

An Ounce of Prevention Is Worth a Pound of Cure: Improving Research Quality Before Data Collection

Herman Aguinis¹ and Robert J. Vandenberg²

¹Department of Management and Entrepreneurship, Kelley School of Business, Indiana University, Bloomington, Indiana 47405; email: haguinis@indiana.edu

²Department of Management, Terry College of Business, The University of Georgia, Athens, Georgia 30602; email: rvandenb@uga.edu

Annu. Rev. Organ. Psychol. Organ. Behav. 2014. 1:569–95

First published online as a Review in Advance on November 8, 2013

The *Annual Review of Organizational Psychology and Organizational Behavior* is online at orgpsych.annualreviews.org

This article's doi:
10.1146/annurev-orgpsych-031413-091231

Copyright © 2014 by Annual Reviews.
All rights reserved



Watch a video lecture online

Keywords

theory development, theory testing, research design, measurement, methodology, research quality, validity, causality

Abstract

We rely on classic as well as recently published sources to offer a review of theory, research design, and measurement issues that should be considered prior to conducting any empirical study. First, we examine theory-related issues that should be addressed before research design and measurement considerations. Specifically, we discuss how to make meaningful theoretical progress including the use of inductive and deductive approaches, address an important issue, and conduct research with a practical end in mind. Second, we offer recommendations regarding research design, including how to address the low statistical power challenge, design studies that strengthen inferences about causal relationships, and use control variables appropriately. Finally, we address measurement issues. Specifically, we discuss how to improve the link between underlying constructs and their observable indicators. Our review offers a checklist for use by researchers to improve research quality prior to data collection and by journal editors and reviewers to evaluate the quality of submitted manuscripts.

INTRODUCTION

Organizational science researchers face important challenges. First, there is a need to produce new knowledge and disseminate it in the form of high-quality journal articles, which is more difficult than ever (Ashkanasy 2010). Indeed, prospective authors compete globally for precious journal space, and journal rejection rates hover around 90% (Certo et al. 2010). A second and closely related challenge is the need to conduct rigorous research that produces knowledge relevant to individuals, organizations, and society at large (Cascio & Aguinis 2008, Rynes et al. 2002). For example, one of the professional impact strategic statements of the Academy of Management is to “encourage our members to make a positive difference in the world by supporting scholarship that matters” (<http://strategicplan.aomonline.org/plan>). Similarly, the Society for Industrial and Organizational Psychology’s (SIOP’s) mission is to “enhance human well-being and performance in organizational and work settings” (<http://www.siop.org/mission.aspx>). To address these two challenges, researchers need more than mere depth of knowledge with respect to a particular subject. Selecting important research questions, adopting adequate research designs, choosing appropriate measures, and undertaking rigorous analyses pertinent to the focal questions are equally important. It is through the proper implementation of these various steps that studies produce strong results that are also relevant. We highlight proper because it is this aspect of the research process that proves particularly challenging to many academics. The purpose of our review is to offer suggestions on how proper implementation may be achieved.

The suggestions offered here are motivated by our combined experience as past journal editors, associate editors, and editorial board members. We noted, sadly, that the majority of manuscripts not accepted for publication after traversing the peer-review process could have actually been rejected before the data were collected. The problems with the majority of rejected manuscripts are related to theory, research design, and/or measurement. Rarely are data analyses grounds for rejection because weak analyses can often be fixed if all of the other components are strong. However, there seems to be a belief among some researchers that using the latest and greatest statistical tools will overcome deficiencies regarding theory, design, and measurement. Perhaps partly to blame for this belief is the overemphasis on analytical issues in our methods journals. This overemphasis may give researchers the impression that data analysis is most important. For instance, a review of almost 200 articles published in *Organizational Research Methods* from 1998 to 2007 revealed that data analysis was addressed by about half of all articles. By contrast, only 15% addressed research design topics, and about 35% addressed measurement issues (Aguinis et al. 2009). We conducted a similar content analysis of articles published in *Psychological Methods* by randomly selecting four years of issues published between 2000 and 2012 (116 total articles). Our results revealed that only 10% of the articles addressed pure design issues and another 19% addressed statistical and measurement issues with clear design implications. Pure statistics, though, were the focus of 71% of the articles. Yet another possible reason for the overemphasis on data analysis is that the methodological training of future scholars emphasizes data analysis over design (Tett et al. 2013). In fact, very few doctoral-level courses address design by itself—a situation that has not changed much over the past three decades (Aiken et al. 1990, 2008). The precise reasons for the lack of relative attention to design and measurement issues are not known. This is likely a result of a combination of factors including doctoral-level training and implicit norms regarding the value-added contribution of manuscripts addressing design and/or measurement issues as compared with those emphasizing data-analytic issues. Also, faculty reward systems emphasizing the publication of a minimum number of articles in journals considered prestigious, and an emphasis on quantity of publications and outlets where this research appears rather than quality and actual content, are another likely contributing factor. From

the perspective of a junior researcher whose goal is to receive a positive tenure decision, publishing in such top-tier journals, and as often as possible, becomes a priority, and this often means cutting corners and not paying sufficient attention to design and measurement issues—which often leads to the opposite effect (i.e., rejection decision) of what is desired.

Benjamin Franklin is credited for having admonished that an ounce of prevention is worth a pound of cure. Our review applies this sage advice to our field and offers a checklist to be used by researchers to improve the quality of their empirical work prior to data collection and by journal reviewers and editors to evaluate the quality of submitted manuscripts. We address issues about theory, research design, and measurement. Each of our recommendations is accompanied by specific and actionable advice that researchers can implement as they strategize and plan empirical studies and that journal editors and reviewers can adopt when they evaluate manuscripts.

Organization and Overview

First, we tackle theory-related issues including how to make meaningful theoretical progress, how to insure a study addresses an important issue, and how to conduct research with a practical end in mind. Second, we discuss issues related to research design, including how to address the low statistical power challenge, how to design studies that strengthen inferences about causal relationships, and how to use control variables appropriately. Finally, we address measurement issues. Specifically, we discuss how to improve the link between underlying constructs and their observable indicators. As a preview, **Table 1** summarizes the issues and recommendations we discuss in detail throughout the article.

THEORY

Making Meaningful Theoretical Progress

Our primary role as social scientists is to make theoretical progress, and indeed, it is the metric used to evaluate the impact of our research within our respective disciplines. It is our belief that current practices are not achieving as much theoretical progress as many believe. As concluded by Edwards (2010, p. 616), the general problem is that

given the value placed on theory and the pressures and rewards that emphasize theory development, it would seem safe to assume that organizational and management research has made great theoretical progress. This assumption is arguably tenable if we equate theoretical progress with the development of new theories, an outcome that would naturally result from the norms and incentives in our field. However, if we broaden the meaning of theoretical progress to include the refinement of theories, such that we expose theories to stringent tests and modify or eliminate contenders that fail such tests, our record of theoretical progress would appear less positive.

To put the issue we discuss in this section in context, we start by briefly delineating between inductive and deductive research approaches because both have important roles in contributing to meaningful theoretical progress. Moreover, as of the writing of this article, there is some movement within the profession to encourage the publication of inductively oriented research. However, because this movement has just started building momentum, the bulk of this section focuses on the deductive approach.

In brief, the inductive approach depends on the data to provide meaningful patterns to the theorist, where meaningfulness is some tension, conflict, or contradiction in those patterns relative

Table 1 Summary of recommendations for improving research quality before data collection

Issue	Reason for importance of issue	Consideration(s) mainly for journal editors and reviewers	Recommendations mainly for authors
1. Making meaningful theoretical progress	The current state of affairs includes theory proliferation and numerous theories that have yet to be meaningfully tested. Until they are, they do not make a contribution to advancing our understanding of organizational phenomena.	Does the study put the focal theory at risk? Does the study take advantage of inductive and deductive theorizing?	Engage in inductive research designs to uncover conflicts, tensions, and contradictions to current theory. Expand the null hypothesis test by stating a threshold above a null effect or relationship. Use effect sizes reported in previously published results as a minimum threshold for claiming meaningful theoretical progress. Precisely specify the functional form of the relationship as something other than linear, and test for its presence. Increase the precision of all measures in the study so as to increase confidence in their validities. When possible, conduct research that follows experimental and quasi-experimental principles. Design studies with the pursuit of failure as the goal—not something to be avoided.
2. Addressing an important question	Addressing an important question improves the chances that results will make a valuable contribution to a body of knowledge and change the conversation—regardless of the nature of the actual results.	What is the evidence that the particular question represents a puzzle, conundrum, or point of confusion? What is the evidence that the chosen research question represents an important challenge in the field? Does the research question aim at creating and/or shifting consensus?	For junior scholars, become involved in as many research projects as possible (beginning in graduate school) and volunteer as a reviewer for conferences. For more senior scholars, volunteer as a reviewer for journals and, eventually, as an editorial board member. Attend research seminars and professional conferences. Consult with colleagues regarding their views on the most important and challenging questions in the field. Conduct reviews of persistent questions and challenges posted on electronic mailing lists. The review process can be facilitated by the use of computer-aided text analysis.

(Continued)

Table 1 (Continued)

Issue	Reason for importance of issue	Consideration(s) mainly for journal editors and reviewers	Recommendations mainly for authors
3. Conducting research with a practical end in mind	<p>Conducting research with a practical end in mind will help accomplish the explicit strategic objectives of organizational science regarding the improvement of human welfare and elevate the status and prestige of the organizational sciences in the eyes of stakeholders outside of the academy, including public policy makers, the media, and society at large.</p>	<p>Does the research adopt a design-science approach that addresses not only what is but also what can be (i.e., preferred futures)?</p> <p>Is the research solution oriented?</p>	<p>Train future scholars regarding the need to adopt a design-science approach.</p> <p>Study dependent variables that are of interest to decision makers and independent variables that can be changed by instituting new policies.</p> <p>Engage in boundary-spanning activities, such as spending sabbaticals in business practice either as “translators” of research results or as researchers on a set of practitioner-oriented research issues.</p> <p>Become involved in senior managerial decision making by serving on boards of directors.</p> <p>Participate in executive-education contexts as a means to develop relationships with practitioners.</p> <p>Write articles targeting a practitioner audience because this will lead to a better awareness of the important issues for stakeholders outside of the academy.</p>
4. Addressing the low statistical power challenge	<p>Conducting research at insufficient levels of statistical power yields inconclusive results when the null hypothesis is not rejected. Specifically, we do not know whether the statistically nonsignificant result was due to a true null hypothesis (i.e., the effect does not exist in the population) or to a research design that was unable to detect the existing effect.</p>	<p>If a null hypothesis is not rejected (i.e., a statistically nonsignificant result), request that authors report results of a statistical power analysis.</p> <p>If a null hypothesis is not rejected and statistical power is not sufficient (e.g., $<.80$), request that authors mention that there is insufficient evidence to support the absence of the effect.</p>	<p>Use power calculators to plan research design and understand the trade-offs involved in terms of resources and practical constraints.</p> <p>Set the a priori α level using a rational process as opposed to using the conventional and arbitrary .05 and .01 values.</p> <p>Adopt a Bayesian instead of a more traditional null hypothesis significance testing approach.</p>

(Continued)

Table 1 (Continued)

Issue	Reason for importance of issue	Consideration(s) mainly for journal editors and reviewers	Recommendations mainly for authors
5. Strengthening inferences about causal relationships	Causal inferences are a desired end state for most researchers because they strengthen the focal theory and make that theory more useful.	How closely does the study adhere to the conditions of causality? Did the authors use design elements that reduce threats to the validity of making causal inferences from the findings? Did the authors attempt to use a quasi-experimental approach?	Thoroughly understand the conditions for causality and use these as decision heuristics when designing the study. Incorporate design features that turn what otherwise might be a nonexperimental study into a quasi-experiment. Utilize a longitudinal design in which a change in X is lagged ahead of a change in Y, and test whether the two change vectors are related to mimic the cause coming before the effect. Utilize latent class tools such as latent profile analysis on the sample to identify groups who may be at different levels of the independent variables, and use this as a manipulation to evaluate whether Y differs in the expected direction between the groups.
6. Using control variables appropriately	Because most research designs are observational in nature, control variables are usually needed to rule out alternative explanations for observed substantive relationships. However, control variables are often used indiscriminately and may, ironically, accentuate the problems they are intended to remedy.	Did the authors offer a clear conceptual reason for the inclusion of particular control variables? Did the authors provide evidence regarding the adequacy of the psychometric properties for the measures used to assess control variables?	Include control variables that rule out key alternative explanations. Provide a sound conceptual basis as to how the variables in those alternatives will operate in the model. Examine correlations between controls and dependent variables, and if they are not significant, then do not include the controls in the models. Include control variables in tables summarizing descriptive statistics. Compare two models, one with and one without the controls, to assess impact of controls on parameter estimates. If the model fits just as well without controls and parameter estimates are similar, then controls may not have a meaningful effect.

(Continued)

Table 1 (Continued)

Issue	Reason for importance of issue	Consideration(s) mainly for journal editors and reviewers	Recommendations mainly for authors
<p>7. Improving the link between underlying constructs and their observable indicators</p>	<p>A small observed relationship can be explained by a weak relationship between constructs or, alternatively, by a weak relationship between predictor indicators and their underlying constructs, outcome indicators and their underlying constructs, or both. Thus, there is a need to gather construct validity evidence; otherwise, substantive results are inconclusive.</p>	<p>What is the evidence regarding the link between constructs and their indicators?</p>	<p>Follow established best-practice recommendations regarding scale construction. Provide a clear and unambiguous definition of each construct. Do not alter items (e.g., change wording or omit items) without gathering additional construct validity evidence because the revised scale may no longer assess the originally intended construct. Gather construct validity evidence using computer-aided text analysis (for text-based measures) and video-based techniques (for constructs that are not self-referential).</p>

to what we know or understand (Shepherd & Sutcliffe 2011). It can be described as letting the data speak, such that some forms of analyses result in patterns or concepts, and the researcher's role is to make the connections within those patterns or among those concepts (Glaser 1992). This emerging theory is compared with existing theories, which may have been generated using a deductive or inductive approach, to evaluate its potential contribution. What knowledge the theorist brings to the task could range from a state of unknowing (having little, if any, preconceived notions about what will emerge) to a state of knowing (being a subject-matter expert in an area and letting the knowledge guide what happens to the emerging patterns next). As noted by Locke (2007), inductive theorizing can lead to important theoretical progress. Evidence of such was provided by Aguinis et al. (2013a), who noted that, compared with theory-testing elements, theory-building elements in an empirical article lead to a greater number of citations (Colquitt & Zapata-Phelan 2007). Despite inductive research's ability to facilitate theoretical progress, the fact remains that few of our mainstream journals regularly publish or directly recognize this type of research in their general calls for papers (Eby et al. 2009). To address this gap, the Academy of Management has announced the creation of the *Academy of Management Discoveries*, a new journal, for now in electronic format only, devoted to inductive research. In addition, there was a recent call for papers on inductive research for a forthcoming special issue of the *Journal of Business and Psychology* (Ryan et al. 2014).

The deductive approach predominates, and it seems it will continue to do so into the foreseeable future. As of this article's writing, few of the major journals have policies that include inductive contributions explicitly, and thus, most describe a meaningful theoretical contribution as one that follows a deductive, top-down approach. Specifically, the theorist (someone typically well ensconced in a specific research literature) discovers some problem/issue/gap within the existing knowledge and sets out to solve the problem (or resolve the issue or close the gap) by providing a potentially superior explanation (Shepherd & Sutcliffe 2011). The explanation is just potentially superior because it is only through its specification of testable hypotheses that it "provides the *potential* to generate new theoretical insights" (p. 361, emphasis in original)—that is, insights that include the old and the new, with the latter representing the changes needed to solve the problem, resolve the issue, or close the gap.

Although the above is a nice ideal toward which our journals strive, forcing every submission to make a meaningful theoretical contribution following a hypothetico-deductive approach has created its own problems. These problems have been thoughtfully documented over many decades, and we do not repeat them here (see, e.g., Davis & Marquis 2005; Edwards 2010; Hambrick 2007; Kacmar & Whitfield 2000; Lakatos 1978; Meehl 1978; Pfeffer 1993, 2007; Starbuck 2004; Weick 1989). As emphasized in a feature topic issue of *Organizational Research Methods* (2010, issue 4), although the problems have been long recognized, little, if anything, has been done to resolve them (Davis 2010, Edwards & Berry 2010, Gray & Cooper 2010, Leavitt et al. 2010).

Leavitt et al. (2010) metaphorically described it as a situation in which we have a very large garden with so much growing that the desired plants (strong generalizable theories) cannot be discerned from the weeds (weak incremental or grandiose theories). As recognized by Gray & Cooper (2010), our inability to prune theories down by subjecting them rigorously to failure has had at least three undesirable outcomes. First, because the organizational sciences provide little incentive to question the value of published theories, our field suffers from an overaccumulation of undigested findings about theories of unknown fidelity. Second, there is very little, if any, incentive to discover the true limits of one's theory beyond just an initial empirical study or two. "Third, theory development seems to have been replaced by eclectic (and sometimes elegant) problem solving that is theoretically agnostic (Davis & Marquis 2005, p. 334; Pfeffer & Fong 2005, p. 372), using whatever constructs are at hand, an approach that, in the pointed phrase of Davis

and Marquis (2005, p. 335) ‘risks being simply business journalism with regressions’” (Gray & Cooper 2010, pp. 622–23).

The reader may be asking at this point, what has any of the above to do with the general theme of improving research quality before data collection? Our response is that it has everything to do with our theme. To understand, we look to Popper (1963) regarding the manner in which a theory contributes to scientific knowledge. In his own words, Popper stated the following about a meaningful theoretical contribution:

1. It is easy to obtain confirmations, or verifications, for nearly every theory—if we look for confirmations.
 2. Confirmations should count only if they are the result of *risky predictions*; that is to say, if, unenlightened by the theory in question, we should have expected an event which was incompatible with the theory—an event which would have refuted the theory.
 3. Every “good” scientific theory is a prohibition: it forbids certain things to happen. The more a theory forbids, the better it is.
 4. A theory which is not refutable by any conceivable event is non-scientific. Irrefutability is not a virtue of a theory (as people often think) but a vice.
 5. Every genuine *test* of a theory is an attempt to falsify it, or to refute it. Testability is falsifiability; but there are degrees of testability: some theories are more testable, more exposed to refutation, than others; they take, as it were, greater risks.
 6. Confirming evidence should not count *except when it is the result of a genuine test of the theory*; and this means that it can be presented as a serious but unsuccessful attempt to falsify the theory. (I now speak in such cases of “corroborating evidence.”)
 7. Some genuinely testable theories, when found to be false, are still upheld by their admirers—for example by introducing *ad hoc* some auxiliary assumption, or by reinterpreting theory *ad hoc* in such a way that it escapes refutation. Such a procedure is always possible, but it rescues the theory from refutation only at the price of destroying, or at least lowering, its scientific status. (I later described such a rescuing operation as a “*conventionalist twist*” or a “*conventionalist stratagem*.”)
- One can sum up all this by saying that the criterion of *the scientific status of a theory is its falsifiability, or refutability, or testability*. (pp. 36–37, emphases in original)

Even with the use of these seven characteristics as benchmarks against which to gauge current practices, it is understandable why the continued proliferation of theories may not be viewed as all that positive in terms of making theoretical progress (Edwards 2010) and why the practice has created a garden full of plants but little is being done to separate the weeds from the truly palatable (Leavitt et al. 2010). Our goal is not to imply that a moratorium be placed on theoretical development per se. However, what constitutes meaningful theoretical progress needs to be redefined. It is through this redefinition that improvements in research quality can occur before data are collected.

One means to improve a study’s meaningful theoretical contribution is through theoretical precision. Specifically, Edwards & Berry (2010) evaluated a set of the most cited theories between 1995 and 2009 against a template of characteristics defining theoretical precision. They found that, for the most part, the theories they reviewed developed propositions that predicted the direction of a relationship but said very little about the form of that relationship or the conditions that might influence the relationship. “Rather, the majority of the propositions essentially stated

that, if one variable increases, another variable will increase or decrease” (p. 670). Edwards and Berry presented six means through which researchers could increase the precision of theoretical propositions. For example, when specifying a theory, researchers should expand the null hypothesis to predict not merely that a parameter differs significantly from zero but that the parameter deviates from zero by some minimum threshold. Another means they suggested for strengthening theoretical precision is to specify the functional form of the relationships between the focal variables. Most theories take a monotonic view, when in reality, the relationships possess some other form such as curvilinear (Pierce & Aguinis 2013).

Although Davis (2010) focused primarily on macro-organizational theory, two of his recommendations on how to make meaningful theoretical progress are also particularly appropriate here. We only briefly touch on these here because we address them more fully below. First, he observed that when developing theories, measurement is typically not taken very seriously, and indeed is often treated sloppily. Given that measurement is the lens through which we operationalize focal constructs, measurement precision should be paramount even in the specification of the theory. This holds true for any control variables as well. Second, Davis recommended that researchers refamiliarize themselves with the validity standards underlying quasi-experimentation. His main point is that in some situations, certain data points could be used to represent one condition, and others to represent other conditions, and the resulting comparison could rule out alternative explanations or prevent the researcher from supporting the obvious. Thus, quasi-experimentation could be a useful conduit for making meaningful theoretical progress.

Gray & Cooper (2010) suggested that one means to make a meaningful theoretical contribution is through the pursuit of failure, as advocated by Popper (1963). One example of pursuing failure is to identify and specify the tacit assumptions a priori (either before initial data collection or after multiple data collections) to avoid the “conventionalist twist” noted above. This will ensure at some level that operationalizations are included or designs are followed that incorporate the ancillary hypotheses emerging from the tacit assumptions. Gray and Cooper noted that pursuing failure may also be accomplished by explaining counterexamples. This means that

the ruthless pursuit of failure should also entail understanding exceptions—cases that do not fit the theory. It is here that the field’s reluctance to pursue failure is most evident. With few exceptions, studies test a theory against the null hypothesis, something that only the feeblest theory can fail to beat (Meehl, 1978, p. 821). When a theory rejects the null hypothesis at some conventional level of significance, it is as if we conclude that that is good enough; we have virtually no tradition of asking about those data points that do not fit, perhaps thinking that it is too much to expect that all cases can be explained. (pp. 630–31)

Finally, Leavitt et al. (2010) also presented several suggestions for improving a study’s contribution to theory. For example, they provided several recommendations as to what comprises a meaningful comparison between theories and as to how theoretical development of those comparisons may be presented in a manuscript’s introduction section. They subsequently extend this presentation into the design characteristics and features of the study itself so as to most appropriately undertake the comparisons.

The major point of this opening section in our article is to encourage studies that really have an impact, instead of supporting the proliferation of empirical pieces that frankly add little to our primary task of making meaningful theoretical progress. This will require rethinking how we design our studies and what form the data will take. Opening avenues for inductive research is an appropriate step in the right direction. Such research is likely to be facilitated given the vast stores of data being collected by firms in response to the analytics (i.e., Big Data) movement as well as the

sharp decrease in the cost of data storage technology (Aguinis et al. 2013a). Although the advice in the following sections addresses more specific issues, it is given with the overall goal of facilitating meaningful theoretical progress.

Addressing an Important Question

The research process and quest for a meaningful theoretical contribution begin with a question—and it should be an important one. Emphasizing how crucial this step is, the Academy of Management chose as its theme for its annual meeting in 2008 “The Questions We Ask.” That year’s call for papers read as follows: “What puzzles, conundrums, points of confusion, and unanswered questions really bedevil you and your close colleagues? Be sure to consider the most meaningful questions. Just because a question has yet to be asked or answered does not mean that we need to address it. Some questions are more important than others” (Acad. Manag. 2008). In his Academy of Management presidential address that year, Walsh (2011, p. 224) noted, “All I can do right now is to ask us to ask the most important questions. It is up to each of us to define what those important questions are. Let’s just not recoil from the challenge and, in the extreme, seek the easy refuge of the minimum publishable unit.”

The work of Richard Hamming (1915–1998), one of the founding fathers of the modern field of computer science, is an illustration of research addressing an important question. Hamming was a member of the Los Alamos team that created the first atomic bomb. In his role as a team member, he asked only one, but a very important, question: Would the detonation of an atomic bomb ignite the entire atmosphere? In other words, the goal of his research was to gather knowledge on whether the detonation of the first atomic bomb would lead to the destruction of all life on our planet. In a 1986 address at the Bell Communications Research Colloquium Seminar, Hamming (1986, p. 8) admonished that “if you do not work on an important problem, it’s unlikely that you will do important work.”

Similarly, Hollenbeck (2008) noted that important questions aim at either consensus shifting or consensus creation. In other words, studying an important question produces results that are contributions by “either increasing consensus about the validity and utility of some idea or changing the consensus away from one idea toward some other idea that everyone agrees is better” (p. 17). Indications that the question being addressed is likely not important include when the justification for a project involves a statement that “this has never been done before” or when results simply confirm obvious and long-held assumptions (Davis 1971).

Addressing important questions is not easy, and this is probably why it is not done as frequently as it should be (Bartunek et al. 2006). Specifically, identifying important questions requires familiarity with large bodies of work—a daunting obstacle for a single individual. However, interacting and seeking advice from knowledgeable colleagues in various research domains can facilitate the task. For example, Pierce & Aguinis (2013) identified a meta-theoretical principle they labeled the “too-much-of-a-good-thing” (TMGT) effect in the particular domains of organizational behavior and human resource management. The TMGT effect states that all seemingly monotonic positive relations reach context-specific inflection points after which the relations turn asymptotic and often negative, resulting in an overall pattern of curvilinearity. Pierce and Aguinis solicited and received advice from several scholars in the fields of entrepreneurship and strategy regarding how research findings in these areas may also be explained by the TMGT effect. These interactions resulted in the identification of several research streams explained by the TMGT effect, not only in organizational behavior and human resource management (e.g., leadership, personality, job design, personnel selection) but also in other fields (e.g., new venture planning, firm growth rate, diversification, and organizational slack). Also, archives

for electronic mailing lists provide an additional source of useful information regarding important questions in a particular field. For example, Aguinis et al. (2013b) searched the archives of RMNET (Research Methods Division of the Academy of Management) and MULTILEVEL (a list specifically devoted to multilevel analysis) to compile a list of the most persistent and challenging questions regarding how to estimate and interpret cross-level interaction effects, and they then used these questions as the impetus for their research.

In understanding the extent to which a study addresses an important question, the main task for journal editors and reviewers is to answer the following questions: (a) What is the evidence that the particular question represents a puzzle, conundrum, or point of confusion? (b) What is the evidence that the chosen research question represents an important challenge in the field? and (c) Does the research question aim at creating and/or shifting consensus? For authors, gathering such evidence should be seen as an ongoing process, which begins for most scholars by taking doctoral seminars and, more importantly, becoming involved in as many research projects as possible in graduate school. For those in more senior career stages, the process involves serving as a reviewer for conferences and journals and, eventually, volunteering to serve on as many journal editorial boards as possible. These activities, which involve reading about research and doing research as much as possible, are important because, to paraphrase Hamming (1986), knowledge is like compound interest. More knowledge leads to more learning, and more learning leads to more opportunities for identifying important questions. Additional recommendations are to consult with colleagues regarding their views on the most important and challenging questions in the field. Such interactions take place by attending research seminars and professional conferences. These activities provide opportunities to hear about research being conducted by a large number of scholars and, hence, are also a good source of knowledge regarding important questions and challenges. As noted earlier, conducting reviews of exchanges posted on electronic mailing lists is another excellent source of information, which can be mined systematically using computer-aided text analysis (e.g., Pollack 2012). We believe that after implementing these recommendations, as noted by Bergh (2008, p. 121), when the time comes to write the resulting manuscript, “the contribution of the manuscript will be obvious and result in no guesswork.” Moreover, such research has the greatest potential to “change the conversation” in a particular research domain.

Conducting Research with a Practical End in Mind

The science-practice gap refers to a lack of connection between the knowledge academics produce and the actionable knowledge practitioners need for addressing their most pressing real-time challenges (Rynes et al. 2002). The concern about the presence of a science-practice gap is not new (Dunnette 1990). For example, Cascio & Aguinis (2008) conducted a content analysis of almost 6,000 articles published in the *Journal of Applied Psychology* and *Personnel Psychology* from 1963 to 2007. Their results revealed that the majority of research does not address contemporary human capital issues of concern at that moment in time to practitioners and society in general. In fact, Cascio & Aguinis (2008, p. 1074) concluded that “if we extrapolate past emphases in published research to the next 10 years, we are confronted with one compelling conclusion, namely, that I-O [industrial-organizational] psychology will not be out front in influencing the debate on issues that are (or will be) of broad organizational and societal appeal.”

The science-practice gap is not unique to any one domain in the organizational sciences but characterizes many (Rousseau 2007, Tushman & O’Reilly 2007). In his Academy of Management presidential address, Hambrick (1994) lamented that we as scholars in the organizational sciences have a “minimalist ethos: . . . minimal visibility, minimal impact.” He noted, “Each August, we come

to talk to each other [at the Academy of Management's annual meetings]; during the rest of the year we read each other's papers in our journals and write our own papers so that we may, in turn, have an audience the following August: an incestuous, closed loop" (p. 13).

We emphasize that the concern about the science-practice gap does not imply that there is no need for basic research without immediate practical application. However, if the vast majority of organizational science research falls into that category, it is unlikely that the knowledge generated will help fulfill the explicit mission statements of professional organizations such as the Academy of Management or SIOP. Specifically, one of the Academy of Management's strategic intent statements includes "Professional Impact: The Academy of Management encourages our members to make a positive difference in the world by supporting scholarship that matters" (<http://strategicplan.aomonline.org/plan>). Also, as noted earlier, SIOP's mission "is to enhance human well-being and performance in organizational and work settings" (<http://www.siop.org/mission.aspx>).

In terms of the research process, a study's contribution can be amplified when there is a practical end in mind (Hakel et al. 1982). However, designing research with a practical end is not easy, and perhaps this is the reason why it is done infrequently (Cascio & Aguinis 2008). There are steps, though, that increase the probability that a research study will have clear implications for practice.

One of those steps involves adopting a design-science approach as proposed by Nobel laureate Herbert Simon [1996 (1969)]. Simon highlighted the need to recognize design and a future orientation in the applied sciences (Van Aken & Romme 2012). In other words, applied sciences are concerned not only about what is but also about what can be. Applied disciplines such as medicine and engineering follow this approach systematically, which consists of not just describing the present, but also creating preferred futures (Van Aken & Romme 2012). In the case of medicine, a design-science approach involves, for example, restoring health to a patient suffering from cancer. In the case of engineering, a design-science approach involves, for example, creating a more fuel-efficient car. In the organizational sciences, a design-science approach may involve creating personnel selection procedures that are equally fair for members of all ethnic groups (e.g., Aguinis et al. 2010a) or a performance management system that maximizes not only individual and firm performance but also individual growth and personal development (Aguinis 2013).

Among other steps authors can take are the following specific actions (Cascio & Aguinis 2008). The first is to train future generations of organizational scholars on how to undertake a design-science approach. As noted by former Academy of Management President Bill Starbuck, "People should do management research because they want to contribute to human welfare. Those who are professors of management are people of superior abilities and they should use these abilities for purposes greater than themselves" (transcribed in Barnett 2007, p. 126). Part of this training should involve honing the skill to choose dependent variables that are of interest to decision makers and independent variables that can be changed by instituting new policies (Ruback & Innes 1988). Second, authors at all career stages can engage in boundary-spanning activities such as spending sabbaticals in business practice either as "translators" of research results or as researchers studying a set of practitioner-oriented research issues (Shapiro et al. 2007). Third, researchers can become involved in senior managerial decision making by serving on boards of directors (Anderson 2007). An additional suggestion is to participate in executive education as a means to develop relationships with practitioners (Tushman & O'Reilly 2007). Finally, researchers can write articles targeting a practitioner audience in outlets such as *Business Horizons* and *Harvard Business Review* because this will allow them to create a better awareness of the important issues for stakeholders outside of the Academy.

We do not advocate that researchers simply chase contemporary trends. However, each of the aforementioned actions will allow researchers to identify important practical problems and offer

valuable solutions leading to the design of preferred futures. The resulting research will not only help accomplish the explicit strategic objectives of organizational science professional organizations, but also help elevate the status and prestige of these fields in the eyes of outside stakeholders, including public policy makers, the media, and society at large.

RESEARCH DESIGN

Addressing the Low Statistical Power Challenge

Organizational science research is almost always based on samples and not populations. However, the goal is to produce theories that generalize to populations. Accordingly, we use inferential statistics to make probabilistic statements about the extent to which results found in samples can indeed be generalized to larger populations. Null hypothesis significance testing (NHST) is the current paradigmatic approach used to infer whether effects exist in a population. Essentially, NHST consists of stating a hypothesis of no effect in a population (i.e., the correlation between variables X and Y is zero; $\rho = 0$) and collecting sample-based data to compute the likelihood of finding an effect of a certain magnitude or larger if the null hypothesis is true. If this probability is very small, as defined by a predetermined threshold labeled α , we reject the null hypothesis and conclude that the population effect is unlikely to be zero. For example, assume we conduct a study to investigate the relationship between pay and job satisfaction and obtain a correlation of ($r =$) .15. Then, based on an inferential statistical test, we find that if the relationship between pay and job satisfaction is zero in the population, the probability of obtaining $r = .15$ or larger based on sample data is ($p =$) .03. Assuming a typical $\alpha = .05$, because the obtained probability is smaller than our α threshold, we reject the null hypothesis and conclude that the relationship between pay and job satisfaction is not zero (i.e., $\rho \neq 0$).

NHST has several known problems (Aguinis et al. 2010b, Kruschke et al. 2012, Lance 2011). Most notably, we would like to know the viability of a null hypothesis given the data [$p(H_0|D)$]. However, NHST tells us the probability of obtaining the data in hand, or more extreme unobserved data, if the null hypothesis were true [$p(D|H_0)$]. Unfortunately, $p(H_0|D) \neq p(D|H_0)$. Accordingly, Cohen (1994, p. 997) eloquently noted that a test of statistical significance “does not tell us what we want to know, and we so much want to know what we want to know that, out of desperation, we nevertheless believe that it does!”

In spite of repeated calls to remove NHST from our journals (e.g., Hunter 1997, Orlitzky 2012), it is unlikely that this will happen any time soon. Thus, in this section we offer recommendations addressing statistical power, which is the probability that we will correctly conclude that the null hypothesis should be rejected. We focus on this issue because an important concern is that the vast majority of studies are conducted at insufficient levels of statistical power (Maxwell 2004, Mone et al. 1996). Stated differently, the probability that a researcher will correctly conclude that there is an effect in the population, when this is in fact true, is usually only around 50% (Maxwell 2004). Moreover, power is even lower when hypotheses involve more complex relationships such as interaction (Aguinis et al. 2005) and multilevel (Mathieu et al. 2012) effects. We focus on statistical power also because this is an issue that can be addressed prior to data collection—there are other suggestions about how to address weaknesses of NHST after data have been collected, such as reporting effect sizes, confidence intervals around effect sizes, and the practical significance of such effects (Aguinis et al. 2010b).

The three factors affecting statistical power are the size of the effect in the population (i.e., larger population effect \rightarrow greater power), sample size (i.e., larger $N \rightarrow$ greater power), and the a priori probability level considered sufficiently small to reject the null hypothesis (i.e., larger $\alpha \rightarrow$ greater power) (Cohen 1988). Obviously, researchers cannot control the size of the effect in the

population—this is precisely what we are trying to assess. Researchers do control to some extent, however, both sample size and the a priori α . To illustrate, we continue with the pay and job satisfaction example from above. Assume the nonzero population correlation is known and is ($\rho =$) .20. If we set $\alpha = .05$ and desire an 80% level of statistical power, we need data from at least 153 individuals. However, if we collect data from only 100 individuals, the probability of correctly rejecting the null hypothesis is only 64%. If we collect data from 68 people, statistical power is 50%; and if sample size is 50, power is only 40%.

There are several specific actions that can address the low statistical power challenge. First, one should understand the consequences of various research design decisions. Fortunately, there are statistical power calculators that can be used for various types of hypotheses and research designs, and many of them are available online free of charge. For example, G*Power 3 (available at <http://www.psych.uni-duesseldorf.de/abteilungen/aap/gpower3/>) allows for power calculations involving the F test, t test, χ^2 test, and other test statistics. For substantive hypotheses involving multilevel relationships, there is Optimal Design (Spybrook et al. 2011) or Power IN Two-level designs (PINT; Bosker et al. 2003). More specialized software is also available to calculate power for tests involving moderator effects in the context of multiple regression (Aguinis 2004) and multilevel modeling (Mathieu et al. 2012) (see programs available at <http://mypage.iu.edu/~haguinis>). As illustrated with the pay and job satisfaction example, a statistical power analysis allows for understanding the consequences of increasing and decreasing sample size. Also, conducting a statistical power analysis prior to data collection provides information about how various decisions, which are usually dictated by resources and logistical constraints, will affect the accuracy of the resulting conclusions. For example, what are the relative trade-offs in terms of power if, in the context of a multilevel investigation, we are able to increase the number of level-two units (i.e., collect data from 30 more teams), but at the expense of having, on average, five instead of seven individuals on each team (i.e., level-one units)?

Increasing the α level is also under the control of researchers (e.g., Cascio & Zedeck 1983). A review of articles published from January 2002 to December 2006 in *Administrative Science Quarterly*, *Academy of Management Journal*, and *Strategic Management Journal* revealed that 99% of authors used the conventional α values of .05 and .01 (Aguinis et al. 2010b). Despite their pervasive use, the .05 and .01 α values are themselves arbitrary (Little 2001). Specifically, the .05 value for α originated with Fisher (1925), and his major reason for selecting it was that “he had no table for other significance levels, partly because his professional enemy, Karl Pearson, refused to let him reprint the tables Pearson had” (Gigerenzer 1998, p. 200). This led Rosnow & Rosenthal (1989, p. 1277) to sarcastically state, “Surely, God loves the .06 nearly as much as the .05.”

Murphy & Myers (1998) suggested a useful way to weigh the pros and cons of increasing α for a specific research situation. It is based on the appropriate balance between α and β (i.e., the probability of incorrectly concluding that there is a population effect). Note that α and β have an inverse proportional relationship and statistical power is formally defined as $1 - \beta$. To set an appropriate level for α , we consider what is called the desired relative seriousness (DRS) of making an α versus a β error. Specifically, the computation of the appropriate value for α is as follows (Murphy & Myers 1998, p. 67):

$$\alpha_{\text{desired}} = \left[\frac{p(H_1)\beta}{1 - p(H_1)} \right] \left(\frac{1}{\text{DRS}} \right), \quad (1)$$

where $p(H_1)$ is the estimated probability that the alternative hypothesis is true (i.e., there is a relationship or effect in the population), β is the probability that one will incorrectly conclude that

the null hypothesis is true, and *DRS* is a judgment of the seriousness of making an α error vis-à-vis the seriousness of making a β error. As noted by Aguinis et al. (2010b), deciding on an appropriate *DRS* value could also be seen as an arbitrary process. Thus, “this decision should be well thought out and argued and not determined by fiat” (Aguinis et al. 2010b, p. 522).

Our third recommendation about how to address the low statistical power challenge is to adopt a Bayesian approach that avoids NHST altogether (Kruschke et al. 2012). This is a seemingly more radical recommendation given the emphasis on NHST in most doctoral-level programs (e.g., Aiken et al. 2008) and the newness of the Bayesian approach in the organizational sciences. In fact, Kruschke et al. (2012) content analyzed more than 10,000 articles published in 15 organizational science journals from January 2001 to December 2010 and reported that only 42 used Bayesian methods. As noted above, NHST provides information on the probability of obtaining the data in hand if the null hypothesis were true, or $p(D|H_0)$. However, a Bayesian view estimates the viability of a null hypothesis given the data, or $p(H_0|D)$. In other words, adopting a Bayesian approach allows us to know the credibility of candidate parameter values given the data that we actually observed. In short, Bayesian analysis is the mathematically normative way to reallocate credibility across parameter values as new data arrive and circumvents the need to use NHST (Kruschke et al. 2012).

In sum, there are fundamental concerns regarding the use of NHST. However, it may take decades until alternatives are widely known and used in organizational science research. Thus, in this section we address the important concern that organizational science research is often conducted at insufficient levels of statistical power. The implication is that, sadly, the typical probability that a researcher will find a hypothesized effect that actually exists is only around 50%—and even lower when hypotheses involve more complex relationships such as interaction and multilevel effects. Conducting research at insufficient levels of statistical power often leads to inconclusive results because we do not know whether lack of support for rejecting a null hypothesis is due to the nonexistence of the effect or to the fact that research design issues prevented the study from accurately detecting the effect. Thus, it is important to design studies such that there will be a good probability that existing population effects will be found.

Strengthening Inferences About Causal Relationships

There are extensive discussions of philosophical views underlying causation (for excellent reviews, see Cook & Campbell 1979, ch. 1; James et al. 1982, ch. 1). For current purposes, we adopt the view of James et al. (1982, p. 19) that

we infer the presence of causal relations when we isolate groups of variables into self-contained systems of functional equations on which the varying values of some variables (causes) appear to determine totally the varying values of other variables (effects) The inference of causation is an inductive inference based on presuming that the functional equations/relations describing the causal variables and effect variables observed in the past will continue to hold in the future.

Making causal inferences from our research findings is a highly desired end state because it is how we make research matter not only to us as researchers but also to others who make practical decisions based on the findings. Better decisions presumably stem from findings for which strong confidence exists as to the cause and effect of a phenomenon. The reality, though, is that such confidence does not exist in the majority of cases because of how the study was originally designed. Conducting experimental research involves many practical constraints and also often results in decreased external validity. Consequently, given the nature of organizational science research foci,

we are constrained for the most part with collecting data from the field. As such, the most frequently used research designs are passive observation studies in which we observe whether the rank order of values in one (or more) variable(s) is associated with the rank order of another variable(s). Causal inferences are improbable under those circumstances.

We believe, though, that with some adjustments to a study's design, it is possible to create conditions permitting stronger inferences about causality. Those adjustments include (*a*) adhering to the conditions of causality in making design decisions, (*b*) using quasi-experimental approaches, and (*c*) undertaking pure experimental research particularly on the focal *X* and *Y* variables. James et al. (1982) laid out the following 10 conditions required for causal inferences to be inferred:

1. Formal statement of theory in terms of a structural model
2. Theoretical rationale for causal hypotheses
3. Specification of causal order
4. Specification of causal direction
5. Self-contained functional equations
6. Specification of boundaries
7. Stability of the structural model
8. Operationalization of variables
9. Empirical confirmation of functional equations/hypothesized paths
10. Empirical confirmation of fit between the structural theoretical model and empirical data

A point of clarification is that the term structural above was not reserved for structural equation modeling as it would be viewed today. Rather, even a simple regression analysis is a structural model as long as the empirics were linked to confirming or testing some theoretical specification. We advocate that every one of these conditions serve as a critical benchmark against which to design a study or to evaluate a study under editorial review. Some are perhaps easier to fulfill than others. For example, as noted in the earlier section on making meaningful theoretical contributions, authors seem to be particularly adept at conditions 1 and 2. Further, they may make statements reflecting causal order and direction (conditions 3 and 4), although the design of the study may not have operationalized that order and direction in its execution. Indeed, an important point in our earlier section on theoretical progress was to lament how our discipline is failing at conditions 6, 7, and 8. The stability of the model (condition 7), for example, is at the heart of the issue of testing the validity of "new" theoretical contributions over time.

To increase the probability of fulfilling these conditions, we also advocate that researchers and editorial gatekeepers become thoroughly grounded in the principles of quasi-experimentation (Cook & Campbell 1979, Grant & Wall 2009, Shadish et al. 2002) and use these principles as benchmarks for study design and evaluation. Most published empirical work employs research designs alternatively labeled correlational, passive observational, or nonexperimental designs (Podsakoff & Dalton 1987, Scandura & Williams 2000). Regardless of the label, these are designs in which the cause and effect may be theoretically explicated, but other structural features of the design are missing, which precludes the researcher from actually making causal, counterfactual inferences from the results (Shadish et al. 2002). Among the missing features, for example, are lack of random assignment, not using pretests, and not having true control conditions. As noted by Cook & Campbell (1979), these conditions result in far too many threats to a study's validity to have confidence in that study's ability to evaluate a cause-effect association. However, as succinctly stated by Campbell & Stanley (1963, p. 34):

There are many natural social settings in which the research person can introduce something like experimental design into his scheduling of data collection procedures (e.g., the when and to whom of measurement), even though he lacks the full control over the scheduling of experimental stimuli (the when and to whom of exposure and the ability to randomize exposures) which makes a true experiment possible. Collectively, such situations can be regarded as quasi-experimental designs.

Cook & Campbell (1979) and Shadish et al. (2002) discussed in detail different design options that may be incorporated to turn a correlational study into a quasi-experiment. Incorporating such design options brings the researcher closer to making causal inferences from the findings even though some design options are better than others at doing so. In addition to the options in those sources, we would also encourage researchers to incorporate longitudinal designs with lagged growth models (Bentein et al. 2005, Ployhart & Vandenberg 2010); that is, one can lag a vector of change in the cause before (in terms of time) the vector of change in the effect. Such longitudinal designs require collecting data at time intervals that are theoretically consistent with the phenomenon being studied. For example, if one studies the socialization of new employees, data collection should take place during the first few weeks, and even days, of employment—collecting data a few months after the beginning of employment is unlikely to yield useful information even if collected at several points in time. Another option is to incorporate a latent class analysis procedure such as latent profile analysis to identify homogeneous groups within a sample (see Meyer et al. 2013). For example, Stanley et al. (2013) identified several commitment profiles in a sample ranging from not being committed at all on any commitment dimension to being highly committed across all dimensions. This allowed them to make theoretical comparisons using a between-group design on turnover intention and turnover behavior. Our major point is that with a bit more effort, researchers could turn “yet another” correlational study into one for which there is greater confidence in the cause and effect.

Using Control Variables Appropriately

Because most organizational science researchers use passive observation designs more frequently than any other type of design, alternative explanations cannot be eliminated through the study design features like they can with experimental and quasi-experimental designs (Shadish et al. 2002). Rather, reliance is placed on measuring alternative explanations and statistically controlling for them. However, the latter practice is actually not effective

unless much is already known about which alternative interpretations are plausible, unless those that are plausible can be validly measured, and unless the substantive model used for statistical adjustment is well-specified. These are difficult conditions to meet in the real world of research practice, and therefore many commentators doubt the potential of such designs to support strong causal inferences in most cases. (Shadish et al. 2002, p. 18)

As illustrated in many comprehensive reviews on the use of control variables (Atinc et al. 2012, Becker 2005, Braugh 2006, Carlson & Wu 2012, Edwards 2008, Spector & Brannick 2011), it is rare that authors of published work in which control variables are used have fulfilled the criteria outlined by Shadish and colleagues. Ironically, this is the same conclusion that Meehl (1971, p. 146) derived over 40 years ago when he stated that using control variables is “the commonest methodological vice in contemporary social sciences.” The concern or issue then, as it remains today, is whether the observed findings concerning the substantive variables of interest are valid or simply artifacts of including invalidly specified control variables.

Based on our current knowledge, researchers using control variables and editorial decision makers reviewing submissions in which control variables are used should adhere to and insist upon the following requirements. There should be (a) a strong conceptual explanation of why the control must be instituted, (b) a strong conceptual explanation of how it may come to influence the substantive variables of interest (both observed and latent) and the hypothesized relations among them, and (c) strong evidence regarding the psychometric properties of the measures used to assess controls—just as strong as that for measures used to assess variables of substantive interest. We purposely use the term requirement and not recommendation. Indeed, we feel so strongly about it that we are calling for editors and reviewers to put a stop to the practice of including control variables unless authors have met the above requirements.

It is not infrequent in published organizational science articles that the inclusion of one or more control variables is justified with language such as “so-and-so authors did so in their previous publications, and thus, I must include them as well because I am studying similar phenomena.” Another frequently used rationale is, “We thought it would simply be prudent to include them just in case they had an unknown effect.” And yet another frequently used justification is, “We found a relationship/difference between gender (or whatever the potential control variables may be) and our independent/predictor variable.” However, these reasons are insufficient justifications in the absence of conceptual evidence indicating why and if the control variables truly possess some influence on the variables of substantive interest and the associations among them (Williams et al. 2009).

At the heart of the control variable issue is their seemingly indiscriminate inclusion. According to Becker [2005; his findings were extended and reinforced recently in Atinc et al. (2012) and Carlson & Wu (2012)], over 63% of articles provided no to very unclear reasons for control variable use. Further, he found that nearly 50% of authors failed to explain how the control variable was operationalized, and nearly as many (48%) failed to discuss the quality of the operationalization’s reliability and validity. The net result is that a skeptical scientific audience does not have a firm conceptual understanding as to why a given control variable was included and why its absence would hinder an unambiguous interpretation of the underlying results. Further, readers are exposed to operationalizations of control variables possessing unknown measurement qualities. Thus, the possibility exists that the measures of the control variables are conceptually invalid and not representative of the underlying control variable constructs (as we discuss in more detail in the next section).

There are also a host of statistical issues that need to be considered. Specifically, the inclusion of control variables changes the substantive meaning of the focal constructs (Breaugh 2006, Edwards 2008). Once the variance of the control variables is removed from the focal variables, the remaining variance of the focal variables is nothing more than a residual term (Edwards 2008). Researchers often interpret these residuals in toto, as if the complete construct of the variables is present. Although the residual variances could be interpreted in various ways, this is certainly not the same as the whole construct before instituting control variables (Edwards 2008). Indeed, it is difficult to know what the residuals are in the absence of further studies examining the construct validity of each residual itself. Edwards (2008) also demonstrated the complications of using control variables with unknown error with respect to unambiguously interpreting the regression coefficients representing the associations among the variables of substantive interest. Also, not accounting for the causal structure among the control and focal variables during the analyses is tantamount to model misspecification. There could be any number of ways the control variables should be modeled, ranging from being main effects to moderators to mediators.

In short, we offer the following suggestions for authors. First, include an explanation as to why a control variable was selected and whether it is biasing in impact (per the reasoning of Spector et al. 2000) or has a substantive impact. In either case, explanations with supporting citations

should be provided, as doing so increases the rigor of the research (Edwards 2008). Second, beware of control variables that are uncorrelated with the endogenous-criterion-dependent variables, or what Becker (2005) referred to as “impotent control variables.” These variables reduce power in some cases, but in other cases they may increase the type I error rate by concluding that there is an effect when in reality there is not one. The latter situation occurs when the control variable also acts as a suppressor. Third, avoid the “throw them all into the analyses” approach. In the end, there may be so little residual variance left over that the substantive predictor variable will account for large proportions of a variance that may be miniscule to begin with. Most important, though, is the change in meaning of the substantive variables of interest with the inclusion of each control variable (Edwards 2008). Fourth, clearly and concisely explain how each control variable was operationalized and why that operationalization represents the construct or conceptual domain of the control variable. In other words, the same standards of validity and reliability should be applied to the control variables as are applied to the substantive variables of interest. Fifth, the analyses should reflect the causal structure among the control variables and the substantive variables of interest. Sixth, include the descriptive statistics for the control variables in the tables of means, standard deviations, and correlations. Additionally, include the effects of the control variables in the same tables or figures used to report the parameter estimates relating the substantive variables to one another. Finally, models with and without the control variables should be evaluated. If the model without the control variables fits as well as that with them in (note that these are not nested models, and thus a chi-square difference test is inappropriate) and the same parameter estimates are statistically significant, then one may move forward with the interpretation of the model without the control variables. However, if that is not the case and the model with the control variables is the optimal one, then the substantive arguments for their inclusion become more persuasive. Thus, in the next iteration of research in this stream, these control variables now become substantive variables.

MEASUREMENT

Improving the Link Between Underlying Constructs and Their Observable Indicators

The majority of organizational science theories include constructs—concepts that are abstract and not directly observable (Bagozzi & Edwards 1998, Boyd et al. 2013). Notable examples include job satisfaction, organizational commitment, and performance. These and many others are often conceptualized as latent constructs, which we measure by collecting data on observable indicators. In other words, we infer the extent to which an individual is satisfied with her job not by directly observing her job satisfaction, but by collecting data on observable indicators that we believe represent the unobservable construct of job satisfaction. Such indicators might include, for example, this employee’s answers to a multi-item survey or interviews with her peers.

Even though we cannot directly observe the constructs, our goal is to draw inferences from the constructs of interest based on the statistical associations we observe among the operationalizations (i.e., measures) of those constructs (Le et al. 2009). An important challenge faced by organizational science researchers is that measures of latent constructs are fallible (Schmidt 1992), and because they are only imperfect indicators of underlying constructs, observed relationships between constructs are usually smaller than the actual relationships. In fact, Aguinis et al. (2011) conducted a review of almost 6,000 correlation coefficients reported in the *Academy of Management Journal*, *Journal of Applied Psychology*, *Journal of Management*, *Personnel Psychology*, and *Strategic Management Journal* from 1982 through August 2009 and found that the median absolute effect across all types of bivariate relationships is only .23. Interestingly, this value is very

close to .20, which is what Lykken (1968) guessed was the expected correlation between any two not necessarily related variables in psychological research. So, observed bivariate effects explain only about 4% of variance in relevant outcomes even as reported in some of the most prestigious organizational science journals.

Construct validity refers to the confidence researchers have that the indicators are good proxies for their presumed underlying constructs (Cook & Campbell 1979, Shadish et al. 2002). It is important to gather construct validity evidence prior to assessing substantive relationships because absent such evidence, the presence of a small effect could be either real or, alternatively, due to the error in the measures (Binning & Barrett 1989). In other words, absent strong construct validity evidence, substantive results are inconclusive. Unfortunately, few published articles in organizational science journals provide sufficient evidence to draw solid conclusions regarding construct validity (Podsakoff & Dalton 1987, Scandura & Williams 2000).

There is a voluminous literature regarding how to strengthen construct validity (e.g., Bagozzi et al. 1991, Churchill 1979, Cronbach & Meehl 1955, Edwards 2003, Hinkin 1998, Nunnally & Bernstein 1994, Schriesheim et al. 1993, Shadish et al. 2002, Stone-Romero 1994). This literature addresses the construction of new scales using a process that maximizes the ability of a scale to assess the targeted construct. In what follows, we offer recommendations that rely on these sources, focusing on more recent advancements. First, however, we emphasize the need to define constructs carefully (MacKenzie 2002). In contrast to the majority of terms used in the physical and biological sciences, terms used to define constructs in organizational science research are also used in daily life. For example, labels such as leadership, intelligence, power, and performance are used in people's work or social conversations and by the media on an ongoing basis. A clear and unambiguous conceptual definition of a construct is required for the development of good measures in part to disentangle it from the everyday implicit definitions held by people. Not having one results in a downward spiral involving measurement deficiency (i.e., the measure does not capture all aspects of the construct) and measurement contamination (i.e., the measure captures aspects that are not part of the construct). In turn, deficiency and contamination lead to disappointing results in the form of no or small observed effects when in reality those effects might be much larger (Dalton & Aguinis 2013). In some cases, contamination can also lead to upwardly biased effect sizes, as in the case of a test with race group differences being validated against supervisory ratings of performance that are racially biased. Defining a construct adequately involves both specifying the conceptual theme in unambiguous terms and consistently with previous research and distinguishing it from related constructs (MacKenzie 2002). Also, once a scale has been created, any alteration such as changing the wording for some items or eliminating others from the scale means that the resulting scale may no longer be accurately measuring the underlying construct that it was originally created to assess (Schriesheim et al. 1993).

In addition to the traditional scale development process, technological advancements have allowed for innovative ways to address the challenge of improving the link between underlying constructs and their indicators. One of these advancements is the use of computer-aided text analysis (CATA) (Short et al. 2010). CATA is actually a family of computer-based tools used to gather text-based information from sources such as organizational mission statements, CEOs' letters to shareholders, and firms' annual reports. Short et al. (2010) illustrated how to use CATA to gather evidence regarding linkages between observable indicators and their underlying constructs. The first step involves deductively deriving a word list, and the second step involves inductively deriving additional word lists. Finally, scores based on the absence/presence and frequency of words are compared with external criteria (e.g., firm performance). Note that CATA can also be used to measure and gather validity evidence regarding multidimensional as well as multilevel constructs (McKenny et al. 2012).

Another technological advancement is the use of video-based technology for creating manipulations that in turn are used to gather evidence regarding the extent to which a measure assesses the intended construct (Podsakoff et al. 2013). Specifically, video-based technology can be used to gather evidence regarding the construct validity of measures used to assess such constructs as leadership behaviors, influence tactics, and various aspects of performance (e.g., task performance, contextual performance). The creation of these videos involves first defining the construct and then creating scripts (e.g., female and male managers behaving assertively or unassertively) (e.g., Aguinis & Adams 1998). Participants are exposed to different scripts (e.g., assertive versus unassertive manager) and then complete the intended measure (e.g., managerial assertiveness scale). Essentially, this step involves assessing the effects of the video manipulations on measures (e.g., leadership assertiveness).

Video-based technology also has distinct benefits over the classical and traditional way to gather construct validity evidence through the multitrait-multimethod matrix method (MTMM; Campbell & Fiske 1959, Woehr et al. 2012). In brief, MTMM consists of constructing a correlation matrix involving scores for at least two constructs collected using at least two methods each. Evidence of good correspondence between a construct and its indicators is found if (a) there is a relationship between different measures assessing the same construct (i.e., convergent validity) and (b) there is no relationship between scores collected using the same type of measure but assessing different constructs (i.e., discriminant validity). A pattern of results that does not support convergent and discriminant validity evidence suggests that there is a distal relationship between a construct and its indicators. Moreover, it may also suggest that the measures not only are inadequate for the construct in question but, even worse, assess a different—unintended—construct. An important advantage of the video-based approach over classic correlation-based approaches for gathering validity evidence is that changes in scores are directly attributed to the manipulations. Also, participants are randomly assigned to conditions (i.e., different manipulation levels), and there is temporal precedence such that video-based manipulations precede the measures. Taken together, these advantages lead to stronger evidence regarding a measure's ability to assess intended construct that are not self-referential in nature.

In sum, a consistent result in organizational science research is that observed effects are smaller than hypothesized ones (Dalton & Aguinis 2013). A likely explanation is that measures are proxies too distal from their underlying constructs. Unfortunately, organizational science researchers do not pay sufficient attention to this crucial issue. In addition to best-practice recommendations regarding scale development that have been available for quite some time (e.g., Hinkin 1998), recent technological advances have allowed for novel approaches such as CATA and video-based techniques, all of which can complement more classic approaches to produce better measures as well as make stronger inferences regarding the validity of existing ones.

CONCLUSIONS

As organizational science researchers, we have an ethical obligation to conduct high-quality research because otherwise we are wasting valuable resources (Rosenthal 1994). In addition, it is our professional obligation to produce evidence-based advice for practitioners and decision makers in general (e.g., policy makers, managers, legislative officials). In our combined experience as journal editors and editorial board members, we have found that too many organizational science researchers begin their data analysis journey with a losing hand. Accordingly, we have offered recommendations on what researchers can do prior to data collection. We are aware that following our recommendations will require investing additional time and effort at the front end of the research process. Moreover, we have chosen an admittedly selected set of issues and have not

addressed others, such as sampling (e.g., representativeness, generalizability) and pilot testing of measures and manipulations. However, investment in theory, design, and measurement issues is likely to yield a much greater return compared with investment in the data analysis stage. As noted by Sir Ronald Fisher (1938, p. 17), “Immensely laborious calculations on inferior data may increase the yield . . . [by] 5 per cent of perhaps a small total. A competent overhauling of the process of collection, or of the experimental design, may often increase the yield ten or twelve fold, for the same cost in time and labour.” We hope our recommendations will lead to higher-quality research that will not only help advance organizational science theories but also improve the impact and effectiveness of subsequent practical applications.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Lillian Eby, Fred Morgeson, Ann Marie Ryan, Neal Schmitt, and Dean Shepherd for highly constructive feedback on previous versions of our article.

LITERATURE CITED

- Acad. Manag. 2008. *The questions we ask, Anaheim, California, August 18–13, 2008*. Progr. Guide, Acad. Manag., Briarcliff Manor, NY. http://program.aomonline.org/2008/pdf/AOM_2008_Annual_Meeting_Program.pdf
- Aguinis H. 2004. *Regression Analysis for Categorical Moderators*. New York: Guilford
- Aguinis H. 2013. *Performance Management*. Upper Saddle River, NJ: Pearson Prentice Hall. 3rd ed.
- Aguinis H, Adams SKR. 1998. Social-role versus structural models of gender and influence use in organizations: a strong inference approach. *Group Organ. Manag.* 23:414–46
- Aguinis H, Beaty JC, Boik RJ, Pierce CA. 2005. Effect size and power in assessing moderating effects of categorical variables using multiple regression: a 30-year review. *J. Appl. Psychol.* 90:94–107
- Aguinis H, Culpepper SA, Pierce CA. 2010a. Revival of test bias research in preemployment testing. *J. Appl. Psychol.* 95:648–80
- Aguinis H, Dalton DR, Bosco FA, Pierce CA, Dalton CM. 2011. Meta-analytic choices and judgment calls: implications for theory building and testing, obtained effect sizes, and scholarly impact. *J. Manag.* 37:5–38
- Aguinis H, Forcum LE, Joo H. 2013a. Using market basket analysis in management research. *J. Manag.* 39:1799–824
- Aguinis H, Gottfredson RK, Culpepper SA. 2013b. Best-practice recommendations for estimating cross-level interaction effects using multilevel modeling. *J. Manag.* 39:1490–528
- Aguinis H, Pierce CA, Bosco FA, Muslin IS. 2009. First decade of *Organizational Research Methods*: trends in design, measurement, and data-analysis topics. *Organ. Res. Methods* 12:69–112
- Aguinis H, Werner S, Abbott JL, Angert C, Park JH, Kohlhausen D. 2010b. Customer-centric science: reporting significant research results with rigor, relevance, and practical impact in mind. *Organ. Res. Methods* 13:515–39
- Aiken LS, West SG, Millsap RE. 2008. Doctoral training in statistics, measurement, and methodology in psychology: replication and extension of the Aiken, West, Sechrest, and Reno (1990) survey of Ph.D. programs in North America. *Am. Psychol.* 63:32–50
- Aiken LS, West SG, Sechrest LB, Reno RR. 1990. Graduate training in statistics, methodology, and measurement in psychology: an international survey. *Am. Psychol.* 45:721–34
- Anderson N. 2007. The practitioner-researcher divide revisited: strategic-level bridges and the roles of IWO psychologists. *J. Occup. Organ. Psychol.* 80:175–83

- Ashkanasy NM. 2010. Publishing today is more difficult than ever. *J. Organ. Behav.* 31:1–3
- Atinc G, Simmering MJ, Kroll MJ. 2012. Control variable use and reporting in macro and micro management research. *Organ. Res. Methods* 15:57–74
- Bagozzi RP, Edwards JR. 1998. A general approach to construct validation in organizational research: application to the measurement of work values. *Organ. Res. Methods* 1:45–87
- Bagozzi RP, Yi Y, Phillips LW. 1991. Assessing construct validity in organizational research. *Adm. Sci. Q.* 36:421–58
- Barnett ML. 2007. (Un)learning and (mis)education through the eyes of Bill Starbuck: an interview with Pandora's playmate. *Acad. Manag. Learn. Educ.* 6:114–27
- Bartunek JM, Rynes SL, Ireland RD. 2006. What makes management research interesting, and why does it matter? *Acad. Manag. J.* 49:9–15
- Becker TE. 2005. Potential problems in the statistical control of variables in organizational research: a qualitative analysis with recommendations. *Organ. Res. Methods* 8:274–89
- Bentein K, Vandenberghe C, Vandenberg R, Stinglhamber F. 2005. The role of change in the relationship between commitment and turnover: A latent growth modeling approach. *J. Appl. Psychol.* 90:468–82
- Bergh DD. 2008. The developmental editor: assessing and detecting manuscript contribution. In *Opening the Black Box of Editorship*, ed. Y Baruch, AM Konrad, H Aguinis, WH Starbuck, pp. 114–23. San Francisco: Jossey Bass
- Binning JF, Barrett GV. 1989. Validity of personnel decisions: a conceptual analysis of the inferential and evidential bases. *J. Appl. Psychol.* 74:478–94
- Bosker RJ, Snijders TAB, Guldemond H. 2003. *PINT (Power IN Two-Level Designs): Estimating Standard Errors of Regression Coefficients in Hierarchical Linear Models for Power Calculations, Version 2.1*. Groningen, Neth.: Rijksuniv. Groningen
- Boyd BK, Bergh DD, Ireland RD, Ketchen DJ. 2013. Constructs in strategic management. *Organ. Res. Methods* 16:3–14
- Breaugh JA. 2006. Rethinking the control of nuisance variables in theory testing. *J. Bus. Psychol.* 20:429–43
- Campbell DT, Fiske DW. 1959. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.* 56:81–105
- Campbell DT, Stanley JC. 1963. *Experimental and Quasi-experimental Designs for Research*. Chicago: Rand McNally
- Carlson KD, Wu J. 2012. The illusion of statistical control: control variable practice in management research. *Organ. Res. Methods* 15:413–35
- Cascio WF, Aguinis H. 2008. Research in industrial and organizational psychology from 1963 to 2007: changes, choices, and trends. *J. Appl. Psychol.* 93:1062–81
- Cascio WF, Zedeck S. 1983. Open a new window in rational research planning: Adjust alpha to maximize statistical power. *Pers. Psychol.* 83:517–26
- Churchill GA. 1979. A paradigm for developing better measures of marketing constructs. *J. Mark. Res.* 16:64–73
- Certo ST, Sirmon DG, Brymer R. 2010. Competition and knowledge creation in management: investigating changes in scholarship from 1988 to 2007. *Acad. Manag. Learn. Educ.* 9:591–606
- Cohen J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum. 2nd ed.
- Cohen J. 1994. The earth is round ($p < 0.05$). *Am. Psychol.* 49:997–1003
- Colquitt JA, Zapata-Phelan CP. 2007. Trends in theory building and theory testing: a five-decade study of the *Academy of Management Journal*. *Acad. Manag. J.* 50:1281–303
- Cook TD, Campbell DT. 1979. *Quasi-experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand McNally
- Cronbach L, Meehl P. 1955. Construct validity in psychological tests. *Psychol. Bull.* 52:281–302
- Dalton DR, Aguinis H. 2013. Measurement malaise in strategic management studies: the case of corporate governance research. *Organ. Res. Methods* 16:88–99
- Davis GF. 2010. Do theories of organizations progress? *Organ. Res. Methods* 13:690–709
- Davis GF, Marquis C. 2005. Prospects for organization theory in the early 21st century: institutional fields and mechanisms. *Organ. Sci.* 16:332–43

- Davis MS. 1971. That's interesting! Towards a phenomenology of sociology and a sociology of phenomenology. *Philos. Soc. Sci.* 1:309–44
- Dunnette MD. 1990. Blending the science and practice of industrial and organizational psychology: Where are we and where are we going? In *Handbook of Industrial and Organizational Psychology*, Vol. 1, ed. MD Dunnette, LM Hough, pp. 1–27. Palo Alto, CA: Consult. Psychol. 2nd ed.
- Eby LT, Hurst CS, Butts MM. 2009. Qualitative research: the redheaded stepchild in organizational and social research? In *Statistical and Methodological Myths and Urban Legends: Received Doctrine, Verity, and Fable in the Organizational and Social Sciences*, ed. CE Lance, RJ Vandenberg, pp. 219–46. New York: Routledge
- Edwards JR. 2003. Construct validation in organizational behavior research. In *Organizational Behavior: The State of the Science*, ed. J Greenberg, pp. 327–71. Mahwah, NJ: Lawrence Erlbaum. 2nd ed.
- Edwards JR. 2008. To prosper, organizational psychology should...overcome methodological barriers to progress. *J. Organ. Behav.* 29:469–91
- Edwards JR. 2010. Reconsidering theoretical progress in organizational and management research. *Organ. Res. Methods* 13:615–19
- Edwards JR, Berry JW. 2010. The presence of something or the absence of nothing: increasing theoretical precision in management research. *Organ. Res. Methods* 13:668–89
- Fisher RA. 1925. *Statistical Methods for Research Workers*. Edinburgh, UK: Oliver & Boyd
- Fisher RA. 1938. Presidential address. *Sankhyā: Indian J. Stat.* 4:14–17
- Gigerenzer G. 1998. We need statistical thinking, not statistical rituals. *Behav. Brain Sci.* 21:199–200
- Glaser BG. 1992. *Basics of Grounded Theory Analysis: Emerging Versus Forcing*. Mill Valley, CA: Sociol. Press
- Grant AM, Wall TD. 2009. The neglected science and art of quasi-experimentation: why-to, when-to, and how-to advice for organizational researchers. *Organ. Res. Methods* 12:653–86
- Gray PH, Cooper WH. 2010. Pursuing failure. *Organ. Res. Methods* 13:620–43
- Hakel MD, Sorcher M, Beer M, Moses JL. 1982. *Making It Happen: Designing Research with Implementation in Mind*. Beverly Hills, CA: Sage
- Hambrick DC. 1994. 1993 Presidential address: What if the Academy actually mattered? *Acad. Manag. Rev.* 19:11–16
- Hambrick DC. 2007. The field of management's devotion to theory: too much of a good thing? *Acad. Manag. J.* 50:1346–52
- Hamming R. 1986. *You and your research*. Presented at Bell Commun. Res. Colloq. Semin., Mar. 7, Morristown, NJ
- Hinkin TR. 1998. A brief tutorial on the development of measures for use in survey questionnaires. *Organ. Res. Methods* 1:104–21
- Hollenbeck JR. 2008. The role of editing in knowledge development: consensus shifting and consensus creation. In *Opening the Black Box of Editorship*, ed. Y Baruch, AM Konrad, H Aguinis, WH Starbuck, pp. 16–26. San Francisco: Jossey Bass
- Hunter JE. 1997. Needed: a ban on the significance test. *Psychol. Sci.* 8:3–7
- James LR, Mulaik SA, Brett JM. 1982. *Causal Analysis: Assumptions, Models and Data*. Beverly Hills, CA: Sage
- Kacmar KM, Whitfield JM. 2000. An additional rating method for journal articles in the field of management. *Organ. Res. Methods* 3:392–406
- Kruschke JK, Aguinis H, Joo H. 2012. The time has come: Bayesian methods for data analysis in the organizational sciences. *Organ. Res. Methods* 15:722–52
- Lakatos I. 1978. Falsification and the methodology of scientific research programmes. In *The Methodology of Scientific Research Programmes*, Vol. 1, ed. J Worrall, G Currie, pp. 170–96. Cambridge, UK: Cambridge Univ. Press
- Lance CE. 2011. More statistical and methodological myths and urban legends. *Organ. Res. Methods* 14:279–86
- Le H, Schmidt FL, Putka DJ. 2009. The multifaceted nature of measurement artifacts and its implications for estimating construct-level relationships. *Organ. Res. Methods* 12:165–200

- Leavitt K, Mitchell RR, Peterson J. 2010. Theory pruning: strategies to reduce our dense theoretical landscape. *Organ. Res. Methods* 13:644–67
- Little J. 2001. Understanding statistical significance: a conceptual history. *J. Tech. Writ. Commun.* 31:363–72
- Locke EA. 2007. The case for inductive theory building. *J. Manag.* 33:867–90
- Lykken DT. 1968. Statistical significance in psychological research. *Psychol. Bull.* 70:151–59
- MacKenzie SB. 2002. The dangers of poor construct conceptualization. *J. Acad. Mark. Sci.* 31:323–26
- Mathieu JE, Aguinis H, Culpepper SA, Chen G. 2012. Understanding and estimating the power to detect cross-level interaction effects in multilevel modeling. *J. Appl. Psychol.* 97:951–66
- McKenny AF, Short JC, Payne GT. 2012. Using computer-aided text analysis to elevate constructs: an illustration using psychological capital. *Organ. Res. Methods* 16:152–84
- Maxwell SE. 2004. The persistence of underpowered studies in psychological research: causes, consequences, and remedies. *Psychol. Methods* 9:147–63
- Meehl PE. 1971. High school yearbooks: a reply to Schwartz. *J. Abnorm. Psychol.* 77:143–48
- Meehl PE. 1978. Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *J. Consult. Clin. Psychol.* 46:806–34
- Meyer JP, Stanley LA, Vandenberg RJ. 2013. A person-centered approach to the study of commitment. *Hum. Resour. Manag. Rev.* 23:190–202
- Mone MA, Mueller GC, Mauland W. 1996. The perceptions and usage of statistical power in applied psychology and management research. *Pers. Psychol.* 49:103–20
- Murphy KR, Myers B. 1998. *Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis Tests*. Mahwah, NJ: Lawrence Erlbaum
- Nunnally JC, Bernstein IH. 1994. *Psychometric Theory*. New York: McGraw-Hill. 3rd ed.
- Orlitzky M. 2012. How can significance tests be deinstitutionalized? *Organ. Res. Methods* 15:199–228
- Pfeffer J. 1993. Barriers to the advance of organizational science: paradigm development as a dependent variable. *Acad. Manag. Rev.* 18:599–620
- Pfeffer J. 2007. A modest proposal: how we might change the process and product of managerial research. *Acad. Manag. J.* 50:1334–45
- Pfeffer J, Fong CT. 2005. Building organization theory from first principles: the self-enhancement motive and understanding power and influence. *Organ. Sci.* 16:372–88
- Pierce JR, Aguinis H. 2013. The too-much-of-a-good-thing effect in management. *J. Manag.* 39:313–38
- Ployhart R, Vandenberg RJ. 2010. A review and synthesis of longitudinal data analytical procedures in the organizational sciences. *J. Manag.* 36:94–120
- Podsakoff PM, Dalton DR. 1987. Research methodology in organizational studies. *J. Manag.* 13:419–41
- Podsakoff NP, Podsakoff PM, MacKenzie SB, Klinger RL. 2013. Are we really measuring what we say we're measuring? Using video techniques to supplement traditional construct validation procedures. *J. Appl. Psychol.* 98:99–113
- Pollack I. 2012. Taming textual data: the contribution of corpus linguistics to computer-aided text analysis. *Organ. Res. Methods* 15:263–87
- Popper K. 1963. *Conjectures and Refutations: The Growth of Scientific Knowledge*. New York: Harper Torchbooks
- Rosenthal R. 1994. Science and ethics in conducting, analyzing, and reporting psychological research. *Psychol. Sci.* 5:127–34
- Rosnow RL, Rosenthal R. 1989. Statistical procedures and the justification of knowledge in psychological science. *Am. Psychol.* 44:1276–84
- Rousseau DM. 2007. A sticky, levering, and scalable strategy for high-quality connections between organizational practice and science. *Acad. Manag. J.* 50:1037–42
- Ruback RA, Innes CA. 1988. The relevance and irrelevance of psychological research: the example of prison crowding. *Am. Psychol.* 43:683–93
- Ryan AM, Schmitt N, Spector P, Zedeck S, Rogelberg S, eds. 2014. Inductive research in organizations. Special issue, *J. Bus. Psychol.* In press
- Rynes SL, Colbert AE, Brown KG. 2002. HR professionals' beliefs about effective human resource practices: correspondence between research and practice. *Hum. Resour. Manag.* 41:149–74

- Scandura TA, Williams EA. 2000. Research methodology in management: current practices, trends, and implications for future research. *Acad. Manag. J.* 43:1248–64
- Schmidt FL. 1992. What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *Am. Psychol.* 47:1173–81
- Schriesheim CA, Powers KJ, Scandura TA, Gardiner CC, Lankau MJ. 1993. Improving construct measurement in management research: comments and a quantitative approach for assessing the theoretical content adequacy of paper-and-pencil survey-type instruments. *J. Manag.* 19:385–417
- Shadish WR, Cook TD, Campbell DT. 2002. *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin
- Shapiro DL, Kirkman BL, Courtney HG. 2007. Perceived causes and solutions of the translation problem in management research. *Acad. Manag. J.* 50:249–66
- Shepherd DA, Sutcliffe KM. 2011. Inductive top-down theorizing: a source of new theories of organization. *Acad. Manag. Rev.* 36:361–80
- Short JC, Broberg JC, Coglisier CC, Brigham KH. 2010. Construct validation using computer-aided text analysis (CATA): an illustration using entrepreneurial orientation. *Organ. Res. Methods* 13:320–47
- Simon H. 1996 (1969). *The Sciences of the Artificial*. Boston: MIT Press. 3rd ed.
- Spector PE, Brannick MT. 2011. Methodological urban legends: the myth of statistical control variables. *Organ. Res. Methods* 14:287–305
- Spector PE, Zapf D, Chen PY, Frese M. 2000. Why negative affectivity should not be controlled in job stress research: Don't throw out the baby with the bath water. *J. Organ. Behav.* 21:79–95
- Spybrook J, Bloom H, Congdon R, Hill C, Martinez A, Raudenbush S. 2011. *Optimal Design Plus Empirical Evidence: Documentation for the "Optimal Design" Software (Version 3.0)*. New York: William T. Grant Found. <http://www.wtgrantfoundation.org/resources/consultation-service-and-optimal-design>
- Stanley LA, Vandenberghe C, Vandenberg RJ, Bentein K. 2013. Commitment profiles and employee turnover. *J. Vocat. Behav.* 82:176–87
- Starbuck WH. 2004. Why I stopped trying to understand the real world. *Organ. Stud.* 25:1233–54
- Stone-Romero EF. 1994. Construct validity issues in organizational behavior research. In *Organizational Behavior: The State of the Science*, ed. J Greenberg, pp. 155–79. Hillsdale, NJ: Lawrence Erlbaum
- Tett RP, Walser B, Brown C, Simonet DV, Tonidandel S. 2013. The 2011 SIOP graduate program benchmarking survey part 3: curriculum and competencies. *Ind.-Organ. Psychol.* 50(4):69–89
- Tushman M, O'Reilly C III. 2007. Research and relevance: implications of Pasteur's quadrant for doctoral programs and faculty development. *Acad. Manag. J.* 50:769–74
- Van Aken JE, Romme AGL. 2012. A design science approach to evidence-based management. In *The Oxford Handbook of Evidence-Based Management*, ed. DM Rousseau, pp. 43–57. New York: Oxford Univ. Press
- Walsh JP. 2011. Presidential address: embracing the sacred in our secular scholarly world. *Acad. Manag. Rev.* 36:215–34
- Weick KE. 1989. Theory construction as disciplined imagination. *Acad. Manag. Rev.* 14:516–31
- Williams LJ, Vandenberg RJ, Edwards JR. 2009. Structural equation modeling in management research: a guide for improved analysis. *Acad. Manag. Ann.* 3:543–604
- Woehr DJ, Putka DJ, Bowler MC. 2012. An examination of G-theory methods for modeling multitrait-multimethod data: clarifying links to construct validity and confirmatory factor analysis. *Organ. Res. Methods* 15:134–61