of the effectiveness of marketing campaigns or other such initiatives could be distilled into a set of defensible numbers based on the application of contemporary econometric techniques to the very large data that Amazon possessed. We need to ensure that management scholars are equipped with and can use analytical tools to explore substantive impacts of manipulations of X to Y, such as what Amazon and other firms are doing with their economist PhD recruits.

MEASURING EFFECT SIZE

A core element to the measurement of effect size is the estimated coefficient in a regression estimation. In a linear regression, coefficient estimates represent the predicted change in the dependent variable corresponding to a unit change in the independent variable, *ceteris paribus*. For example, in a study of Standard & Poor's 1500 firms in the 1992 to 2006 period, Dezsö and Ross (2012) find that "...[a female member sits] in top management leads to an increase of $42 million in firm value…." Effectively, this reporting of the substantive nature of the effect size of female representation provides a clear interpretation of the outcome in terms of firm value.

When it comes to non-linear relationships, the calculation is more complicated. Sampson (2007: 377) employed a negative binomial model to identify a U-shaped relationship between a firm's technological diversity as measured by its alliances, and its innovative outcomes, as measured by its patents. In the reporting of the effect size of results, we find the statement that "…moderately diverse alliances contribute over 13 times more than alliances with very low diversity… and over 3 times more than alliances with very high diversity…."

Sometimes, researchers can use mean differences between groups to compare effect sizes. David Card, a Nobel Laureate in Economics in 2021 reported in his classic study on minimum wage and employment that: "… the rise in the California minimum wage reduced the fraction of California worker earning \$3.35-4.24 per hour by some 5 percentage points…[and] raised the wage 15% for 5% of California workers" (Card, 1992: 42).

These three examples clearly illustrate how researchers can provide an indication of the practical significance of their research. This communication can aid managers and policy makers in the design of their organizations or programs for developing evidence-based initiatives. The reporting found in the above comes from calculations based on the substitution of coefficient estimates into the relevant regression equation to determine the prediction of the dependent variable, given the relevant unit changes in the independent variables. We note that this is an important way of reporting effect sizes in research which we recommend. It is what we identify as describing the effect size in words, but we also want to draw attention to other ways in which research has developed measures for effect size.

Table 1 depicts our summary of measures of effect size. We note there are over 40 measures of effect size, which challenges researchers to know which is the correct one to use in a particular circumstance. Our Table 1 aids this choice by identifying the effect size measure we

recommend according to the type of commonly used statistical analysis employed in the research. Table 1 provides our effect size related recommendations for numerous commonly used statistical analyses in management research. These effect size measures are developed based on the statistical attributes of the sample used in an analysis and its corresponding statistics. To help convey the meaning of different effect size measures, we identify what constitutes a small, medium, and large effect size for each measure. Table 1 is rich in information, making it a key reference for management researchers. We describe several specific areas of Table 1 as follows.

INSERT TABLE I ABOUT HERE

As shown in Table 1, researchers have used estimated mean differences (Nakagawa & Cuthill, 2007; Rosenthal, Cooper & Hedges, 1994), correlations (Cohen, 1992), variances (Fern & Monroe, 1996), and regression estimates (Cohen, 1988; Sullivan & Feinn, 2012) to measure and report effect size. Accordingly, a clear implication of the information in Table 1 is that researchers should choose a measure of effect size according to their research design and the statistics that are available.

As an example, researchers can use Cohen's *d* as an effect size measure when the independent variable is categorical, and the dependent variable is continuous. Pearson's *r* can be used as an effect size measure when both the independent variable and the dependent variable are continuous.

When considering effect sizes, a natural question to ask is what constitutes a large, medium, and small effect size. Cohen was one of the earliest scholars to develop the concept of effect size. His work remains important to the present day. Among Cohen's (1998, 1992)

12

impression-based guidelines for different type of statistics, his recommendations for classifications of effect sizes for mean differences and correlations are especially widely cited today. For a comparison of mean differences, effect sizes between 0.20 and 0.49 are said to be small, effect sizes between 0.50 and 0.79 are medium, and effect sizes of 0.80 or greater are large. For a Pearson *r*, a correlation of 0.10 to 0.29 is a small effect size, a Pearson *r* value of 0.30 to 0.49 is classified as a medium effect size, and a Pearson *r* value of 0.50 or greater is a large effect size.

To provide some context of what these different effect sizes mean Cohen (1988) suggested that an effect size of 0.2 would be yielded by comparing the height of 15- and 16-year-old girls in the United States. In other words, the effect size is not very large or easy to observe. He further suggested that unlike the 15- and 16-year-old comparison, a phenomenon leading to a 0.50 (medium) effect size should be visible to the naked eye. Cohen also warned he was trying to provide approximate guidelines that could be useful across different behavioral sciences. However, interpretations of effect sizes should also consider the limitations in using specific cut-offs to identify the magnitude of an effect size.

Cohen (1988) recommended that researchers consider the levels that have been observed and deemed reasonable in past studies in a specific area, as an additional point for identifying the size of an effect size. Relatedly, studies such as Sun, Pan, and Wang (2010) warn against a blind reliance on effect size cut-offs. Just like the critique related to *p* values that the difference between $p=0.05$ and $p=0.06$ is just 0.01 and should not substantially change the interpretation of a statistical analysis, effect sizes should be considered as a continual evolution from small to large with cut-offs providing guidance.

This point connects to the reasons why Cohen (1988: 352) offered benchmark guidelines. He did so "because they were needed in a research climate characterized by a neglect of attention to issues of [effect size] magnitude." Such a climate persists today given that effect sizes are not discussed in most published research. Thus, the recommended cut-offs for effect size continue to serve a useful function, but the context in which the research was conducted should be considered in their application.

It is useful to consider how well Cohen's (1988, 1992) cut-offs apply to recent research. Based on a review of two meta-analyses in psychology, Hemphill (2003) found that one-third of the correlations in the meta-analysis were less than 0.20, one-third were between 0.20 and 0.30, and one-third were greater than 0.30. Hemphill (2003) proposed accordingly these three ranges be recast as small, medium, and large effect sizes for correlations, which are similar but slightly lower thresholds than those recommended by Cohen (1988, 1992). More recently, Gignac and Szodorai (2016) found that only 2.7% of correlations were 0.50 or larger, which suggests that this previous cut-off of 0.50 (Cohen, 1988, 1992) seems to be too high for a large effect size. As such, they suggest that a small effect size should be 0.10 to 0.19, a medium effect size should be 0.20 to 0.29 and a large effect size should be 0.30 or larger. Given the realities of observed inter-item correlations found in most management research, the recommendations for the qualitative categorization of effect size by Gignac and Szodorai (2016) have the most substantive and analytical traction.

Next, Acock (2014: 272) suggested that standardized regression coefficients can be considered at similarly corresponding levels as those we have discussed for Pearson correlations. More specifically, he suggests that a standardized β less than 0.2 should be considered as a weak association between the independent and dependent variables; whereas a β that is greater than 0.2

but less than 0.5 be considered as moderate, and a β greater than 0.5 should be considered a strong effect.

It is worth noting that caution should be exerted when considering β as a measure of effect size, as various factors affect the standard deviation of a variable and standard deviations are used to calculate a β. A common measure of the collective effect size of all the explanatory variables (the global effect size) in a multiple regression is Cohen's f^2 which is $R^2/(1-R^2)$. Cohen (1992) states that small effect sizes for Cohen's f^2 are between 0.02 and 0.14; medium are between 0.15 and 0.34; and large effect sizes are greater than or equal to 0.35. In their thirty-year-observation review of categorical moderators, Aguinis et al. (2005) found that Cohen's f^2 averaged only 0.002. However, they found that 72% of the moderator tests had a power of 0.80 or more, which provides some evidence that smaller effect sizes for categorical moderators may be reasonable to consider.

Cohen's f^2 can also be modified to measure the effect size of an independent variable of interest (see Table 1 for details). This is called Cohen's incremental f^2 and has the formula of: $f^2 = (R^2_{included} - R^2_{excluded})/(1 - R^2_{included})$. It has the same effect size cut-offs as mentioned for the global/regular Cohen's f^2 above. This effects size measure can be helpful when building models incrementally or when one wants to identify the amount of information or the practical significance of a focal variable.

We want to point out that some effect size measures are standardized like β and f^2. We normally advocate for the use of effect size measures that are standardized as they facilitate comparison across variables and studies. However, a limitation of such measures is that they can be less intuitive to understand due to not being presented in conventional units. Thus, it can be useful for researchers to also present non-standardized effect size measures to better understand

results from a particular study. Further, researchers are encouraged to report both local effect size measures that focus on the effect of one variable, such as by using Cohen's incremental f^2 for key independent variables of interest, and global effect size measures like Cohen's f^2 for an entire model.

Cautions on Using Effect Size

First, researchers should keep in mind that the estimation of effect sizes is dependent on a *ceteris paribus* assumption. This assessment of biases and the likelihood that a bias exists is not disconnected from how accurate a measure of effect size is, as appropriate estimation will minimize the chances for substantial biases to emerge. We need to be primarily concerned about whether the specification we employ under-estimates or over-estimates the value of a coefficient.

Second, the estimation of effect size will be always dependent on the specifications for the model which are determined by the researchers. A second consideration emerges from the consideration of the specification in that a coefficient estimate, or the meaning of any variable, will vary depending upon the nomological network in which it is embedded. A nomological network is a representation of latent constructs, the observable items that measure them, and the relationships between them. In other words, we do not derive estimates of effect sizes from simple correlation coefficients, which can also be conceived as an under-specified model. As such, the specification of the model will influence the values that coefficient estimates take in the analysis and consequently how much increasing X by one standard deviation, leads to x units of increase in Y.

Third, researchers should learn to interpret and explain effect size for readers. We note that researchers should strive to convey effect size statistics in representations of economic

significance to connect to managers or policy makers, such that they can understand the practical implications of a particular study. Critical to this step of communication is not only how is a result presented to be immediately and readily digestible to the appropriate audience, but also how do we present estimates of the margin of error around any given point estimate of effect.

The consequences of interpreting an effect, even as consistently positive within a given error band, can vary considerably depending on where in the range of plausible estimates the researchers gauge their effect size to be. As economists interpret the economic meaning of their findings; for example, to what extent an increase in the minimum wage can lead to a decrease in employment rates, management scholars also need to clearly articulate the extent to which strategic actions like an acquisition, an investment in human resource capital, or a decision to hire or fire an executive can bring as consequences to firm performance. Given that policies and managerial action can connect directly to these estimates of effect size, the considerations around this issue are of no small matter; that is, we need to treat the implications of effect size predictions and depictions seriously.

BEST PRACTICES IN REPORTING EFFECT SIZE

We make two recommendations on how researchers should report effect sizes: (1) they should include a detailed textual description in the paper; and (2) they should have a clear presentation in graphs. In the text of the paper, authors should report the estimated coefficients of a statistical regression, along with confidence intervals and/or standard errors of the coefficients. Using these statistics, researchers can explicitly describe the effect size, compare in qualitative terms (small, medium, large) and explain the substantive impact in words (Bettis et al., 2016; Ellis, 2010).

Textual descriptions not only communicate the academic findings of a study but also let readers associate academic findings with real-word experience.

A recent paper published in MOR by Li-Ying, Zhang, and Long (2018) is a good example of the implementation of this practice. Their research concerns how hidden knowledge facilitators, about whom a firm's employees may be unaware, can promote knowledge sharing within a firm. They report the estimated coefficients of the focal independent variable, *the quantity of facilitators*, to depict their findings about how facilitators help knowledge sharing. Thus, after establishing evidence that this relationship is reflected in their data and analyses, they go further to test the moderating effect of *the quality of facilitators* and report on page 808 that: "…The [standard mean difference] effect size for the moderating effect is -0.156 (with an estimation of [-0.209, -0.10] at the 95% confidence interval) …" With this effect size analysis, the authors announce a failure to finding a moderating effect.

Gupta and Misangyi (2018: 41) provide a good example of explaining effect size in words when they write, "Low (−1. SD) vs. high (+1. SD) values of recent international diversification actions by charismatic CEOs predicted a difference of 3% (2% for narcissistic CEOs) change in the DV, which, for a company with median sales of all firms in our sample, represents a difference of \$300 million (\$200 million for narcissistic CEOs) in revenues from international business segments." We also encourage researchers to report conventional effect size measures (e.g., *Cohen's d*, η^2, pseudo-R^2, and ω^2) as depicted in detail in Table 1. These conventional measures should be discussed and interpreted in plain, easily accessible (non-technical) language. Song et al. (2018) provide a good example of this practice in their study of how prosocial intervention on employees can reduce perceived customer mistreatments. They use a pseudo-R^2 to measure the effect sizes from the estimated coefficients in the model and

18

report on page 1005: "Predictors … accounted for 13% of the within-person variance in employee daily experience of customer mistreatment, 16% in afternoon negative mood, 24% in evening rumination…" Moreover, they use Cohen's d and ω^2 for cross validation. By providing rich information and by using multiple effect size measures, the authors show the power of their field experiments when they compare their findings with conventional standards and previous studies.

Finally, we suggest that authors indicate if the effect size corresponds to a small, medium, or large scale in established classifications. Song et al. (2018) again provides a good example as they refer to conventional cut-offs when reporting effect size measures: "…$\omega^2 = .10$, … [which] indicates a medium effect size (typically ranging from 0.06 to 0.15) for the repeated-measure design".

Next, the visualization of effect size is becoming an increasingly common practice in the reporting of results in research in management and organizations. Visualization in reporting effect size provides a graphical description, and it helps one to quickly understand what is being hypothesized and observed. Further, it enables readers to see beyond a conceptual model and more clearly recognize the substantive impact of the key findings of a study.

An important initial consideration before constructing a visual representation of the effect sizes of an analysis is to supplement information on point estimates with reported standard errors to identify confidence intervals for non-standardized regression coefficients. Reporting confidence intervals explicitly allows us to understand the level of precision in coefficient estimates. A common standard is to use a 95% confidence interval, which captures values within approximately +/- 2 times the standard error. When reporting a confidence interval, it is important to identify if the confidence interval includes 0. If it does, then we cannot confidently

state that we have a positive or negative relationship, or even a relationship at all. Alternatively, if we are plotting the effects of two positive or negative coefficients and we believe these to be different in value from one another, we do not want to see overlap in the confidence intervals as projected for both. We will see this point more clearly in the example we present later in this section.

Next, a good presentation using graphs also necessitates a consideration of the units that are used, and the quantities or scales depicted. Further, it provides a platform for researchers to demonstrate their creativity in reporting their findings. That said, we still believe that extant papers infrequently use graphical descriptions on effect size.

Chatterji et al. (2019) provide a good example in their research on whether start-up firms can benefit from receiving consulting advice from established entrepreneurs. They find start-ups can have an additional size growth rate of 28% as measured by the number of employees, and a 10% lower chance of failure, if firms receive advice from entrepreneurs (as compared to those who do not receive advice). The authors visually present the effect size of the main effects, which is the benefit of receiving entrepreneurial advice as related to the growth and survival of start-ups. They divide their presentation of effect sizes according to whether the firm is an MBA start-up or a non-MBA start-up (see Figure 2).

It is noteworthy that in this figure they provide point estimates with confidence intervals, which is an emerging but not common standard in graphical representations. As reprinted herein, Chatterji et al. (2019) is a best practice example of how to visualize effect sizes, which we recommend MOR authors to follow. Another good example is Vanneste and Gulati (2021) in their study of a firm's use of trust in coping with economic downturns. The authors provide graphs of effect sizes (Figure 3 of their paper) across regression models for the illustration of the

results of their four hypotheses. They include point estimates and confidence interval depictions, which provide an immediate and clear visualization of the effect sizes for readers.

INSERT FIGURE I ABOUT HERE

Within MOR, we note several relevant examples of effect size depictions. Chen and his colleagues (Chen, Dai & Li, 2019) provide a detailed discussion of how a firm's innovation output is associated with its affiliated R&D consortia. The authors used standard errors and confidence intervals in their discussion of the main effects and moderating effects. Prashantham, Zhou, and Dhanaraj (2020) report a global effect size for a Structural Equation Model and discuss effect sizes for quantitative empirical articles when reviewing the literature. Both steps are good practices that help readers to interpret their study and past research with accuracy. Another example is Hu, Zhang, and Yao (2018) who study the labor productivity of family firms. Their Figures 1 to 3 provide good examples of how to plot the effects of interaction terms, as related to effect size.

A CHECK LIST FOR REPORTING EFFECT SIZE

We strongly recommend authors to use both textual descriptions and visual aids to report and interpret effect sizes. We provide two check lists as guidance on how to implement this practice.

I. When reporting the effect size measures and discussing effect sizes in a textual description, researchers should:

1) Provide a description of the estimated coefficients;

2) Provide confidence intervals or report the standard deviations;

3) Provide measures of effect sizes, which can include a comparison of the effect sizes observed in relation to those found in previous related studies;

4) When appropriate, make reference to established norms for cut-offs to identify if an effect size is large, medium, or small;

5) Provide an interpretation of the effect size in words such as an interpretation of the economic meaning, if any.

6) Report both the effect of key independent variables of interest using effect size measures such as Cohen's incremental f^2, and the global effect size of the entire model using measures such as Cohen's regular f^2.

II. When presenting a visualization of effect size, researchers should:

1) Provide point estimates of effect sizes;

2) Provide a visual depiction of confidence intervals around the point estimates;

3) Develop fitted lines that link points and show the trend of estimates for the dependent variable along values of the independent variable;

4) All values within the figure should be within the ranges of the independent and dependent variables as found in the data;

5) The plots should be developed based on fitted values of the dependent variable. The calculation uses the estimated regression equation, where specific values in the range of the key independent variable are used, alongside the mean values of all other variables in the regression equation (Makino & Delios, 1996: 918).

6) Identify the confidence intervals around critical isolated point estimates found in the plots, resulting in a figure mirroring our reprint of Chatterji et al. (2018) as in our Figure 1.

These two check lists refer to ideal presentations of effect size, which should become part of standard practice in reporting. Admittedly, space can be scarce in journals. Researchers must balance the need to have effect size discussions in the main text of the paper versus the inclusion of other material, where at times effect size visuals might be presented in appendices. Kaiser et. al (2018) and Asgari et. al. (2018) recommend publishing graphs and critical statistics in appendices that are available online. In so doing, space constraints become less of a valid excuse for not reporting effect sizes. Researchers are thus empowered to provide a complete set of textual and graphical descriptions of effect sizes.

CONCLUSION

Effect size reporting is a critical step in the communication of the findings from empirical research. An effect size identifies the size and thus relevance of the relationships between the independent variable(s) and dependent variable(s) in a study. It can also convey information on the practical importance of the relationship. A large effect size has more practical importance than a small effect size because it indicates that the relationship being studied has real, substantive meaning for managers or policy makers.

In this editorial essay, we emphasized how effect size marks practical significance as an important complement to the ubiquitous reporting of statistical significance as found in reports of p values. We stress the importance of providing clear information on practical significance alongside the consideration of confidence intervals either for point estimates or for graphical

representations of results. Our Table 1 helps guide researchers to choose the correct measure of effect size with the choice of measure dependent on the type of statistical estimation employed. Figure 1 also provides guidance on cut-offs that are commonly used in research to identify small, medium, and large effect sizes. This coalescence also led to our presentation of best practice examples, where we introduce standards of reporting effect sizes as captured in words in the text of a paper or as represented by a figure. As such, our editorial essay yields clear and practical advice on how to standardize effect size reporting. We trust that these ideas can serve as a useful reference for management researchers.

REFERENCES

Acock, A. C. 2014. *A gentle introduction to stata,* Texas: Stata Press.

Aguinis, H., Beaty, J. C., Boik, R. J. & Pierce, C. A. 2005. Effect size and power in assessing moderating effects of categorical variables using multiple regression: a 30-year review. *Journal of Applied Psychology,* 90(1): 94.

Albers, C. & Lakens, D. 2018. When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias. *Journal of Experimental Social Psychology,* 74: 187-195.

American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications. *Educational Researcher*, 35(6), 33–40.

APA. 2020. *Publication manual of the American Psychological Association*, (7th ed.). Washington, DC: American Psychological Association.

Asgari, N., Tandon, V., Singh, K. & Mitchell, W. 2018. Creating and taming discord: How firms manage embedded competition in alliance portfolios to limit alliance termination. *Strategic Management Journal,* 39(12): 3273-3299.

Baum, J. & Bromiley, P. 2019. P-hacking in top-tier management journals. *Academy of Management Annual Meeting*. Academy of Management.

Bettis, R. A., Ethiraj, S., Gambardella, A., Helfat, C. & Mitchell, W. 2016. Creating Repeatable Cumulative Knowledge in Strategic Management. *Strategic Management Journal,* 37: 257-26.

Boßow-Thies, S. & Albers, S. 2010. Application of PLS in marketing: content strategies on the internet. In Vinzi, V. E., Chin, W. W., Henseler, J. & Wang, H. (Eds.) *Handbook of Partial Least Squares:* 589-604. Berlin: Springer.

Card, D. 1992. Do minimum wages reduce employment? A case study of California, 1987–89. *Industrial and Labor Relations Review,* 46(1): 38-54.

Chatterji, A., Delecourt, S., Hasan, S. & Koning, R. 2019. When does advice impact startup performance? *Strategic Management Journal,* 40(3): 331-356.

Chen, D., Dai, L. & Li, D. 2019. A delicate balance for innovation: Competition and collaboration in R&D consortia. *Management and Organization Review,* 15(1): 145-176.

Chin, W. W. 1998. The partial least squares approach to structural equation modeling. *Modern Methods for Business Research,* 295(2): 295-336.

Cohen, J. 1988. *Statistical power analysis for the behavioral sciences,* Hillsdale, NJ: Erlbaum.

Cohen, J. 1992. Quantitative methods in psychology: A power primer. *Psychological Bulletin.* Citeseer.

Dezsö, C. L. & Ross, D. G. 2012. Does female representation in top management improve firm performance? A panel data investigation. *Strategic Management Journal,* 33(9): 1072-1089.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B. & Burke, M. J. 1996. Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods,* 1(2): 170.

Ellis, P. D. 2010. Effect sizes and the interpretation of research results in international business. *Journal of International Business Studies,* 41(9): 1581-1588.

Essman, S. M., Schepker, D. J., Nyberg, A. J. & Ray, C. 2021. Signaling a successor? A theoretical and empirical analysis of the executive compensation-chief executive officer succession relationship. *Strategic Management Journal,* 42(1): 185-201.

Ferguson, C. J. 2016. An effect size primer: a guide for clinicians and researchers.

Fern, E. F. & Monroe, K. B. 1996. Effect-size estimates: Issues and problems in interpretation. *Journal of Consumer Research,* 23(2): 89-105.

Fisher, R. A. 1946. Statistical methods for research workers. *Statistical Methods for Research Workers,* (10th. ed.).

Gignac, G. E. & Szodorai, E. T. 2016. Effect size guidelines for individual differences researchers. *Personality and Individual Differences,* 102: 74-78.

Gupta, A. & Misangyi, V.F. 2018. Follow the leader (or not): The influence of peer CEOs' characateristics on interorganizational imitation. *Strategic Managemnt Jouranl,* 39: 1437-1472.

Hemphill, J. F. 2003. Interpreting the magnitudes of correlation coefficients. *American Psychologist,* 58(1): 78-79.

Hu, Q., Zhang, Y. & Yao, J. 2018. Family involvement in middle management and its impact on the labor productivity of family firms. *Management and Organization Review,* 14(2): 249-274.

Kaiser, U., Kongsted, H. C., Laursen, K. & Ejsing, A. K. 2018. Experience matters: The role of academic scientist mobility for industrial innovation. *Strategic Management Journal,* 39(7): 1935-1958.

Kelley, K. & Preacher, K. J. 2012. On Effect Size. *Psychological Methods,* 17(2): 137.

Lewin, A. Y., Chiu, C.-Y., Fey, C. F., Levine, S. S., Mcdermott, G., Murmann, J. P. & Tsang, E. 2016. The critique of empirical social science: New policies at management and organization review. *Management and Organization Review,* 12(4): 649-658.

Li-Ying, J., Zhang, Z. & Long, Q. 2018. An alternative way to make knowledge sharing work in online communities? The effects of hidden knowledge facilitators. *Management and Organization Review,* 14(4): 781-825.

Makino, S. & Delios, A. 1996. Local knowledge transfer and performance: Implications for alliance formation in Asia. *Journal of international business studies,* 27(5): 905-927.

Meyer, K. E., Van Witteloostuijn, A. & Beugelsdijk, S. 2017. What's in a p? Reassessing best practices for conducting and reporting hypothesis-testing research. *Journal of International Business Studies,* 48(5): 535-551.

Nakagawa, S. & Cuthill, I. C. 2007. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews,* 82(4): 591-605.

Nieminen, P., Lehtiniemi, H., Vähäkangas, K., Huusko, A. & Rautio, A. 2013. Standardised regression coefficient as an effect size index in summarising findings in epidemiological studies. *Epidemiology, Biostatistics and Public Health,* 10(4).

OBHDP Editors. 2021. Author guidelines for *Organizational Behavior and Human Decision Processes.www.elsevier.com/wps/find/journaldescription.cws_home/622929?generatepdf=true. Accessed December 24, 2021.*

Okada, K. 2013. Is omega squared less biased? A comparison of three major effect size indices in one-way ANOVA. *Behaviormetrika,* 40(2): 129-147.

Olejnik, S. & Algina, J. 2003. Generalized eta and omega squared statistics: measures of effect size for some common research designs. *Psychological Methods,* 8(4): 434.

Pearson, K. 1911. On a correction to be made to the correlation ratio η. *Biometrika,* 8(1/2): 254-256.

Prashantham, S., Zhou, A. J. & Dhanaraj, C. 2020. Depth vs. breadth: Network strategy in emerging markets. *Management and Organization Review,* 16(2): 229-260.

Rea, L. M. & Parker, R. A. 2014. *Designing and conducting survey research: A comprehensive guide*: Jossey-Bass.

Rosenthal, R., Cooper, H. & Hedges, L. 1994. Parametric measures of effect size. *The Handbook of Research Synthesis,* 621(2): 231-244.

Sampson, R. C. 2007. R&D alliances and firm performance: The impact of technological diversity and alliance organization on innovation. *Academy of Management Journal,* 50(2): 364-386.

Song, Y., Liu, Y., Wang, M., Lanaj, K., Johnson, R. E. & Shi, J. 2018. A Social Mindfulness Approach to Understanding Experienced Customer Mistreatment: A Within-person Field Experiment. *Academy of Management Journal,* 61(3): 994-1020.

Sullivan, G. M. & Feinn, R. 2012. Using effect size—or why the P value is not enough. *Journal of Graduate Medical Education,* 4(3): 279-282.

Sun, S., Pan, W. & Wang, L. L. 2010. A comprehensive review of effect size reporting and interpreting practices in academic journals in education and psychology. *Journal of Educational Psychology,* 102(4): 989.

Thompson, D., Bhatt, R., Lazarus, M. M., Cullen, M. K., Baranowski, D. J. & Baranowski, M. T. 1994. Guidelines for authors. *Educational and Psychological Measurement,* 54: 837–847.

Vanneste, B. & Gulati, R. 2021. Generalized trust, external sourcing, and firm performance in economic downturns. *Organization Science,* Forthcomming.

Wetzels, M., Odekerken-Schröder, G. & Van Oppen, C. 2009. Using PLS Path Modeling for Assessing Hierarchical Construct Models: Guidelines and Empirical Illustration. *Management Information Systems Quarterly*: 177-195.

Wilkinson, L. 1999. Statistical Methods in Psychology Journals: Guidelines and Explanations. *American Psychologist,* 54(8): 594.

Wilson, D. B. 2010. Meta-analysis. *Handbook of quantitative criminology:* 181-208. Springer.

Yigit, S. & Mendes, M. 2018. Which effect size measure is appropriate for one-way and two-way ANOVA models? A monte carlo simulation study. ***Revstat Statistical Jorunal***, 16(3): 295-313.

Table 1. Examples of Commonly Used Effect Size Measures

Index	Calculations	Recommended Cut-offs			Notes and References
		Small	Medium	Large	
Measures of effect size used for differences between groups					
Cohen's d	$d = \dfrac{M_1 - M_2}{s}$ $M_1 - M_2$ = difference between group means M s = Standard Deviation of either group				This is also referred as the Standardized Mean Difference. Statistics are for absolute values. It can be used when the independent variable is categorical, and the dependent variable is continuous. From Cohen (1988,1992)
	$d = t_{unpaired}\sqrt{\dfrac{n_1 + n_2}{n_1 n_2}}$ t values from unpaired t-test	0.2	0.5	0.8	This can be used for comparing two independent groups. From Rosenthal (1994), as summarized by Nakagawa and Cuthill (2007)
	$d = t_{paired}\sqrt{\dfrac{2(1 - r_{12})}{n}}$ t values from paired t-test, r_{12} is the correlation coefficient between two groups. $n = n_1 = n_2$				This can be used for comparing two dependent groups. From Dunlap et al. (1996), as summarized by Nakagawa and Cuthill (2007)
Measures of effect size used for correlations					
Pearson's r		0.1	0.30	0.50	This is a commonly used measure for the degree of linear relationship, typically used when the independent and dependent variables are continuous. The cut-offs are suggested by Cohen (1992).
	$r = \dfrac{\sum Z_x Z_y}{N}$	<0.20	0.20	0.30	The cut-offs are suggested by Hemphill (2003).
	Absolute values of correlations are used in the equation.	0.1	0.20	0.30	The cut-offs are suggested by Gignac and Szodorai (2016). We recommend using these cut-offs when using Pearson's r.
Fisher's Z_r	$q = Z_{rA} - Z_{rB}; \quad Z_r = 0.5 log_e(\frac{1+r}{1-r})$ The difference between two independent rs (group A and B), in the Fisher's Z transformation	0.10	0.30	0.50	An adjusted measure based on Pearson's r. Z_r follows a standard normal distribution. The calculation of this measure is summarized by Nakagawa and Cuthill (2007). The cut-offs are suggested by Cohen (1992),

Spearman's ρ	$\rho = 1 - \dfrac{6\sum D^2}{N^3 - N}$ D is the difference between two ranks of observations.	0.20	0.40	0.60	This is a special case of *r* applied to rank data, which can be to assess monotonic relationships (either linear or not). The cut-offs are suggested by Rea and Parker (2014).

Measures of effect size used for ANOVA

η^2	$\eta^2 = \dfrac{SS_B}{SS_B + SS_w}$ Eta squared. SS_B is the between-group sum of squares. SS_w is the within-group sum of squares.	0.01	0.06	0.14	Eta squared is the percentage of variance explained by the membership of being with one group as to the full sample. It is a traditional measure for effect sizes dates back to Pearson (1911) and Fisher (1946). The cut-offs are suggested by Cohen (1988). It is commonly reported by mainstream statistical software in a one-way or factorial ANOVA analysis. However, η^2 is considered as being more biased than ε^2 and ω^2 Thus, we recommend ε^2 as a measure of effect size for ANOVA followed closely by ω^2 in preference to η^2 or η_p^2 as ε^2 and ω^2 are less biased (Yigit & Mendes, 2018).
η_p^2	$\eta_p^2 = \dfrac{SS_B}{SS_B + SS_{error}}$ Partial eta squared. SS_{error} denotes effect and error sums of squares.	0.01	0.06	0.14	Partial eta squared is a variation of eta squared, which comes out the same value if used in a one-way ANOVA. The values could be larger than 1. It is usually used for one-way or factorial ANOVA analysis, but obtainable in multiple-way ANOVA analysis. We recommend ε^2 as a measure of effect size for ANOVA followed closely by ω^2 in preference to η^2 or η_p^2 as ε^2 and ω^2 are less biased (Yigit & Mendes, 2018).
ε^2	$\epsilon^2 = \dfrac{SS_B - (k-1)MS_W}{SS_T}$	0.04	0.25	0.64	This is a frequently used measure that evaluates the strength of a relationship in one-way, two-way, or factorial ANOVA design. This measure is also referred to as adjusted η^2, but less biased than η^2, especially, with small samples. The cut-offs are suggested by Ferguson (2009). We recommend ε^2 as a measure of effect size for ANOVA followed very closely by ω^2 in preference to η^2 or η_p^2 as ε^2 and ω^2 are less biased (Yigit & Mendes, 2018).
ω^2	$\omega^2 = \dfrac{SS_B - (k-1)MS_W}{SS_T + MS_W}$	0.04	0.25	0.64	This is another commonly used measure in a one-way, two-way, or factorial ANOVA analysis. The values of ω^2 is smaller than ε^2. This measure is less biased than η^2, especially, with small samples. The cut-offs are suggested by Ferguson (2009). We recommend ε^2 for a measure of effect size for ANOVA followed closely by ω^2 in preference to η^2 or η_p^2 as ε^2 and ω^2 are less biased (Yigit & Mendes, 2018).

Measures of effect size used for different types of regression models

β	β	<0.2	0.2	0.5	This is the standardized regression coefficients in absolute values, which can be used for most of the fitting models. See Acock (2014:272).
Weighted β	$M = \dfrac{\sum_{i=1}^{k} w_i \beta_i}{\sum_{i=1}^{k} w_i}$ k=number studies, w_i is the standard regression coefficient from study *i*.	--	--	--	These are weighted standardised regression coefficients (in absolute values), which can be used in Meta Analysis when there is a need to report combined results from multiple studies. From Nieminen et.al. (2013).

R^2	R^2	0.04	0.25	0.64	This is the percentage of variance in the dependent variable that is explained by that in the independent variable(s), also referred to as goodness of fit. From Sullivan and Feinn (2012).
Cohen's f^2	$f^2 = \dfrac{R^2}{(1-R^2)}$	0.02	0.15	0.35	We recommend this for a global measure of effect size in multiple regressions. This measure is the multiple partial correlations, commonly used for multiple-covariate regressions. From Cohen (1992).
Cohen's incremental f^2	$f^2 = \dfrac{R^2_{included}-R^2_{excluded}}{(1-R^2_{included})}$ Included means with all variables in the model including the focal variable. Excluded is without the focal variable.	0.02	0.15	0.35	We recommend this measure for effect size of an independent variable in a regression. From Cohen (1992). This can be used tfor an independent variable or moderator.
Chi-square goodness of fit (w)	$w = \sqrt{\sum_{i=1}^{k} \dfrac{(P_{1i} - P_{0i})^2}{P_{0i}}}$	0.1	0.3	0.5	This is a measure based on goodness of fit. P_{1i}m and P_{0i} are the hypothetical and alternate hypothetical population proportion. From Cohen (1988).
Odds ratio (OR)	$\dfrac{\text{Group 1 odds of outcome}}{\text{group 2 odds of outcome}}$	1.5	2	3	This measure is a comparison of odds of one outcome compared to another outcome occurring. It can be used for binary outcome variables, in logistic, loglinear and survival analyses. From Sullivan and Feinn (2012).

Measures of effect size used for Structural Equation Models (SEMs) and PLS

|					
Goodness of fit (GOF)	$GOF = \sqrt{\overline{AVE} \times \overline{R^2}}$	0.1	0.25	0.36	This is a goodness-of-fit measure for Structural Equation Models (not PLS). It is from Wetzels, Odekerken-Schröder, and Van Oppen (2009).
Cohen's f^2	$f^2 = \dfrac{R^2}{(1-R^2)}$	0.02	0.15	0.35	This is the effect size of the part of the model associated with a particular dependent variable. The cut-offs are from Chin (1988: 317) for PLS based on Cohen (1992), but are also appropriate for SEM.
Cohen's incremental f^2	$f^2 = \dfrac{R^2_{included}-R^2_{excluded}}{(1-R^2_{included})}$ Included means with all variables in the model including the focal variable. Excluded is without the focal variable.	0.02	0.15	0.35	We recommend this measure for effect size of an independent variable (Boßow-Thies and Albers,2010) or moderator (Wilson, 2010). This can be used for an independent variable or moderator. The cut-offs come from Chin (1998: 317) for PLS but are also applicable to SEM.
Path Coefficient	Path Coefficient	<0.2	0.2	0.5	We recommend Cohen's f^2 as above to measure the effect size of path coefficients. However, the path coefficient itself is also a measure of effect size. Cut-offs are for absolute values and regression models from Acock (2014:272), but applicable to SEM and PLS.
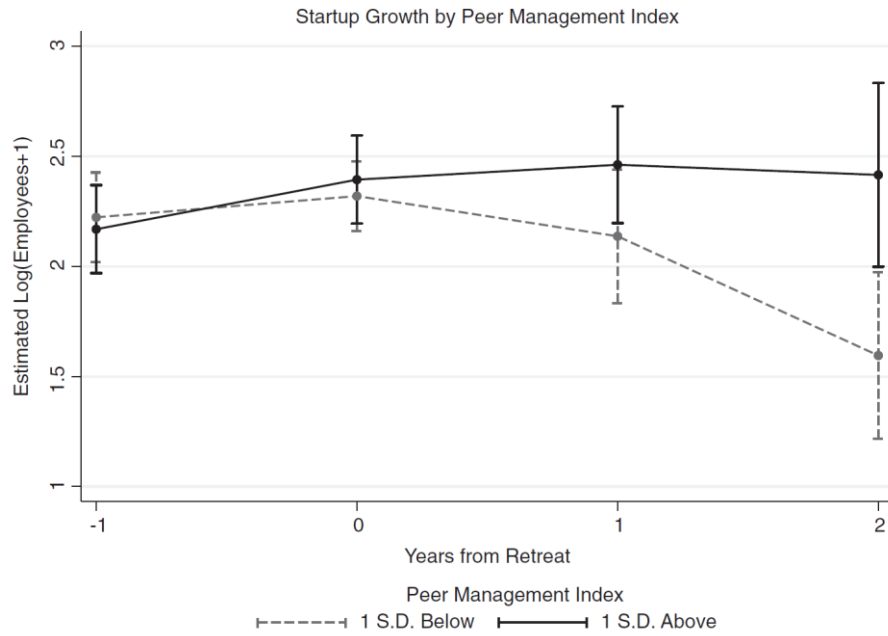
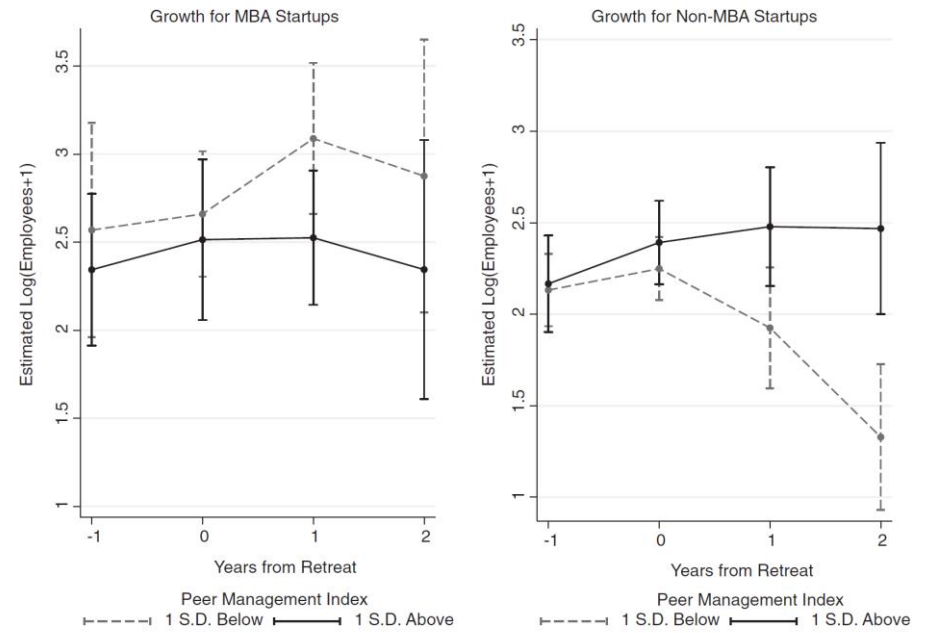FIGURE 1 The impact of management advice on startup growth

FIGURE 2 Estimated startup size by treatment status for startups with MBA founders (23 startups) and for non-MBA founders (67 startups)

Figure1: Examples of plots of effect size as reproduced from Chatterji (2019: Figure 1 and Figure 2)

AUTHOR BIOGRAPHIES

Carl F. Fey is Professor of International Business at Aalto University School of Business. From 2011-2015 he was Dean of Nottingham University Business School China. Professor Fey's research focuses on how cultural and institutional differences affect which management practices work best in different countries especially China and Russia. More specifically, his research explores drivers of innovation, international knowledge transfer, informal networks, organizational culture, leadership, etc. Fey received the 2013 Journal of International Business Studies Decade Award.

Tianyou Hu is assistant professor in the School of Business, Maynooth University. He earned his Ph.D. in Management from National University of Singapore (NUS Business School), and two Bachelor's Degrees in Economics and Literature and an MBA from Peking University. His research interests span the areas of strategic management and international business. Recently, Tianyou is studying new industries, interfirm competition and cooperation, and institutional environment for foreign direct investment.

Andrew Delios is Professor and Vice-Dean MSc Programs at NUS Business School. He was Head of the Department of Strategy and Policy at NUS for 8 years. His research looks at strategy, governance and global competition in emerging economies, with a particular focus on international business issues for companies operating in East and Southeast Asia. Andrew has worked and lived in Asia for 25 years, including three years in Hong Kong and two in Japan.