

Methodological Wishes for the Next Decade and How to Make Wishes Come True

Herman Aguinis and Jeffrey R. Edwards

Indiana University; University of North Carolina at Chapel Hill

ABSTRACT We discuss seven methodological improvements that would stimulate important advancements in management research. We refer to these improvements as ‘wishes’ that we hope will materialize within the next decade. To promote the implementation of these improvements, we offer concrete and actionable recommendations that researchers can apply when designing and conducting empirical research and that journal editors and reviewers can consider when evaluating manuscripts for publication. These improvements address: (1) accelerating theoretical progress; (2) improving the construct validity of measures; (3) strengthening causal inferences; (4) incorporating multilevel design, measurement, and analysis; (5) balancing trade-offs between internal and external validity, (6) understanding the nature and impact of outliers; and (7) offering a realistic and useful description of a study’s limitations.

Keywords: empirical research, measurement, methodology, research design

INTRODUCTION

Since the establishment of the *Journal of Management Studies* (*JMS*) 50 years ago, the field of management has made important advancements. In the intervening 50 years, we have developed and tested numerous theories in strategy, organizational behaviour, human resource management, entrepreneurship, and other domains. Nevertheless, compared to other scientific fields such as physics, biology, and astronomy, management still has yet to establish scientific paradigms that garner widespread consensus (Pfeffer, 1993) or make significant theoretical progress (Davis, 2010). There are many reasons for this state of affairs. One reason is that management is a relatively young scientific domain. As a consequence, management is still in the process of developing and refining methodological procedures and tools that would guide the development of strong paradigms and facilitate theoretical advancements (Boyd et al., 2008; Edwards, 2008; Hitt et al., 2004).

Address for reprints: Herman Aguinis, Department of Management and Entrepreneurship, Kelley School of Business, Indiana University, 1309 E. 10th Street, Bloomington, IN 47405-1701, USA (haguinis@indiana.edu).

Methodological improvements are essential for the progress of management research (Van Maanen et al., 2007). The importance of research methodology has prompted several institutional developments in the field of management. For instance, in 1985 the Academy of Management (AOM) established an interest group devoted to methodology that developed into the Research Methods Division, which is now one of the largest divisions in the AOM. In addition, *Organizational Research Methods*, founded in 1998, is an AOM-sponsored journal exclusively devoted to methodology. Moreover, substantive journals in both micro and macro domains have published articles and special issues on methodology (e.g., Aguinis et al., 2009; Podsakoff and Dalton, 1987; Scandura and Williams, 2000). These are important developments that have arguably helped the field move forward. Nonetheless, methodological advancements tend to disseminate slowly, given that we as management researchers tend to have our own methodological comfort zones (Aguinis et al., 2009). Consequently, there seems to be a 'scientific community's persistence in the use of particular methods' (Podsakoff and Dalton, 1987, p. 433) that becomes habitual, if not ritualistic, and can be difficult to dislodge.

As we initially framed this article, we developed and discussed an extensive 'wish list' of methodological improvements that, if implemented, would allow the field of management to make important advancements. We subsequently requested feedback on our 'wish list' from several senior scholars in our field, including current and past editors of *JMS* and other journals. This process distilled our list into the following seven methodological improvements: accelerating theoretical progress (wish 1); improving the construct validity of measures (wish 2); strengthening causal inferences (wish 3); incorporating multilevel design, measurement, and analysis (wish 4); balancing trade-offs between internal and external validity (wish 5); understanding the nature and impact of outliers (wish 6); and offering a realistic and useful description of a study's limitations (wish 7). We refer to these potential improvements as 'wishes' that we hope will come true within the coming decade. Our wishes reflect the sequence of steps involved in conducting empirical research (e.g., Runkel and McGrath, 1972), which include formulating hypotheses (i.e., wish 1), developing and evaluating measures (i.e., wish 2), making crucial design decisions (i.e., wishes 3 to 5), analysing data (i.e., wish 6), and drawing appropriate inferences (i.e., wish 7). To help management researchers implement these improvements, we offer concrete and actionable recommendations to guide the design and conduct of empirical research and inform criteria used to evaluate manuscripts during the review process. As a preview, Table I includes our list of these wishes as well as a brief summary of our recommendations.

Before proceeding, we should clarify the epistemological assumptions that underlie our recommendations. Stated succinctly, we adopt the perspective that the purpose of management research is to generate knowledge that is valid as well as relevant from a practical standpoint. To these ends, the questions pursued by management research should draw from relevant theory and be anchored in issues relevant to the practice of management. We view management research as a scientific enterprise that grapples with methodological issues that confront the social sciences in general, many of which are rooted in the complexity and malleability of the phenomena that characterize individuals, groups, and organizations. We do not presume that management research can match the rigour and precision of research in the hard sciences, nor do we think

Table I. Summary of methodological wishes and recommendations for researchers as well as journal editors and reviewers for making wishes come true

<i>Methodological wish</i>	<i>Recommendations</i>
1. Researchers will strive to accelerate the progress of management theories.	<ul style="list-style-type: none"> • Develop theories that yield hypotheses that are more precise in describing the nature of the effect rather than merely stating that an effect will differ from zero. • Expose theories to stringent tests that put them at genuine risk of falsification. • When possible, test hypotheses that posit lower and upper bounds for effects, non-null effects, zones of indifference for effects, competitive predictions (i.e., strong inference), and a 'good enough' belt for effects. • Move beyond null hypothesis significance testing by reporting effect sizes along with confidence intervals and applying Bayesian approaches to data analysis.
2. Researchers will improve the construct validity of measures used in management research.	<ul style="list-style-type: none"> • Clearly define constructs and evaluate the conceptual linkages between constructs and measures. • Carefully specify the nature and direction of the relationships between constructs and measures. • Empirically evaluate the construct validity of measures using confirmatory factor analysis.
3. Researchers will strengthen the causal inferences that can be drawn from management research.	<ul style="list-style-type: none"> • When possible, conduct randomized experiments, which are the gold standard for drawing causal inferences. • Apply matching techniques, in which groups are formed that contain units with similar scores on variables that represent alternative explanations for the causal effect of interest. • Statistically control for variables believed to represent alternative explanations for the causal effect of interest. • Use instrumental variables to address the effects of omitting variables that are correlated with the causal variables of interest.
4. Researchers will design empirical studies, measure constructs, and analyse the resulting data adopting multilevel approaches.	<ul style="list-style-type: none"> • Understand whether there is dependence in a data structure, even if entities are not formally embedded within collectives. • Consider hypothesizing and testing conceptual models that involve antecedent and outcome variables residing at different levels of analysis. • Consider hypothesizing and testing relationship involving different levels of analysis such as cross-level direct and cross-level interaction effects and including outcome variables at the lower and higher levels of analysis.
5. Researchers will design empirical studies that will balance usual trade-offs between internal (i.e., confidence in inferences about causal relationships) and external validity (i.e., confidence in inferences about generalizability of relationships).	<ul style="list-style-type: none"> • Design a study so as to not be put in a dilemma maximizing internal validity (e.g., experimental manipulation and control) at the expense of external validity (e.g., ability to generalize results), or vice versa. • Be aware of the trade-offs involved between internal and external validity • Become aware of technological advancements that allow for the design of studies that strike a good balance between internal and external validity such as experience sampling methodology, eLancing, and virtual reality. • Take advantage of technological advancements to design studies that are strong regarding both internal and external validity with the goal of bridging the micro-macro gap and making contributions to research streams such as upper echelons, micro-foundations, and strategy-as-practice.
6. Researchers will use systematic and standardized procedures for defining, identifying, and handling outliers and these procedures will be described clearly, perhaps in a section titled 'Outlier Detection and Management', in all manuscripts.	<ul style="list-style-type: none"> • Describe each choice and procedure implemented regarding outliers in detail to ensure transparency and include a rationale for the particular procedures that have been implemented. • Describe and acknowledge the particular type(s) of outliers in which there is an interest (i.e., error, interesting, influential) and then use an identification technique that is congruent with the outlier definition. • Follow best-practice recommendations regarding how to identify outliers by following a sequence of steps including first using procedures for error outliers, next for interesting outliers and, finally, for influential outliers. • Follow best-practice recommendations regarding how to handle outliers. For error outliers, adjust or omit cases. For interesting outliers, study them by using quantitative and qualitative approaches. For influential outliers, there are three options: model respecification, case deletion, and implementing robust techniques. For all approaches, report results with and without the implementation of a given approach to maximize transparency and replicability.
7. Researchers will include a separate section offering an honest and realistic assessment of the limitations of the research together with consequences of such limitations for the interpretation of results and implications in all manuscripts.	<ul style="list-style-type: none"> • Revise journal editorial policies to include a requirement that limitations be described clearly and realistically in a separate section in all submitted manuscripts. • Revise journal reviewer evaluation forms to include a separate set of questions asking reviewers about limitations and the extent to which they affect the study in terms of substantive conclusions. • Focus on limitations that really matter in terms of the degree of confidence placed on results and conclusions and not on generic statements such as 'our study was based on one sample only – future research should confirm the degree to which our results generalize to other samples'. • Accompany limitations with a clear explanation of their consequences in terms of theory and practice. • Accompany limitations with information necessary to make them informative for a sceptical scientific audience rather than justifications or statements aimed at discounting them as minor issues that do not threaten the interpretation of results and implications.

management researchers should be held to these standards. Nonetheless, we assert that progress in management research depends on applying the best methods available to us, and we should regularly review, reconsider, and strive to improve the methodological tools we use in our field. Although these assumptions might not seem particularly controversial, it is important to make them explicit, as they form the foundation for the wishes we present in this article.

WISH 1: ACCELERATING THEORETICAL PROGRESS

As with most social sciences, management research follows a cycle in which theories generate directional hypotheses, such as whether an effect will be positive or negative, and these hypotheses are empirically evaluated with null hypothesis significance testing, which determines whether an effect is likely to differ from zero. Within this cycle, empirical results are rarely drawn upon to calibrate theoretical hypotheses to predict the magnitude of an effect, and in like fashion, the directional hypotheses set forth by theories provide little reason to conduct statistical tests that detect anything other than differences from zero (Gigerenzer, 1998). Research that follows this cycle is caught in a trap that yields little theoretical progress, because theories are stated with such imprecision that they are difficult to refute (Edwards and Berry, 2010), and empirical research does not home in on non-zero effects that are conceptually and practically meaningful (Kirk, 1996). As a result, the progress of theories in management tends to follow a pattern described by Meehl (1978, p. 807) as such:

There is a period of enthusiasm about a new theory, a period of attempted application to several fact domains, a period of disillusionment as the negative data come in, a growing bafflement about inconsistent and unreplicable empirical results, multiple resort to ad hoc excuses, and then finally people just sort of lose interest in the thing and pursue other endeavors.

To break from this cycle, our wish for the next decade is that management researchers accelerate theoretical progress by developing theories with increased precision and exposing them to stringent tests that put them at genuine risk of falsification. Theoretical precision can be enhanced by moving beyond the directional predictions that dominate management theories (Edwards and Berry, 2010). For instance, researchers can specify lower and upper bounds within which a theoretical parameter should fall. Lower bounds can refer to a value below which an effect is considered trivial. To illustrate, correlations between variables that are unrelated from a conceptual standpoint tend to hover around 0.20 to 0.30 (Aguinis et al., 2011b; Lykken, 1968; Meehl, 1990). Small effects such as these could be used to construct a 'zone of indifference' around zero that comprises values that would be considered essentially equivalent to the null from a theoretical and practical standpoint (Binder, 1963). Upper bounds can refer to the maximum value a parameter could attain on conceptual grounds, such as a correlation large enough to question whether two constructs are conceptually distinct (Bagozzi et al., 1991). Alternately, upper limits can be derived from formulas that yield the largest value a parameter can obtain in light of other parameters in a theoretical model (McNemar, 1962).

Theoretical precision can also be increased by stating predictions as comparisons, such that one effect will be larger or smaller than another (Edwards and Berry, 2010). Comparative predictions are consistent with the principle of strong inference (Platt, 1964) in which competing hypotheses are derived and empirical tests are conducted such that evidence supporting one hypothesis concurrently refutes one or more alternative hypotheses. The merits of this approach were underscored by Hitt et al. (1998a), who noted that testing competing predictions can strengthen the logical conclusions yielded by empirical research. Other ways to refine theoretical predictions include identifying moderators that influence the strength and direction of relationships between constructs (Edwards and Berry, 2010) and explicitly stating whether relationships are expected to be linear versus curvilinear (Ferris et al., 2006; Pierce and Aguinis, 2013).

Perhaps the most effective way to increase theoretical precision is to derive non-nil predictions, such that theories predict the presence of a non-zero effect rather than the mere absence of a zero effect (Edwards and Berry, 2010; Meehl, 1967, 1990). Although non-nil predictions can be found in the hard sciences, such as physics and chemistry, they are rare in social sciences such as management. Predictions stated as point estimates can be difficult to justify in management research, given the uncertainties inherent in the phenomena under investigation. However, it can be feasible to state predictions as ranges of values, akin to the 'good-enough' belt advocated by Serlin and Lapsley (1985). Values used to derive non-nil predictions can be based on formal theorizing (Adner et al., 2009) or findings from previous empirical research, such as individual studies deemed relevant to the theory at hand (Mulaik et al., 1997) or meta-analyses that summarize effect sizes across multiple studies (Aguinis et al., 2011b; Schmidt, 1992).

In addition to developing theories with greater precision, theoretical progress can be accelerated by adopting alternatives to null hypothesis significance testing. The shortcomings of null hypothesis significance testing have been lamented for years (Bakan, 1966; Carver, 1978; Cohen, 1994; Krueger, 2001; Nickerson, 2000; Rozeboom, 1960), yet it retains a firm foothold in management research (Schwab and Starbuck, 2009). The persistence of null hypothesis significance testing can be attributed to several factors, such as statistical training in graduate programmes, tolerance of null hypothesis significance tests in the review process, and the preponderance of directional hypotheses that require nothing more than testing whether an effect differs from zero in the expected direction (Edwards and Berry, 2010; Orlitzky, 2012; Schmidt, 1996).

Various alternatives to null hypothesis significance testing have been proposed. Perhaps the most common recommendation is to report effect sizes along with confidence intervals that document the uncertainty surrounding parameter estimates (Cohen, 1994; Nickerson, 2000; Schmidt, 1996). Although this approach provides more information than null hypothesis significance tests, studies that report confidence intervals often focus on whether they exclude zero, which is logically equivalent to null hypothesis significance testing itself (Cortina and Dunlap, 1997). The merits of effect sizes and confidence intervals are better realized when they are used to evaluate hypotheses that predict non-nil effects, which in turn can be derived using principles for increasing theoretical precision (Edwards and Berry, 2010). When non-nil predictions are stated as range hypotheses, confidence intervals can be modified accordingly (Nickerson, 2000; Serlin, 1993; Serlin and Lapsley, 1985; Tryon, 2001).

Another alternative to null hypothesis significance testing is Bayesian analysis, in which the researcher specifies the prior distribution of an effect before conducting a study and uses Bayes' theorem to derive a posterior distribution, which adjusts the prior distribution using the obtained data (Howard et al., 2000; Kruschke et al., 2012; Pruzek, 1997). Bayesian analysis can help accelerate theoretical progress because it invites researchers to draw from available theory and evidence when formulating hypotheses and use results to update existing knowledge, thereby creating a mutually reinforcing cycle that connects theory with empiricism. Although Bayesian analysis is rare in management research (Kruschke et al., 2012), its use is growing in the social sciences, facilitated by accessible tutorials and the growing availability of software to conduct Bayesian analysis (Kruschke, 2010; Kruschke et al., 2012; Lee, 2004; Ntzoufras, 2009; Sivia and Skilling, 2006).

To summarize, theoretical progress in management research can be accelerated using a two-pronged approach. First, we encourage researchers to develop theories with greater precision, such that they yield predictions that are more specific and refined than merely stating that an effect will differ from zero in a positive or negative direction (Edwards and Berry, 2010). Second, we urge researchers to adopt alternatives to null hypothesis significance testing, such as reporting effect sizes along with confidence intervals and conducting Bayesian analysis, which incorporates prior information that is refined with each successive empirical study. These recommendations go hand-in-hand, because the non-nil predictions yielded by precise theories motivate alternatives to null hypotheses significance testing, and vice versa (Gigerenzer, 1998).

WISH 2: IMPROVING THE CONSTRUCT VALIDITY OF MEASURES

Construct validity is a perennial concern in management research. In essence, construct validity refers to the extent to which a measure represents the intended construct (Edwards, 2003; Schwab, 1980). Methodological reviews of management research have repeatedly emphasized the importance of construct validity and the need to improve the measurement of theoretical constructs (Aguinis and Vandenberg, in press). For instance, Venkatraman and Grant (1986) reviewed measurement practices in strategic management research and identified several key shortcomings, including the use of categorical measures to represent complex constructs, relying on single indicators, inadequate assessment of reliability, and insufficient attention to levels of measurement. These issues were re-examined by Boyd et al. (2005), who found that most studies in strategic management research continued to rely on single indicators, and those studies that used multiple indicators typically failed to report reliability estimates. These observations led Boyd et al. (2005) to conclude that 'poor construct measurement is one of the most serious threats to strategic management research' (p. 240). These concerns were revisited by Boyd et al. (2013), who found that many of the measurement shortcomings identified by Venkatraman and Grant (1986) and revisited by Boyd et al. (2005) have not been resolved.

To address these concerns, our wish for the next decade is that management researchers actively strive to improve the construct validity of measures used in empirical research. To this end, we offer several recommendations. First, at the conceptual level,

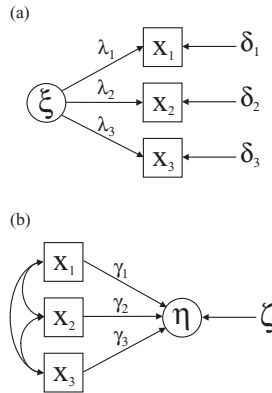


Figure 1. (a) reflective measurement model; (b) formative measurement model

it is essential to establish that measures meaningfully correspond to the intended constructs. This correspondence can be strengthened by clearly defining the construct of interest and selecting measures whose meaning unequivocally matches the definition of the construct. Although this principle might seem obvious, there is no shortage of studies in which the correspondence between constructs and measures is tenuous. For example, Ketchen et al. (2013) reviewed studies that used the same archival measures to represent substantively different constructs. As Ketchen et al. (2013, p. 37) noted, ‘even if a proxy appears to be reasonable *within the context of an individual study*, its use *among different studies* to represent different constructs creates serious problems’ (emphasis in original). The correspondence between constructs and measures can be evaluated by providing definitions of the constructs under investigation to informed judges and asking them to rate the extent to which each measure represents each construct. Good correspondence is evidenced by measures with high mean ratings on the intended construct, low mean ratings on other constructs, and low variance in ratings (cf. Hinkin and Tracey, 1999). This pattern would indicate that the judges agree that the measure represents only the intended construct. For an illustration of this approach, see Edwards and O’Neill (1998).

Second, the nature and direction of the relationships between constructs and measures should be specified explicitly (Edwards and Bagozzi, 2000). Typically, constructs are treated as direct causes of their measures, consistent with the reflective measurement model in Figure 1a. This model underlies common factor analysis (Mulaik, 2009a) and internal consistency reliability estimation (Cronbach, 1951) and captures the premise that variation in a construct produces variation in its measures (Borsboom et al., 2004; Edwards and Bagozzi, 2000). An alternative specification is shown in Figure 1b, which depicts a formative measurement model (Blalock, 1964; MacCallum and Browne, 1993; Podsakoff et al., 2006). Although formative measurement has received increased attention in management research, it rests on several dubious assumptions, perhaps the foremost of which is that measures have the capacity to cause constructs (Edwards, 2011). Stripped to their essence, measures are scores taken as empirical analogues of constructs. These scores have no causal potency in their own right, but instead are visible traces of the unobserved phenomena to which constructs refer (Borsboom et al., 2004; Edwards

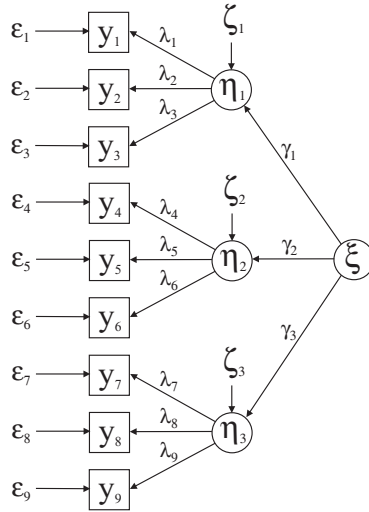


Figure 2. Superordinate multidimensional construct model

and Bagozzi, 2000; Messick, 1995). For these reasons, we believe it makes little sense to specify models in which measures cause constructs (Edwards, 2011; Howell et al., 2007).

Despite our misgivings about formative measurement, we understand their intuitive appeal. For instance, formative measurement models might seem to provide a succinct and convenient way to combine measures of different concepts into an overall composite. Indeed, guidelines for creating formative measures recommend selecting indicators that are conceptually distinct from one another (Diamantopoulos and Siguaw, 2006; Podsakoff et al., 2006). However, when measures represent distinct dimensions, the dimensions themselves should be conceived as constructs in their own right, and the relationship between each measure and its associated dimension should be carefully considered. In most cases, we believe this process will yield a multidimensional construct model in which the measures are reflective indicators of their dimensions and the dimensions are reflective or formative indicators of a broader construct. A model that treats the dimensions as reflective relative to a broader construct is illustrated in Figure 2, which shows a superordinate multidimensional construct model (Edwards, 2001). This model is consistent with conventional second-order factor models, which have a long history in the factor analysis literature (Cattell, 1978; Gorsuch, 1983; Rindskopf and Rose, 1988). A model that specifies the dimensions as formative relative to a broader construct is depicted in Figure 3, which illustrates an aggregate multidimensional construct model (Edwards, 2001). The choice between these two models is a conceptual matter that depends on whether the broader construct is seen as producing its dimensions, as in Figure 2, or instead is a result of its dimensions, as in Figure 3. Further details regarding the distinctions between these models, along with methods for analysing and interpreting the models, are provided by Edwards (2001).

Third, researchers should thoroughly evaluate measurement models using confirmatory factor analysis. Although a comprehensive treatment of confirmatory factor analysis is well beyond the scope of this article, we will emphasize several key points (for

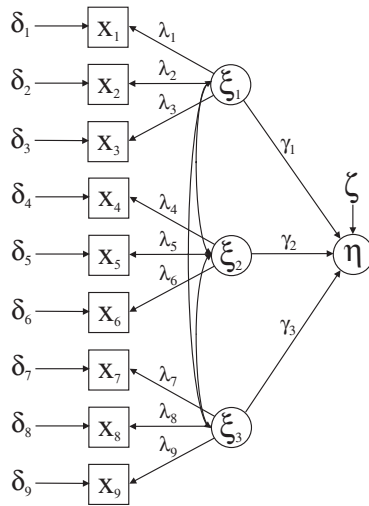


Figure 3. Aggregate multidimensional construct model

additional information, see Bollen, 1989; Brown, 2006; Kline, 2010; Schumacker and Lomax, 2004). When conducting a confirmatory factor analysis, an initial step is to determine whether the data to be analysed satisfy the distributional assumptions of the estimation procedures involved (e.g., whether the data follow a multivariate normal distribution). When these assumptions are violated, alternative methods of estimation can be used (Browne, 1984), or remedial procedures can be applied, such as methods to offset the effects of non-normality (Satorra and Bentler, 1994). Next, the measurement model should be specified based on conceptual reasoning regarding the relationships between the constructs and their measures (Edwards and Bagozzi, 2000). As part of specifying the model, it is essential to ensure that the measurement and structural portions of the model are identified (Bollen, 1989). After the model is estimated, the resulting parameters should be examined to ensure they are consistent with the model and fall within logical bounds, such as item loadings that differ appreciably from zero and factor correlations that are significantly and substantively greater than -1.0 and less than 1.0 (Bagozzi et al., 1991). In addition, sources of misfit should be examined, such as standardized residuals and modification indices that refer to constrained measurement errors and item cross-loadings. These results should not be used to re-specify the model (MacCallum et al., 1992), but instead to reveal where the sources of misfit are most pronounced. Finally, the target measurement model should be compared to alternative models that are theoretically meaningful and justifiable. We do not recommend the routine comparison of the target model to models that collapse factors, which has become commonplace in management research, because tests that compare these models merely indicate whether the correlations among the factors in the target model are less than unity. This condition can be assessed by simply determining whether the confidence intervals around the correlations exclude unity. For a good discussion of best practices in evaluating and reporting results from confirmatory factor analysis models, see Jackson et al. (2009).

In sum, construct measurement is one of the most pressing concerns in management research (Aguinis and Vandenberg, in press; Boyd et al., 2005), and we have offered recommendations that should help management researchers improve the construct validity of their measures. These recommendations emphasize the importance of clearly defining constructs, ensuring that constructs are linked to their measures from a conceptual standpoint, carefully specifying the nature and direction of the relationships between constructs and measures, and thoroughly evaluating measurement models using confirmatory factor analysis. By following these guidelines, we believe that the construct validity of measurement in management research can be improved substantially.

WISH 3: STRENGTHENING CAUSAL INFERENCES

Management theories are inherently causal, such that they describe conditions and events believed to affect outcomes of interest and explain mechanisms by which these effects occur (Whetten, 1989). In order to test causal claims, empirical studies must satisfy conditions that justify causal inferences. In the philosophy of science, conditions for causality have been debated for centuries (for reviews, see Illari et al., 2011; Mulaik, 2009b; Shadish et al., 2002). Within the social sciences, such as management, psychology, and sociology, conditions for causation usually focus on three criteria (Shadish et al., 2002): (a) the cause and effect must be associated; (b) the cause must precede the effect in time; and (c) alternative explanations for the presumed causal effect can be ruled out.

In management research, the first of the three conditions for causality is most often satisfied, as this condition merely requires that the relationship between the presumed cause and effect differs from zero in the expected direction (Aguinis and Vandenberg, in press). The second condition is satisfied less frequently because many studies rely on cross-sectional designs that do not establish temporal precedence of a hypothesized cause relative to its effect (Bowen and Wiersema, 1999). This shortcoming can be overcome by studies that use longitudinal designs in which units are observed on multiple occasions (Ployhart and Vandenberg, 2010). The third condition is rarely satisfied and presents the most significant challenges for establishing causality, due to the multitude of alternative explanations that can often account for the relationship between a presumed cause and effect.

Our wish for the next decade is that management researchers design studies and implement procedures with the aim of strengthening inferences and ruling out alternative explanations regarding causal relationships. One effective approach to rule out alternative causal explanations is the counterfactual model of causation developed by Rubin and colleagues (Holland, 1986; Little and Rubin, 2000; Rubin, 1974, 1978). This model defines a causal effect as the difference in an outcome variable if the same unit of observation had experienced different levels of a causal variable at the same point in time. This view of causation is consistent with thought experiments researchers conduct by imagining different potential outcomes across levels of a causal variable (Little and Rubin, 2000), such as how the performance of a firm would vary if it implemented different strategies. Thinking about causation in this manner is counterfactual because the same unit of observation cannot simultaneously exist at different levels of a causal variable (e.g., a single firm cannot implement different strategies at the same time). This

approach to causation has become prevalent in social sciences that include psychology and education (Collins and Winship, 2007; Pearl, 2009; Shadish et al., 2002) and is gaining attention in management research (Aguinis and Vandenberg, in press; Antonakis et al., 2010; Edwards, 2008).

To resolve the dilemma posed by the counterfactual model of causation, researchers can attempt to identify units of observation that are identical in all possible respects except their standing on the predictor variable. This goal can be achieved by random assignment, as when units of observation are randomly placed in different experimental conditions. Unfortunately, random assignment is often impractical in management research, particularly when studies are conducted in organizational settings or involve units of observation at higher levels of analysis than the individual, such as work groups, departments, or firms (Bascle, 2008; Grant and Wall, 2009; Hitt et al., 1998a). In such cases, other approaches can be used that attempt to approximate the effects of random assignment.

One approach is matching, in which groups are formed that contain units with similar scores on one or more matching variables that represent alternative explanations for the causal effect of interest (Echambadi et al., 2006; Rosenbaum and Rubin, 1985; Shadish et al., 2002). The resulting groups are similar on the matching variables but differ on the causal variable, which rules out the phenomena represented by the matching variables as alternative explanations of the causal effect. Although matching has a long history in non-experimental research (Cochran, 1968), it presents several difficulties, such as the loss of information when continuous matching variables are categorized to form groups, bias introduced when matching variables are measured with error, and difficulties forming matched groups when the number of matching variables is large (Shadish et al., 2002). To address the difficulties of matching on numerous variables, the variables can be combined into a multivariate composite, which is then used to create groups. This principle underlies propensity scores (Li, 2013; Rosenbaum and Rubin, 1985), which are constructed by using logistic regression to predict group membership based on the matching variables.

Another approach is to statistically control for variables that are believed to represent alternative explanations for the causal effect of interest (Atinc et al., 2011; Becker, 2005; Carlson and Wu, 2012; Rubin, 1974; Spector and Brannick, 2011). This approach is based on the premise that non-random sampling can be viewed as a type of omitted variables problem (Heckman, 1979; Hitt et al., 2004; Li, 2013). In particular, when units are not randomly assigned to levels of a causal variable, they can differ for reasons other than their standing on the causal variable. If these differences correlate with the causal variable and also influence the dependent variable, the coefficient relating the causal variable to the dependent variable will be biased.

This situation is illustrated by Figure 4. In Figure 4a, X is specified as the sole cause of Y , whereas in Figure 4b, X and W are portrayed as two correlated causes of Y (for simplicity and without loss of generality, we assume X , W , and Y are standardized). If the correct model is that shown in Figure 4b but the model in Figure 4a is estimated, such that W is omitted, the effect of X on Y will be biased by an amount equal to the product of the correlation between W and X times the path from W to Y (i.e., $b' = b + ac$). This bias is removed by controlling for W , as in Figure 4b. It should be noted that the

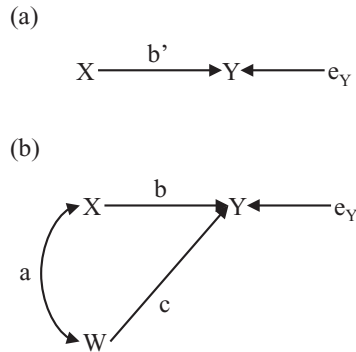


Figure 4. Models relating X to Y with and without the correlated cause W: (a) X as a cause of Y; (b) X and W as correlated causes of Y

effectiveness of this approach depends on the availability of control variables that serve as proxies for the omitted causes of Y that are correlated with X. Moreover, measurement error in W can bias the path from X to Y, which can counteract the benefits of using W to remove bias due to the omitted variables problem. Furthermore, the choice of control variables should not be taken lightly, because adding control variables effectively alters the causal model under investigation. Thus, the role of control variables in a model should be subject to the same conceptual scrutiny as that accorded to the focal causal variables (Aguinis and Vandenberg, in press; Carlson and Wu, 2012; Edwards, 2008).

A third approach to ruling out alternative causal explanations involves the use of instrumental variables (Angrist et al., 1996; Bollen, 2012; Martens et al., 2006; Semadeni et al., in press). Instrumental variable estimation has been recommended as a method for addressing endogeneity, which exists when one or more predictors in an equation are correlated with the residual term (Antonakis et al., 2010; Bascle, 2008; Hamilton and Nickerson, 2003). Endogeneity can arise from several sources, one of which is the aforementioned omitted variables problem that occurs when predictor variables are correlated with other causes of the dependent variable, which are effectively collapsed into the residual term. This situation is depicted in Figure 5a, where the causal variable X is correlated with the residual e_Y , which includes the causes of Y not explicitly included in the model (as before, the prime on b means this coefficient is biased). Endogeneity can be addressed by finding an instrumental variable for X, which satisfies three conditions: (a) the instrumental variable must be correlated with X; (b) the instrumental variable must be uncorrelated with e_Y ; and (c) the instrumental variable is correlated with Y only through its relationship with X. When an instrumental variable is located, it can be used in a two-stage least squares analysis in which X is regressed on the instrumental variable, the predicted values of X are saved, and these values are then used to predict Y. This approach is depicted in Figure 5b, which shows the instrumental variable I as a cause of the predicted value of X (i.e., \hat{X}), which in turn is a predictor of Y. Although instrumental variable estimation provides a viable solution to the endogeneity problem, the conditions that instrumental vari-

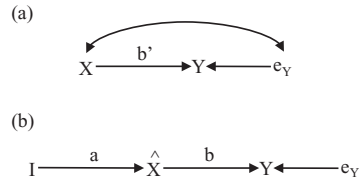


Figure 5. Models depicting X as endogenous and adding the instrumental variable I to address endogeneity: (a) X as a cause of Y that is correlated with e_Y ; (b) using I as an instrumental variable for X in predicting Y

ables must satisfy are stringent, and failure to satisfy the conditions can produce estimates that are biased, inconsistent, and inefficient (Bollen, 2012; Martens et al., 2006).

In summary, although randomized experiments are often impractical or unfeasible in management research, methods are available that can help researchers draw stronger causal inferences from non-experimental research. We have highlighted the fundamental dilemma that arises from non-experimental research using the counterfactual model of causation and suggested various approaches by which this dilemma can be addressed. By implementing these guidelines, our wish is that management researchers can draw stronger causal inferences from empirical research, thereby improving our ability to test theory and provide useful recommendations for practice.

WISH 4: INCORPORATING MULTILEVEL DESIGN, MEASUREMENT, AND ANALYSIS

Regardless of whether a researcher is interested in studying individuals within teams, teams within firms, firms within industries, or industries within countries, the nature of data structures in the field of management is such that entities are usually embedded within larger collectives. As a result, data structures usually contain built-in dependence because entities within the same collective are exposed to similar processes, which leads to greater similarity regarding outcome variables, compared to entities across collectives. In other words, there is co-variation between higher- and lower-level variables (Aguinis et al., 2013b).

Although in some cases such dependence is acknowledged because of formal structures such as an organizational hierarchy that includes individuals in different departments or known classifications such as firms belonging in different industries, dependence often occurs without entities being formally clustered into larger units. As noted by Kenny and Judd (1996, p. 138), ‘observations may be dependent, for instance, because they share some common feature, come from some common source, are affected by social interaction, or are arranged spatially or sequentially in time’. Stated differently, the resulting data structure includes dependence of observations due to shared experiences such as firms facing similar market-based challenges or being members of the same group of firms (e.g., Fortune 500) even if they are in different industries. In fact, Mathieu et al. (2012) reported that such dependence is quite pervasive and yet remains unacknowledged in most published articles.

There are two important and negative consequences when dependence is not properly acknowledged and modelled. First, co-variation of variables across levels leads to gross errors of prediction if a researcher uses popular statistical approaches such as ordinary least squares (OLS) regression, which are not designed to model data structures that include dependence due to clustering of entities (Hox, 2010; Snijders and Bosker, 2012). Specifically, when using techniques that rely on the independence assumption, as is the case for OLS regression, the resulting standard errors will be downwardly biased, resulting in Type I statistical errors (i.e., false positives). In addition, it is possible that ignoring dependence may also lead to increased Type II error rates (i.e., insufficient statistical power to detect existing population effects). However, such decreases in statistical power take place when researchers test models including variables at the lower level of analysis only (Bliese and Hanges, 2004).

The second negative consequence, which perhaps is even more important in terms of theoretical progress, is that when dependence is not acknowledged explicitly, researchers miss an opportunity to examine cross-level direct and interaction effects. Specifically, in addition to examining same-level effects such as the relationship between a firm's top management team diversity and firm performance, acknowledging the various levels present in the data structure allows for tests of cross-level direct hypotheses such as the direct impact of team cohesion (i.e., a higher-level construct that varies across firms) on firm performance (e.g., Wei and Wu, 2013). In addition, modelling different levels also allows for tests of cross-level interaction hypotheses such as the moderating effect of a particular industry characteristic on the relationship between two firm-level variables (Aguinis et al., 2013b). Because of their higher degree of specificity and precision, cross-level hypotheses have great potential in terms of producing important theoretical advancements (Edwards and Berry, 2010). Figure 6a includes a generic multilevel model and hypotheses addressing a same-level direct effect between Level 1 variables (Hypothesis 1), a cross-level direct effect of a Level 2 on a Level 1 variable (Hypothesis 2), and a cross-level interaction effect between two variables at different levels on a Level 1 outcome (Hypothesis 3).

Our wish for the next decade is that researchers will design empirical studies, measure constructs, and analyse the resulting data adopting multilevel approaches. In other words, multilevel considerations need to be taken into account regarding conceptual, measurement, and analytic issues. Conceptually, it is necessary to establish a priori hypotheses regarding the specific nature of the cross-level effects. From a measurement perspective, it will be necessary to use appropriate tools that capture constructs at different levels of analysis and also anticipate the relative appropriateness of aggregating scores to higher levels. Finally, from an analytic perspective, it will be necessary to use multilevel analytic approaches, which can be implemented using software programs such as HLM, MPlus, SAS, and R.

We anticipate many benefits of the broader adoption of multilevel design, measurement, and analysis in the field of management. Specifically, researchers in organizational behaviour, human resource management, strategy, and organizational theory now recognize that integrating micro and macro levels of analysis is one of the biggest challenges in the field of management (Aguinis et al., 2011a; Buckley et al., 2011). An additional indicator of the need to bridge micro and macro domains is that the Strategic Manage-

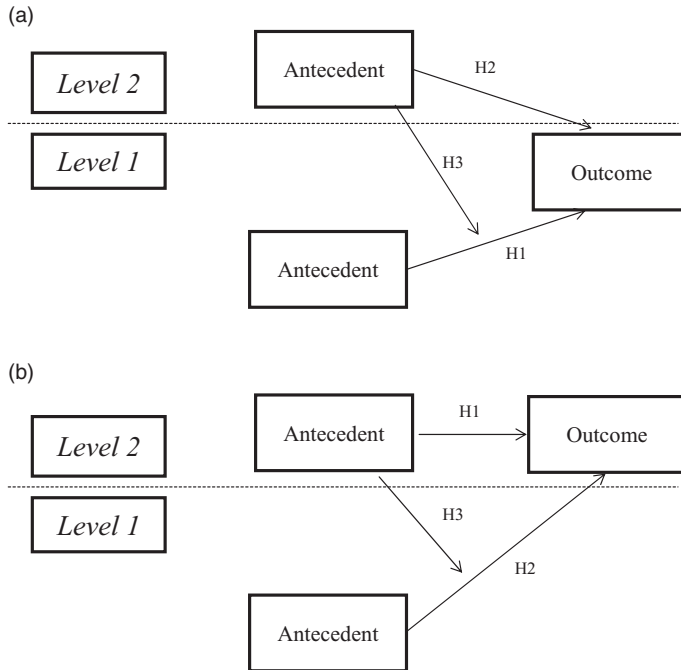


Figure 6. (a) Graphic representation of generic multilevel model including same-level direct (Hypothesis 1), cross-level direct (Hypothesis 2), and cross-level interaction (Hypothesis 3) hypotheses and an outcome residing at the lower level of analysis. (b) Graphic representation of generic multilevel model including same-level direct (Hypothesis 1), cross-level direct (Hypothesis 2), and cross-level interaction (Hypothesis 3) hypotheses and an outcome residing at the higher level of analysis

ment Society (SMS) organized a special conference in Copenhagen in 2014 addressing microfoundations for strategic management research, and the *Journal of Management Studies* published a special issue in 2012 addressing microfoundations of routines and capabilities (Felin et al., 2012). The call for proposals for the SMS conference stated that ‘Microfoundations have become increasingly influential in strategic management research. The basic argument is that individuals and their interaction matter fundamentally to strategic issues.’ As recently noted by Devinney (2013), ‘microfoundations can be the key platform in moving the management field forward. It opens up the possibilities of bridging the macro–micro divide that pervades management research by serving as a conceptual forum to debate whether or not it is possible for us to come up with a more unified and parsimonious characterization of our field’ (p. 84). The task of integrating micro and macro domains by, for example, examining microfoundations for strategy, is inherently multilevel in nature (Felin and Foss, 2006). Accordingly, multilevel perspectives will be necessary to gain insights regarding this important challenge.

Once again, let us consider the generic multilevel model included in Figure 6a. This multilevel conceptual framework could be used to assess the relationship between human resource management practices (Level 2 antecedent), employee organizational commitment (Level 1 antecedent), and employee performance (Level 1 outcome) (e.g., Bal et al., in press). Specifically, one can test not only the direct relationship between organizational

commitment and performance, but also the extent to which human resource management practices moderates that relationship. In other words, a multilevel conceptual model such as this one can be used to test contingency theories involving more than one level of analysis.

Finally, although our illustration thus far referred to a Level 1 outcome, it is certainly possible to test multilevel models including Level 2 outcomes as well (e.g., Croon and van Veldhoven, 2007). Figure 6b includes a generic multilevel model with an outcome at the higher level of analysis. In this model, the outcome could be firm performance, the Level 1 predictor could be individual characteristics of members of the top management team (e.g., personality, experience), and the Level 2 predictor could be the firm's executive compensation system. In other words, a model such as the one included in Figure 6b would allow for tests of a direct relationship between two Level 2 variables as well as the potential moderating effect of a Level 1 predictor on the relationship between a Level 2 predictor and a Level 2 outcome. These types of cross-level relationships involving outcome variables at the higher level of analysis are critical for several theories in strategic management studies including all of those addressing upper echelons (e.g., Carpenter and Reilly, 2006).

In sum, although not acknowledged explicitly, non-independence of observations, be it at the individual, group, or firm levels, is pervasive in management research. However, such non-independence is often not acknowledged explicitly. Not acknowledging dependence explicitly leads to errors in predictions due to a violation of statistical assumptions and, perhaps more importantly, to opportunity cost in terms of theoretical progress. In other words, not adopting a multilevel lens precludes researchers from hypothesizing and testing specific and precise relationship involving variables at different levels of analysis. In contrast, explicitly acknowledging and modelling multilevel relationships has great potential in terms of making meaningful theoretical progress regarding the need to bridge micro and macro research domains.

WISH 5: BALANCING TRADE-OFFS BETWEEN INTERNAL AND EXTERNAL VALIDITY

As discussed earlier in the section about strengthening causal inferences (wish 1), internal validity refers to the extent to which we can make conclusions regarding whether one variable causes another one. In other words, how confident are we that variable X indeed causes outcome Y? External validity refers to the extent to which we can make conclusions regarding whether a particular effect or relationship generalizes across persons, settings, and time (Cook and Campbell, 1979). In other words, how confident are we about the generalizability of a particular relationship beyond, for example, a particular sample, research setting, and set of measures?

Management researchers often face a difficult dilemma. First, research designs that allow for researcher control, random assignment, and the manipulation of variables yield high levels of confidence regarding internal validity but are usually weaker regarding external validity (e.g., due to the use of sophomores in university laboratory settings). Alternatively, research designs that involve collecting data in naturalistic (i.e., field) settings usually afford greater external validity, but less confidence regarding internal

validity (e.g., due to the use of data drawn from larger databases including information on publicly traded firms). It seems that researchers are put in the uncomfortable position of designing either causally strong studies that are dubious in terms of their ability of being replicated in other settings, or studies based on large and *presumably* random and generalizable samples for which there is only weak and indirect evidence regarding causal relationships (Highhouse, 2009).

In the paragraph above, we highlighted the word *presumably* because, although it is often assumed that using a large sample drawn from a database provides assurances of generalizability, this is not necessarily the case. Reassuringly, the median sample size of articles published in *Strategic Management Journal (SMJ)* was $N = 207$ in the 1980–82 time period and $N = 142$ during 1990–92, but increased substantially to $N = 1282$ in the 2000–02 window (Ketchen et al., 2008). However, Short et al. (2002) content-analysed 437 articles published in *Academy of Management Journal (AMJ)*, *Administrative Science Quarterly (ASQ)*, *Journal of Management (JOM)*, *Organization Science*, and *SMJ* between 1980 and 1999 and found that that fewer than one in five studies used a random sample. In other words, the sampling strategy is typically based on data availability – what is called ‘convenience sampling’, raising obvious questions regarding the generalizability of results (Dipboye and Flanagan, 1979). As an illustration, even though samples may be large, using firms drawn from COMPUSTAT versus TRINET usually leads to different substantive conclusions (Davis and Duhaime, 1992).

Our wish for the next decade is that researchers will design studies that maximize both internal and external validity, and not one at the expense of the other. What can researchers do to balance trade-offs between internal and external validity? The solution would have to involve experimental research design features (e.g., temporal sequencing, ability to manipulate variables), as well as naturalistic research design features (i.e., collecting data in naturally occurring environments as opposed to artificial settings). Technological advancements have allowed researchers in fields such as physics to implement novel research designs that resulted in important theoretical advancements. Notable exemplars are the Large Hadron Collider and the Hubble telescope. Similarly, technological advancements now allow management researchers to implement innovative research designs that allow for striking a good balance in terms of internal and external validity.

Experience sampling methodology, eLancing, and virtual reality are examples of methodological tools that have great potential to help the field move forward. These are mostly applicable to research addressing individual-level issues. However, these tools can be used to make contributions to all theories in strategic management studies that address a combination of micro and macro topics including upper echelons, micro-foundations, and strategy-as-practice (Floyd and Sputtek, 2011). As we noted in the section regarding multi-level issues (wish 5), bridging the micro–macro gap is an important challenge in the field of management. So, the use of methods that allow for a balance of internal and external validity trade-offs in the particular context of bridging the micro–macro gap has great potential in terms of advancing the field. The novel methodological tools we discuss in this section can be used to study individual-level phenomena that have important implications for many research streams in the field of strategy (Floyd and Sputtek, 2011). For example, these methods can be used to under-

stand micro-foundations of corporate social responsibility – individual perceptions, attitudes, and behaviours that underlie the success or failure of corporate social responsibility policies and practices (e.g., Aguinis and Glavas, 2013; Morgeson et al., in press). They can also be used to understand micro-foundations of human capital – individual behaviours that serve as an intermediate mechanism between human resource management policies and practices and firm performance (e.g., Foss, 2011). Similarly, they can be used for examining upper echelons questions such as why and how top executives behave the way they do and the consequences of such behaviour for their firms. Finally, they can also be used to conduct empirical research on strategy as something that people do in organizations rather than something organizations have – what has been called the ‘strategy-as-practice’ perspective (e.g., Whittington, 2006).

Experience Sampling Methodology

Experience Sampling Methodology (ESM), also labelled ecological momentary assessment, allows researchers to gather detailed accounts of people’s daily experiences over time and capture these experiences as they occur in a person’s natural environment (Ilies and Judge, 2002; Uy et al., 2010). As such, ESM allows for a longitudinal examination of the nature and causal directions among the constructs investigated because temporal sequencing is a necessary condition for inferring causal effects (Beal and Weiss, 2003; Bergh and Holbein, 1997). Regarding external validity, ESM studies are conducted in a participant’s natural setting and as they go about their daily activities thereby capturing participants’ attitudes, emotions, and behaviours, as they occur naturally and not in artificial laboratory settings.

One of the pioneer studies using experience sampling methodology involved participants who carried notebooks and a blood pressure reader around their arms (Marco and Suls, 1993). In fact, Herman Aguinis happened to be one of the participants in the Marco and Suls study when he was a graduate student and, as part of this study, he was required to answer questions at random intervals as prompted by a portable beeper. At present, participants are no longer required to carry notebooks with them and ESM studies investigating entrepreneurs can be conducted using a mobile phone interface (e.g., Foo et al., 2009).

eLancing

eLancing, or internet freelancing, is a type of work arrangement that takes place through an eLancing ‘marketplace’, which is a website where individuals interested in being hired and clients looking for individuals to perform some type of work meet. Popular eLancing marketplaces include eLance.com, freelancer.com, guru.com, oDesk.com, and microworkers.com and they involve millions of people around the world (Aguinis and Lawal, 2013). Numerous tasks can be contracted out through eLancing; for example, programmers are available through freelancer.com to develop applications for websites and mobile phones. Aguinis and Lawal (2012) illustrated how to use the eLancing environment to conduct field experimental research that maximizes both internal and external validity. Specifically, a researcher issues a call for work (i.e., experiment), then

manipulates variables (e.g., change the nature of the task, composition of teams, amount and type of information shared with individuals and/or teams), and then measures the dependent variables (e.g., team performance, individual job satisfaction with the team). So, eLancing allows researchers to use random assignment within the context of a natural (i.e., eLancing) work environment. For example, Aguinis and Lawal (2012) described the use of eLancing to conduct a study involving teams. The main research question was whether a team's collective efficacy towards entrepreneurial behaviours was higher for teams with more entrepreneurial experience compared to teams with less experience.

Virtual Reality

Virtual reality (VR), often referred to as virtual environment, is a computer-generated, multi-sensory environment in which a user experiences telepresence. Telepresence, or full-immersion VR, means that users lose awareness of being present at the site of the human–computer interface, and instead feel fully immersed in the VR environment experience (Stanney, 2002). Once immersed in a VR environment, users can perform surgery, walk on the surface of Mars, fly an aircraft, or stroll through the Sistine Chapel without really being there (Aguinis et al., 2001).

At present, VR is most often used for training (e.g., pilots and surgeons), architectural design (e.g., identifying design problems prior to actual construction), and entertainment (e.g., virtual worlds) purposes. Pierce and Aguinis (1997) described how virtual reality technology can be used for designing experiments in which one can draw inferences about causality due to random assignment of participants to conditions, and also external validity inferences due to participants' immersion in natural (albeit virtual) environments. Moreover, VR can also be used to study sensitive topics such as workplace romance and unethical work behaviours. For example, a VR user could be immersed in a simulated organization and observe supervisor–subordinate and peer romances within the same work group (i.e., type of romance is the manipulated variable), and then we can measure study participants' affective and attitudinal reactions to each type of romance (i.e., outcome variables). VR also allows for the manipulation of other variables such as a CEO's motives for engaging in unethical behaviour (e.g., desire to receive personal gains versus protecting her company's reputation). Obviously, these are variables that are not easy to study experimentally in natural settings.

To summarize, technological advancements are offering management researchers the opportunity to implement research designs that strike a good balance in terms of usual trade-offs between internal and external validity. Experience sampling methodology, eLancing, and VR are three examples of such innovative research tools. These innovations have allowed for important and meaningful theoretical progress in several research domains (e.g., stress, job satisfaction, personnel selection and placement, training and development, team dynamics) and have great potential in terms of many other research domains such as upper echelons, micro-foundations, and strategy-as-practice (Floyd and Sputtek, 2011), among others (e.g., Vergne and Durand, 2010).

WISH 6: UNDERSTANDING THE NATURE AND IMPACT OF OUTLIERS

An outlier is an individual, team, firm, or any other unit that deviates markedly from others. Outliers, by virtue of being different from other cases, usually, but not always, exert disproportionate influence on substantive conclusions regarding relationships among variables. Outliers are quite pervasive, although the majority of authors do not discuss how they are identified or addressed in articles published in management journals (Aguinis et al., 2013a). Due to their pervasiveness, outliers constitute one of the most enduring and pervasive methodological challenges in the entire field of management ranging from micro- (Orr et al., 1991) to macro-level (Hitt et al., 1998b) research.

In a review of the literature on outliers, Aguinis et al. (2013a) offered several illustrations of their role in terms of influencing substantive conclusions. For example, Hollenbeck et al. (2006) re-analysed data collected by Peterson et al. (2003), who investigated the relationships among CEO personality, team dynamics, and firm performance. Peterson et al. (2003) had reported 17 statistically significant correlations. However, Hollenbeck et al. (2006) showed that only one was actually significant for all 17 analyses after they removed each of the individual data points (i.e., 17 CEOs) one at a time. Stated differently, substantive conclusions regarding relationships among CEO personality, team dynamics, and firm performance changed almost completely depending on the procedure used to identify influential outliers.

Aguinis et al. (2013a) and Aguinis and Joo (in press) identified several concerns regarding outliers. Specifically, it is common for authors to be either vague or not transparent in how outliers are defined and there is also vagueness in how a particular outlier identification technique was chosen and used. In addition, outliers are typically treated as ‘problems’ that must be ‘fixed’, usually by removing particular cases from the analyses. Stated differently, articles published in management journals often include idiosyncratic, non-transparent, and difficult-to-replicate practices regarding outliers. Moreover, as noted by Aguinis et al. (2013a, p. 297), ‘a cynical view is that outliers are treated in such a way that their inclusion or exclusion from a dataset is not based on sound and standardized practices, but on whether results favor one’s preferred hypotheses’.

Our wish for the next decade is that researchers will use systematic and standardized procedures for defining, identifying, and handling outliers. In addition, we wish that a clear description of these procedures, perhaps in a section titled ‘Outlier Detection and Management’, would be included in each manuscript submitted for publication consideration.

Aguinis et al. (2013a) offered recommendations on how to define, then identify, and finally, handle outliers. Also, Aguinis and Joo (in press) offered recommendations specifically regarding how to identify influential outliers. Next, we describe these recommendations. In addition, the sequential steps for identifying and handling outliers are summarized in Figure 7. It is important to emphasize that there are two principles underlying these recommendations. First, each procedure for dealing with outliers should be described in detail to ensure transparency and include a rationale for each particular choice. Second, researchers should clearly and explicitly acknowledge the type of outlier in which they are interested, and then use identification and handling techniques that are congruent with the outlier definition.

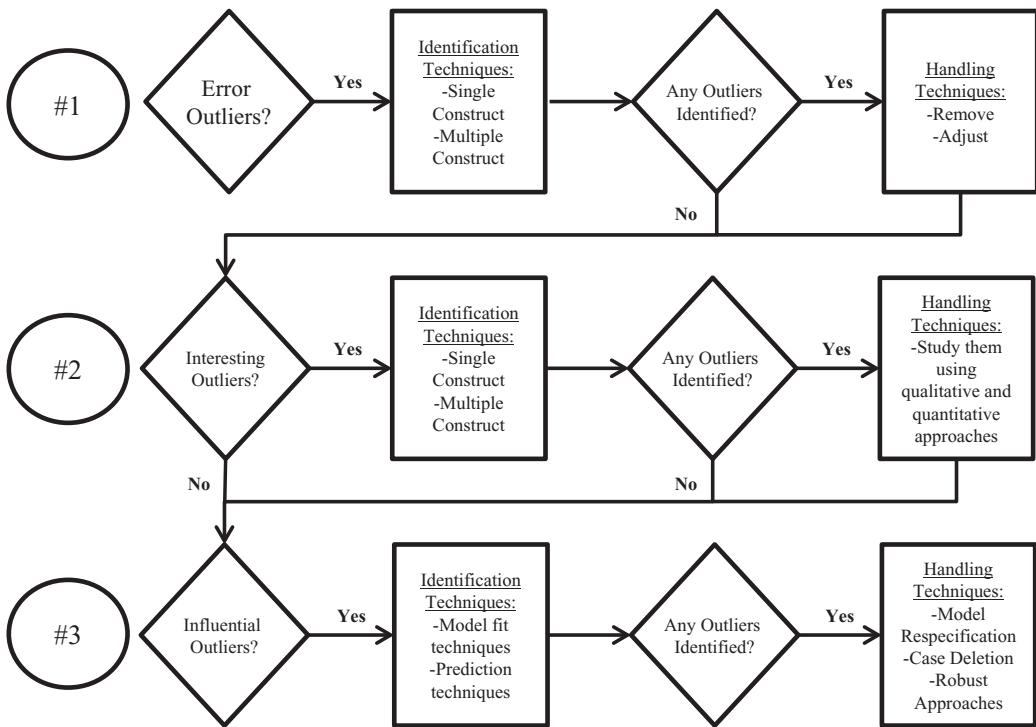


Figure 7. Graphic representation of simplified sequential steps for defining, identifying, and handling (1) error, (2) interesting, and (3) influential outliers

First, regarding their definition, there are three types of outliers: error, interesting, and influential. Error outliers are data points that lie at a distance from others because they are the result of inaccuracies. Interesting outliers are outlying data points that are accurate – that is, data points that have been identified as outlying observations (i.e., potential error outliers), but not confirmed as actual error outliers. Influential outliers include two sub-types: (a) model fit outliers, and (b) prediction outliers. Model fit outliers are data points whose presence alters the fit of a model, and prediction outliers are data points whose presence alters parameter estimates.

Second, the identification of outliers proceeds using the following systematic sequence. First, identifying error outliers involves the first step of locating outlying observations (i.e., identification of *potential* error outliers – candidates for error outliers), and then the second step of separately investigating whether the outlyingness of such data points was caused by errors (i.e., identification of *actual* error outliers). Identifying potential error outliers involves visual and quantitative techniques and using more than one approach is necessary to zero in on as many potential error outliers as possible. All data points that are outlying but were not identified as error outliers are then considered potential interesting outliers. Similar to error outliers, the identification of interesting outliers involves two steps. The first step is to identify *potential* interesting outliers, and the second step is to identify which outliers are *actual* interesting outliers. The first step is already completed based on the identification of error outliers. In the second step, the particular

research domain determines the specific type of technique that will be used to identify interesting outliers. For example, if there is an interest in identifying certain firms that outperform the competition (e.g., Aharoni, 1993), then potential interesting outliers identified through single-construct techniques would be considered actual interesting outliers. If there is an interest in studying the relationship between two constructs, such as firms that are outliers in annual profit and annual cost in research and development, or strategic groups of firms that share high performance and other characteristics (e.g., Athanassopoulos, 2003), then potential interesting outliers would be identified through multiple-construct identification techniques. Finally, in contrast to the procedures for identifying error and interesting outliers, which are fairly invariant across data-analytic approaches used to address a study's substantive questions and hypotheses (e.g., regression, structural equation modelling), influential outliers are identified using different techniques depending on the particular data-analytic approach in question. For example, if a study involves the use of multiple regression, influential outliers are identified based on their impact on the estimated R^2 (i.e., model fit outlier) or regression coefficients (i.e., prediction outlier) in the model.

In terms of how to handle outliers after they are identified, error outliers are corrected or removed from the dataset. The rationale is that this type of outlier is always undesirable because they are due to inaccuracies and often produce bias, either systematic or random, in substantive results. Interesting outliers are studied further using quantitative (e.g., St. John and Harrison, 1999) and qualitative (e.g., Hitt et al., 1998b) approaches. For example, Hitt et al. (1998b) examined firm acquisitions that were either highly successful or highly unsuccessful and identified highly successful acquisitions as 12 pairs of firms that showed increases in terms of return on assets and industry-adjusted research and development intensity. Similarly, they also identified 12 pairs of highly unsuccessful acquisitions. Subsequently, Hitt et al. (1998b) used a case study method on the 24 pairs of firms identified as interesting outliers. Doing so resulted in important theoretical advancements because they were able to derive potential predictors of outliers in the research domain of acquisitions. Finally, in terms of handling influential outliers, there are three possible courses of action (e.g., Cohen et al., 2003): (a) respecification, (b) deletion, and (c) robust approaches. Respecification involves adding additional terms in the conceptual model such as product terms carrying information about interaction effects or polynomial terms representing non-linear relationships. If the respecified model is supported (i.e., if the terms added post hoc significantly improve model fit or prediction), then authors can also build new theoretical models that can be tested, confirmed, or disconfirmed in future research. Deletion is the second option, which serves the purpose of completely removing the influence of a particular case – given that the case is deleted from the analyses. Finally, using robust techniques, which is a less extreme version of deletion because it reduces the influence of outlying cases but does not eliminate it completely, is the third option. Regardless of the use of respecification, deletion, or robust techniques, it is important to report results with and without the implementation of each of these three approaches for handling influential outliers. The rationale is that deleting cases and using robust approaches may lead to improper inferences about a population. This is why results should be reported with and without the use of these approaches to ensure transparency.

In summary, the presence of outliers seems to be a fact of life in management research. However, there is a high degree of vagueness and lack of transparency in how authors address outliers. Implementing systematic and standardized procedures for defining, identifying, and handling outliers and offering a detailed description of these procedures in all manuscripts is likely to enhance the transparency, replicability, and credibility of future research.

WISH 7: OFFERING A REALISTIC AND USEFUL DESCRIPTION OF A STUDY'S LIMITATIONS

The field of management has made important progress in the methodological arena. This progress has been documented in several literature reviews focusing on methodological practices that, collectively, have offered analyses of the state of our knowledge (e.g., Aguinis et al., 2009; Bergh et al., 2004; Hitt et al., 2004; Scandura and Williams, 2000).

In spite of the progress made, there is no such thing as a perfectly flawless study that is able to rule out every single possible threat regarding internal, external, construct, and statistical conclusion validity (Cook and Campbell, 1979; Shadish et al., 2002). Accordingly, a common feature of peer-reviewed journal articles in the field of management is the description of a study's limitations (Brutus et al., 2010, 2013). In fact, each of the wishes we described in our manuscript was motivated by what we regard as limitations in current management research – and our desire that such limitations will be overcome in the future. As noted by Aguinis and Vandenberg (in press), many of these limitations should be addressed prior to data collection, thereby minimizing them as threats to a study's internal validity (e.g., Bergh et al., 2004), construct validity (e.g., Boyd et al., 2013), external validity (e.g., Short et al., 2002), or statistical conclusion validity (e.g., George, 2011). However, many such threats are usually not addressed prior to data collection due to a variety of reasons involving practical and logistical constraints as well as an overemphasis on data-analytic issues compared to research design and measurement issues in doctoral-level education (Aiken et al., 2008; Tett et al., 2013). Consequently, these limitations need to be acknowledged in a post-hoc fashion. For example, regarding our wish 1, authors may offer reasons why observed effects may not be as strong as desired and, regarding wish 2, authors may offer a description of the extent to which endogeneity may threaten the validity of results.

Ideally, a manuscript's limitations section should provide '... a realistic (and adequately self-critical) delineation of limitations and weaknesses' of the research presented (Campion, 1993, p. 717). Limitation sections are potentially useful for understanding the importance of the weaknesses of the particular research effort, placing the study in context, and attributing a credibility level to it (Ioannidis, 2007). However, authors are aware that the material included in a manuscript's limitations section is likely to affect the publication decision. In other words, there is an agency problem: authors are not likely to voluntarily reveal information that may lead to the rejection of their manuscript (Brutus et al., 2013). Pressures stemming from the increasingly low acceptance rates for peer-reviewed journals and university reward systems relying heavily on

such publications (Certo et al., 2010) represent 'clear motives for not acknowledging limitations and for offering only benign directions for future research' (Brutus et al., 2013, p. 49). To investigate this issue empirically, Brutus et al. (2013) conducted a content analysis of 1276 articles published between 1982 and 2007 in *AMJ*, *ASQ*, *JOM*, *SMJ*, and *Journal of Applied Psychology*. Not surprisingly, they found that, on average, each published article included a description of only 1.27 limitations. This result was similar to what has been found in other closely related fields. For example, Aguinis and Lawal (2012) reported a similar average of 1.94 self-reported limitations per article in entrepreneurship and Brutus et al. (2010) reported an average of 1.66 self-reported limitations per article in industrial and organizational psychology. Given our knowledge of dozens of potential threats to the validity of a study's conclusions (Cook and Campbell, 1979; Shadish et al., 2002), do we really believe that, on average, the accuracy of results from empirical studies in the field of management is potentially threatened by an average of between one and two potential limitations *only*?

Our wish for the next decade is that all manuscripts will include a separate section offering an honest and realistic assessment of the limitations of the research together with consequences of such limitations for the interpretation of results and implications. In other words, this separate section would offer a clear and detailed description of each limitation together with a clear and detailed description of how results and conclusions are affected by this limitation.

Due to the agency problem noted earlier, it is doubtful that authors will volunteer information that greatly threatens the possibility of successful publication. Accordingly, the onus falls primarily on journal reviewers and editors to identify these limitations and request that authors include them in their manuscript if they are not. Editorial policies of journals in other fields such as the *Journal of the American Medical Association* and the *British Medical Journal* require that submissions adhere to a very strict structure (Doherty and Smith, 1999). Similar to Brutus et al. (2013), our first recommendation in terms of making our wish come true is that journal policies be revised to include a requirement about a separate section describing limitations. This change in editorial policies seems necessary given that Brutus et al. (2013) found that 38 per cent of articles did not report any limitation at all. In addition, Brutus et al. (2013) found an important difference across management sub-fields regarding a separate section addressing limitations. Specifically, while 68 per cent of articles addressing organizational behaviour issues included a separate limitations section, only 35 per cent of articles addressing strategic management topics did.

Second, related to the first recommendation, Brutus et al. (2013) suggested that reviewers be asked explicitly to identify limitations and their implications. To do so, it would be useful to revise reviewer evaluation forms to include a separate set of questions asking them to list limitations and the extent to which the study is affected by them in terms of substantive conclusions.

Third, in describing limitations, authors may be tempted to describe limitations that do not really matter, or limitations that are perceived to not be 'fatal flaws'. For example, a manuscript may include a generic statement about external validity such as 'our study addressed one sample only, so future research should be conducted using additional samples'. In this way, authors may feel they have 'checked off' the limitations box. Thus,

once again, it is up to reviewers and editors to point to issues that really matter in terms of placing a limit on the degree of confidence in a study's results and implications.

Fourth, it is not sufficient to describe limitations. Specifically, each limitation should be accompanied with a statement regarding its impact on a study's conclusions. For example, it is not sufficient to note that 'due to the lack of manipulation of variables and random assignment, results regarding causal relationships should be interpreted with caution'. A more useful statement is that 'our results do not provide evidence of whether A causes B, B causes A, or both A and B are caused by a third variable; accordingly, interventions based on A should not necessarily lead to B'. Such a statement offers a more clear explanation of the consequences of the limitation in terms of theory and practice.

Finally, Brutus et al.'s (2013) content analysis discovered that many authors do not offer a description and explanation of a study's limitations. Rather, many authors offer justifications in an attempt to minimize the impact of a limitation – possibly as a 'preemptive strike' in seeking a positive publication decision. For example, Brutus et al. (2013) reported that a common rhetorical tool is to describe a limitation but immediately discount it as an issue that is minor and does not threaten the interpretation of results. Many readers will recognize statements along the lines of: 'The study had limitation X, but X does not really matter that much because of Y and Z.' So, as a final recommendation, reviewers and editors should play an important role in ensuring that 'reported limitations contain the details necessary to make them informative for readers' (Brutus et al., 2013, p. 67).

In sum, following the recommendations offered in our manuscript as well as elsewhere (e.g., Aguinis and Vandenberg, in press) will strengthen the confidence in the results of our research. Nevertheless, situational constraints preclude empirical studies from being absolutely flawless in terms of internal, external, construct, and statistical conclusion validity. Including a separate section regarding limitations, and their consequences for a study's results and implications, will allow a sceptical scientific audience as well as other consumers of our research to allocate a more appropriate degree of confidence to the reported results.

CONCLUSIONS

As noted earlier, much progress has been made in the field of management since the first issue of *JMS* 50 years ago. However, management is still a young scientific field compared to physics, biology, and astronomy, which have developed over hundreds of years. In contrast to researchers in those domains, management researchers do not enjoy the benefit of methodological tools such as the Large Hadron Collider used by physicists or the Hubble telescope used by astronomers, which allow for strong tests of precise hypotheses with a high degree of measurement precision. Nonetheless, management researchers can strive to improve the methodology used in empirical research, which is a prerequisite for theoretical progress and the accumulation of knowledge. Accordingly, we hope that our methodological wishes, together with actionable recommendations we have offered for authors, reviewers, and journal editors, will serve as catalysts for the advancement of management research.

ACKNOWLEDGMENTS

We thank Don Bergh, Rich Bettis, Andrew Corbett, Dan Dalton, Andrew Delios, Bill Harley, and Atul Nerkar for stimulating exchanges that allowed us to improve our manuscript substantially.

REFERENCES

- Adner, R., Polos, L., Ryall, M. and Sorenson, O. (2009). 'The case for formal theory'. *Academy of Management Review*, **34**, 201–8.
- Aguinis, H. and Glavas, A. (2013). 'Embedded versus peripheral corporate social responsibility: psychological foundations'. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, **6**, 314–32.
- Aguinis, H. and Joo, H. (in press). 'Debunking myths and urban legends about how to identify influential outliers'. In Lance, C. E. and Vandenberg, R. J. (Eds), *Statistical and Methodological Myths and Urban Legends: Received Doctrine, Verity, and Fable in the Organizational and Social Sciences*, 2nd edition. New York: Routledge.
- Aguinis, H. and Lawal, S. O. (2012). 'Conducting field experiments using eLancing's natural environment'. *Journal of Business Venturing*, **27**, 493–505.
- Aguinis, H. and Lawal, S. O. (2013). 'eLancing: a review and research agenda for bridging the science-practice gap'. *Human Resource Management Review*, **23**, 6–17.
- Aguinis, H. and Vandenberg, R. J. (in press). 'An ounce of prevention is worth a pound of cure: improving research quality before data collection'. *Annual Review of Organizational Psychology and Organizational Behavior*.
- Aguinis, H., Henle, C. A. and Beaty, J. C. (2001). 'Virtual reality technology: a new tool for personnel selection'. *International Journal of Selection and Assessment*, **9**, 70–83.
- Aguinis, H., Pierce, C. A., Bosco, F. A. and Muslin, I. S. (2009). 'First decade of Organizational Research Methods: trends in design, measurement, and data-analysis topics'. *Organizational Research Methods*, **12**, 69–112.
- Aguinis, H., Boyd, B. K., Pierce, C. A. and Short, J. C. (2011a). 'Walking new avenues in management research methods and theories: bridging micro and macro domains'. *Journal of Management*, **37**, 395–403.
- Aguinis, H., Dalton, D. R., Bosco, F. A., Pierce, C. A. and Dalton, C. M. (2011b). 'Meta-analytic choices and judgment calls: implications for theory building and testing, obtained effect sizes, and scholarly impact'. *Journal of Management*, **37**, 5–38.
- Aguinis, H., Gottfredson, R. K. and Joo, H. (2013a). 'Best-practice recommendations for defining, identifying, and handling outliers'. *Organizational Research Methods*, **16**, 270–301.
- Aguinis, H., Gottfredson, R. K. and Culpepper, S. A. (2013b). 'Best-practice recommendations for estimating cross-level interaction effects using multilevel modeling'. *Journal of Management*, **39**, 1490–528.
- Aharoni, Y. (1993). 'In search for the unique: can firm specific advantages be evaluated?'. *Journal of Management Studies*, **30**, 31–47.
- Aiken, L. S., West, S. G. and Millsap, R. E. (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'. *American Psychologist*, **63**, 32–50.
- Angrist, J. D., Imbens, G. W. and Rubin, D. B. (1996). 'Identification of causal effects using instrumental variables'. *Journal of the American Statistical Association*, **91**, 444–55.
- Antonakis, J., Bendahan, S., Jacquart, P. and Lalive, R. (2010). 'On making causal claims: a review and recommendations'. *Leadership Quarterly*, **21**, 1086–120.
- Athanassopoulos, A. D. (2003). 'Strategic groups, frontier benchmarking and performance differences: evidence from the UK retail grocery industry'. *Journal of Management Studies*, **40**, 921–53.
- Atinc, G., Simmering, M. J. and Kroll, M. J. (2011). 'Control variable use and reporting in macro and micro management research'. *Organizational Research Methods*, **15**, 57–74.
- Bagozzi, R. P., Yi, Y. and Phillips, L. W. (1991). 'Assessing construct validity in organizational research'. *Administrative Science Quarterly*, **36**, 421–58.
- Bakan, D. (1966). 'The test of significance in psychological research'. *Psychological Bulletin*, **66**, 423–37.
- Bal, M. P., Kooij, D. T. A. M. and De Jong, S. B. (2013). 'How do developmental and accommodative HRM enhance employee engagement and commitment? The role of psychological contract and SOC-strategies'. *Journal of Management Studies*, **50**, 545–72.
- Bascle, G. (2008). 'Controlling for endogeneity with instrumental variables in strategic management research'. *Strategic Organization*, **6**, 285–327.

- Beal, D. J. and Weiss, H. M. (2003). 'Methods of ecological momentary assessment in organizational research'. *Organizational Research Methods*, **6**, 440–64.
- Becker, T. E. (2005). 'Potential problems in the statistical control of variables in organizational research: a qualitative analysis with recommendations'. *Organizational Research Methods*, **8**, 274–89.
- Bergh, D. D. and Holbein, G. (1997). 'Assessing and redirecting longitudinal analysis: demonstration with a study of the diversification and divestiture relationship'. *Strategic Management Journal*, **18**, 557–71.
- Bergh, D. D., Hanke, R., Balkundi, P., Brown, M. and Chen, X. (2004). 'An assessment of research design in strategic management research: the frequency of threats to internal validity'. In Ketchen, D. J. and Bergh, D. D. (Eds), *Research Methodology in Strategy and Management*. Vol. 1. Oxford: Emerald Group Publishing, 347–63.
- Binder, A. (1963). 'Further considerations on testing the null hypothesis and the strategy and tactics of investigating theoretical models'. *Psychological Review*, **70**, 107–15.
- Blalock, H. M. (1964). *Causal Inferences in Nonexperimental Research*. Chapel Hill, NC: University of North Carolina Press.
- Bliese, P. D. and Hanges, P. J. (2004). 'Being both too liberal and too conservative: the perils of treating grouped data as though they were independent'. *Organizational Research Methods*, **7**, 400–17.
- Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: Wiley.
- Bollen, K. A. (2012). 'Instrumental variables in sociology and the social sciences'. *Annual Review of Sociology*, **38**, 37–72.
- Borsboom, D., Mellenbergh, G. J. and van Heerden, J. (2004). 'The concept of validity'. *Psychological Review*, **111**, 1061–71.
- Bowen, H. B. and Wiersema, M. F. (1999). 'Matching method to paradigm in strategy research: limitations of cross-sectional analysis and some methodological alternatives'. *Strategic Management Journal*, **20**, 625–36.
- Boyd, B. K., Gove, S. and Hitt, M. A. (2005). 'Construct measurement in strategic management research: illusion or reality?'. *Strategic Management Journal*, **26**, 239–57.
- Boyd, B. K., Ketchen, D. J. and Bergh, D. D. (2008). 'Research methodology in strategic management. Past accomplishments and future challenges'. *Organizational Research Methods*, **11**, 643–58.
- Boyd, B. K., Bergh, D. D., Ireland, R. D. and Ketchen, D. J. (2013). 'Constructs in strategic management'. *Organizational Research Methods*, **16**, 3–14.
- Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York: The Guilford Press.
- Browne, M. W. (1984). 'Asymptotic distribution free methods in analysis of covariance structures'. *British Journal of Mathematical and Statistical Psychology*, **37**, 62–83.
- Brutus, S., Gill, H. and Duniewicz, K. (2010). 'Self-reported limitations in industrial and organizational psychology'. *Personnel Psychology*, **63**, 907–36.
- Brutus, S., Aguinis, H. and Wassmer, U. (2013). 'Self-reported limitations and future directions in scholarly reports: analysis and recommendations'. *Journal of Management*, **39**, 48–75.
- Buckley, M. R., Riaz Hamdani, M., Klotz, A. C. and Valcea, S. (2011). 'Into the great wide open: bridging the micro-macro divide in the organizational sciences'. In Ketchen, D. J. and Bergh, D. D. (Eds), *Research Methodology in Strategy and Management*. Vol. 6. Oxford: Emerald Group Publishing, 31–68.
- Campion, M. A. (1993). 'Article review checklist: a criterion checklist for reviewing research articles in applied psychology'. *Personnel Psychology*, **46**, 705–18.
- Carlson, K. D. and Wu, J. (2012). 'The illusion of statistical control: control variable practice in management research'. *Organizational Research Methods*, **15**, 415–35.
- Carpenter, M. A. and Reilly, G. P. (2006). 'Constructs and construct measurement in upper echelons research'. In Ketchen, D. J. and Bergh, D. D. (Eds), *Research Methodology in Strategy and Management*. Vol. 3. Oxford: Emerald Group Publishing, 17–35.
- Carver, R. P. (1978). 'The case against statistical significance testing'. *Harvard Education Review*, **48**, 378–99.
- Cattell, R. B. (1978). *Scientific Use of Factor Analysis in Behavioral and Life Sciences*. New York: Plenum Press.
- Certo, S. T., Sirmon, D. G. and Brymer, R. (2010). 'Competition and scholarly productivity in management: investigating changes in scholarship from 1988 to 2008'. *Academy of Management Learning and Education*, **9**, 591–606.
- Cochran, W. G. (1968). 'The effectiveness of adjustment by subclassification in removing bias in observational studies'. *Biometrics*, **24**, 295–313.
- Cohen, J. (1994). 'The earth is round ($p < .05$)'. *American Psychologist*, **49**, 997–1003.
- Cohen, J., Cohen, P., West, S. G. and Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 3rd edition. Mahwah, NJ: Erlbaum.

- Collins, S. L. and Winship, C. (2007). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. New York: Cambridge University Press.
- Cook, T. D. and Campbell, D. T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago, IL: Rand McNally.
- Cortina, J. M. and Dunlap, W. P. (1997). 'On the logic and purpose of significance testing'. *Psychological Methods*, **2**, 161–72.
- Cronbach, L. J. (1951). 'Coefficient alpha and the internal structure of tests'. *Psychometrika*, **16**, 297–334.
- Croon, M. A. and van Veldhoven, M. J. P. M. (2007). 'Predicting group-level outcome variables from variables measured at the individual level: a latent variable multilevel model'. *Psychological Methods*, **12**, 45–57.
- Davis, G. F. (2010). 'Do theories of organizations progress?'. *Organizational Research Methods*, **13**, 690–709.
- Davis, R. and Duhaime, I. M. (1992). 'Diversification, vertical integration, and industry analysis: new perspectives and measurement'. *Strategic Management Journal*, **13**, 511–24.
- Devinney, T. M. (2013). 'Is microfoundational thinking critical to management thought and practice?'. *Academy of Management Perspectives*, **27**, 81–4.
- Diamantopoulos, A. and Siguaw, J. A. (2006). 'Formative versus reflective indicators in organizational measure development: a comparison and empirical illustration'. *British Journal of Management*, **17**, 263–82.
- Dipboye, R. L. and Flanagan, M. F. (1979). 'Research settings in industrial and organizational psychology: are findings in the field more generalizable than in the laboratory?'. *American Psychologist*, **34**, 141–50.
- Doherty, M. and Smith, R. (1999). 'The case for structuring the discussion of scientific papers'. *British Medical Journal*, **318**, 1224–25.
- Echambadi, R., Campbell, B. and Agarwal, R. (2006). 'Encouraging best practice in quantitative management research: an incomplete list of opportunities'. *Journal of Management Studies*, **43**, 1801–20.
- Edwards, J. R. (2001). 'Multidimensional constructs in organizational behavior research: an integrative analytical framework'. *Organizational Research Methods*, **4**, 144–92.
- Edwards, J. R. (2003). 'Construct validation in organizational behavior research'. In Greenberg, J. (Ed.), *Organizational Behavior: The State of the Science*, 2nd edition. Mahwah, NJ: Erlbaum, 327–71.
- Edwards, J. R. (2008). 'To prosper organizational psychology should . . . overcome methodological barriers to progress'. *Journal of Organizational Behavior*, **29**, 469–91.
- Edwards, J. R. (2011). 'The fallacy of formative measurement'. *Organizational Research Methods*, **14**, 370–88.
- Edwards, J. R. and Bagozzi, R. P. (2000). 'On the nature and direction of the relationship between constructs and measures'. *Psychological Methods*, **5**, 155–74.
- Edwards, J. R. and Berry, J. W. (2010). 'The presence of something or the absence of nothing: increasing theoretical precision in management research'. *Organizational Research Methods*, **13**, 668–89.
- Edwards, J. R. and O'Neill, R. M. (1998). 'The construct validity of scores on the Ways of Coping Questionnaire: confirmatory analysis of alternative factor structures'. *Educational and Psychological Measurement*, **58**, 955–83.
- Felin, T. and Foss, N. (2006). 'Individuals and organizations: thoughts on a micro-foundations project for strategic management and organizational analysis'. In Ketchen, D. J. and Bergh, D. D. (Eds), *Research Methodology in Strategy and Management*. Vol. 3. Oxford: Emerald Group Publishing, 253–88.
- Felin, T., Foss, N. J., Heimeriks, K. H. and Madsen, T. L. (2012). 'Microfoundations of routines and capabilities: individuals, processes, and structure'. *Journal of Management Studies*, **49**, 1351–74.
- Ferris, G. R., Bowen, M. G., Treadway, D. C., Hochwarter, W. A., Hall, A. T. and Perrewé, P. L. (2006). 'The assumed linearity of organizational phenomena: implications for occupational stress and well-being'. In Perrewé, P. L. and Ganster, D. C. (Eds), *Research in Occupational Stress and Well-Being*. Vol. 5. Oxford: Elsevier Science, 205–32.
- Floyd, S. W. and Splettek, R. (2011). 'Rediscovering the individual in strategy: methodological challenges, strategies, and prospects'. In Ketchen, D. J. and Bergh, D. D. (Eds), *Research Methodology in Strategy and Management*. Vol. 6. Oxford: Emerald Group Publishing, 3–30.
- Foo, M., Uy, M. and Baron, R. A. (2009). 'How do feelings influence effort? An empirical study of entrepreneurs' affect and venture effort'. *Journal of Applied Psychology*, **94**, 1086–94.
- Foss, N. J. (2011). 'Why micro-foundations for resource-based theory are needed and what they may look like'. *Journal of Management*, **37**, 1413–28.
- George, B. A. (2011). 'Entrepreneurial orientation: a theoretical and empirical examination of the consequences of differing construct representations'. *Journal of Management Studies*, **48**, 1291–313.
- Gigerenzer, G. (1998). 'Surrogates for theories'. *Theory and Psychology*, **8**, 195–204.
- Gorsuch, R. L. (1983). *Factor Analysis*, 2nd edition. Hillsdale, NJ: Erlbaum.

- Grant, A. M. and Wall, T. D. (2009). 'The neglected science and art of quasi-experimentation'. *Organizational Research Methods*, **12**, 653–86.
- Hamilton, B. H. and Nickerson, J. A. (2003). 'Correcting for endogeneity in strategic management research'. *Strategic Organization*, **1**, 51–78.
- Heckman, J. J. (1979). 'Sample selection bias as a specification error'. *Econometrica*, **47**, 153–61.
- Highhouse, S. (2009). 'Designing experiments that generalize'. *Organizational Research Methods*, **12**, 554–66.
- Hinkin, T. R. and Tracey, J. B. (1999). 'An analysis of variance approach to content validation'. *Organizational Research Methods*, **2**, 175–86.
- Hitt, M. A., Gimeno, J. and Hoskisson, R. E. (1998a). 'Current and future research methods in strategic management'. *Organizational Research Methods*, **1**, 6–44.
- Hitt, M. A., Harrison, J. S., Ireland, R. D. and Best, A. (1998b). 'Attributes of successful and unsuccessful acquisitions of US firms'. *British Journal of Management*, **9**, 91–114.
- Hitt, M. A., Boyd, B. K. and Li, D. (2004). 'The state of strategic management research and a vision of the future'. *Research Methodology in Strategy and Management*, **1**, 1–31.
- Holland, P. W. (1986). 'Statistics and causal inference'. *Journal of the American Statistical Association*, **81**, 945–60.
- Hollenbeck, J. R., DeRue, D. S. and Mannor, M. (2006). 'Statistical power and parameter stability when subjects are few and tests are many: comment on Peterson, Smith, Martorana, and Owens (2003)'. *Journal of Applied Psychology*, **91**, 1–5.
- Howard, G. S., Maxwell, S. E. and Fleming, K. J. (2000). 'The proof of the pudding: an illustration of the relative strengths of null hypothesis, meta-analysis, and Bayesian analysis'. *Psychological Methods*, **5**, 315–32.
- Howell, R. D., Breivik, E. and Wilcox, J. B. (2007). 'Reconsidering formative measurement'. *Psychological Methods*, **12**, 205–18.
- Hox, J. (2010). *Multilevel Analysis: Techniques and Applications*, 2nd edition. New York: Routledge.
- Ilies, R. and Judge, T. A. (2002). 'Understanding the dynamic relationships among personality, mood, and job satisfaction: a field experience sampling study'. *Organizational Behavior and Human Decision Processes*, **89**, 1119–39.
- Illari, P. M., Russo, F. and Williamson, J. (Eds) (2011). *Causality in the Sciences*. Oxford: Oxford University Press.
- Ioannidis, J. P. A. (2007). 'Limitations are not properly acknowledged in the scientific literature'. *Journal of Clinical Epidemiology*, **60**, 324–9.
- Jackson, D. L., Gillaspay, J. A., Jr and Purc-Stephenson, R. (2009). 'Reporting practices in confirmatory factor analysis: an overview and some recommendations'. *Psychological Methods*, **14**, 6–23.
- Kenny, D. A. and Judd, C. M. (1996). 'A general procedure for the estimation of interdependence'. *Psychological Bulletin*, **119**, 138–48.
- Ketchen, D. J., Boyd, B. K. and Bergh, D. D. (2008). 'Research methodology in strategic management: past accomplishments and future challenges'. *Organizational Research Methods*, **11**, 643–58.
- Ketchen, D. J., Jr, Ireland, R. D. and Baker, L. T. (2013). 'The use of archival proxies in strategic management studies: castles made of sand?'. *Organizational Research Methods*, **16**, 32–42.
- Kirk, R. E. (1996). 'Practical significance: a concept whose time has come'. *Educational and Psychological Measurement*, **56**, 746–59.
- Kline, R. B. (2010). *Principles and Practice of Structural Equation Modeling*, 4th edition. New York: Guilford Press.
- Krueger, J. (2001). 'Null hypothesis significance testing: on the survival of a flawed method'. *American Psychologist*, **56**, 16–26.
- Kruschke, J. K. (2010). *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*. New York: Academic Press.
- Kruschke, J. K., Aguinis, H. and Joo, H. (2012). 'The time has come: Bayesian methods for data analysis in the organizational sciences'. *Organizational Research Methods*, **15**, 722–52.
- Lee, P. M. (2004). *Bayesian Statistics: An Introduction*, 3rd edition. New York: Wiley.
- Li, M. (2013). 'Using the propensity score method to estimate causal effects: a review and practical guide'. *Organizational Research Methods*, **16**, 188–226.
- Little, R. J. and Rubin, D. B. (2000). 'Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches'. *Annual Review of Public Health*, **21**, 121–45.
- Lykken, D. T. (1968). 'Statistical significance in psychological research'. *Psychological Bulletin*, **70**, 151–9.
- MacCallum, R. and Browne, M. W. (1993). 'The use of causal indicators in covariance structure models: some practical issues'. *Psychological Bulletin*, **114**, 533–41.
- MacCallum, R., Roznowski, M. and Necowitz, L. B. (1992). 'Model modification in covariance structure analysis: the problem of capitalization on chance'. *Psychological Bulletin*, **111**, 490–504.

- Marco, C. M. and Suls, J. (1993). 'Daily stress and the trajectory of mood: spillover, contrast, and trait negative affectivity'. *Journal of Personality and Social Psychology*, **64**, 1053–63.
- Martens, E. P., Pestman, W. R., de Boer, A., Belitser, S. V. and Klungel, O. H. (2006). 'Instrumental variables: applications and limitations'. *Epidemiology*, **17**, 260–7.
- Mathieu, J. E., Aguinis, H., Culpepper, S. A. and Chen, G. (2012). 'Understanding and estimating the power to detect cross-level interaction effects in multilevel modeling'. *Journal of Applied Psychology*, **97**, 951–66.
- McNemar, Q. (1962). *Psychological Statistics*, 3rd edition. New York: John Wiley.
- Meehl, P. E. (1967). 'Theory-testing in psychology and physics: a methodological paradox'. *Philosophy of Science*, **34**, 103–15.
- Meehl, P. E. (1978). 'Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology'. *Journal of Consulting and Clinical Psychology*, **46**, 806–34.
- Meehl, P. E. (1990). 'Appraising and amending theories: the strategy of Lakatosian defense and two principles that warrant it'. *Psychological Inquiry*, **1**, 108–41.
- Messick, S. (1995). 'Validity of psychological assessment'. *American Psychologist*, **50**, 741–9.
- Morgeson, F. P., Aguinis, H., Waldman, D. A. and Siegel, D. S. (in press). 'Extending corporate social responsibility research to the human resource management and organizational behavior domains: a look to the future'. *Personnel Psychology*.
- Mulaik, S. A. (2009a). *Foundations of Factor Analysis*, 2nd edition. London: Taylor and Francis.
- Mulaik, S. A. (2009b). *Linear Causal Modeling with Structural Equations*. Boca Raton, FL: Chapman and Hall/CRC of Taylor and Francis Group.
- Mulaik, S. A., Raju, N. S. and Harshman, R. A. (1997). 'There is a time and place for significance testing'. In Harlow, L. L., Mulaik, S. A. and Steiger, J. H. (Eds), *What If There Were No Significance Tests?* Hillsdale, NJ: Erlbaum, 65–116.
- Nickerson, R. S. (2000). 'Null hypothesis significance testing: a review of an old and continuing controversy'. *Psychological Methods*, **5**, 241–301.
- Ntzoufras, I. (2009). *Bayesian Modeling Using WinBUGS*. Hoboken, NJ: Wiley.
- Orlitzky, M. (2012). 'How can significance tests be deinstitutionalized?'. *Organizational Research Methods*, **15**, 199–228.
- Orr, J. M., Sackett, P. R. and DuBois, C. L. Z. (1991). 'Outlier detection and treatment in I/O psychology: a survey of researcher beliefs and an empirical illustration'. *Personnel Psychology*, **44**, 473–86.
- Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*, 2nd edition. New York: Cambridge University Press.
- Peterson, R. S., Smith, D. B., Martorana, P. V. and Owens, P. D. (2003). 'The impact of chief executive officer personality on top management team dynamics: one mechanism by which leadership affects organizational performance'. *Journal of Applied Psychology*, **88**, 795–808.
- Pfeffer, J. (1993). 'Barriers to the advance of organizational science: paradigm development as a dependent variable'. *Academy of Management Review*, **18**, 599–620.
- Pierce, C. A. and Aguinis, H. (1997). 'Using virtual reality technology in organizational behavior research'. *Journal of Organizational Behavior*, **18**, 407–10.
- Pierce, J. A. and Aguinis, H. (2013). 'The too-much-of-a-good-thing effect in management'. *Journal of Management*, **39**, 313–38.
- Platt, J. R. (1964). 'Strong inference'. *Science*, **146**, 347–53.
- Ployhart, R. E. and Vandenberg, R. J. (2010). 'Longitudinal research: the theory, design, and analysis of change'. *Journal of Management*, **36**, 94–120.
- Podsakoff, N. P., Shen, W. and Podsakoff, P. M. (2006). 'The role of formative measurement models in strategic management research: review, critique and implications for future research'. In Ketchen, D. J. and Bergh, D. D. (Eds), *Research Methodology in Strategy and Management*. Vol. 3. Burlington, MA: Elsevier, 197–252.
- Podsakoff, P. M. and Dalton, D. R. (1987). 'Research methodology in organizational studies'. *Journal of Management*, **13**, 419–41.
- Pruzek, R. M. (1997). 'An introduction to Bayesian inference and its application'. In Harlow, L. L., Mulaik, S. A. and Steiger, J. J. (Eds), *What If There Were No Significance Tests?* Hillsdale, NJ: Erlbaum, 287–318.
- Rindskopf, D. and Rose, T. (1988). 'Some theory and applications of confirmatory second-order factor analysis'. *Multivariate Behavioral Research*, **23**, 51–67.
- Rosenbaum, P. R. and Rubin, D. B. (1985). 'Constructing a control group using multivariate matched sampling methods that incorporate the propensity score'. *American Statistician*, **39**, 33–8.

- Rozeboom, W. W. (1960). 'The fallacy of the null-hypothesis significance test'. *Psychological Bulletin*, **57**, 416–28.
- Rubin, D. B. (1974). 'Estimating causal effects of treatments in randomized and nonrandomized studies'. *Journal of Educational Psychology*, **66**, 688–701.
- Rubin, D. B. (1978). 'Bayesian inference for causal effects: the role of randomization'. *Annals of Statistics*, **6**, 34–58.
- Runkel, P. J. and McGrath, J. E. (1972). *Research on Human Behavior: A Systematic Guide to Method*. New York: Holt.
- Satorra, A. and Bentler, P. M. (1994). 'Corrections to test statistics and standard errors in covariance structure analysis'. In von Wye, A. and Clogg, C. C. (Eds), *Latent Variable Analysis: Applications for Developmental Research*. Thousand Oaks, CA: Sage, 399–419.
- Scandura, T. and Williams, E. (2000). 'Research methodology in management: current practices, trends, and implications for future research'. *Academy of Management Journal*, **43**, 1248–64.
- Schmidt, F. L. (1992). 'What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology'. *American Psychologist*, **47**, 1173–81.
- Schmidt, F. L. (1996). 'Statistical significance testing and cumulative knowledge in psychology: implications for training of researchers'. *Psychological Methods*, **1**, 115–29.
- Schumacker, R. E. and Lomax, R. G. (2004). *A Beginner's Guide to Structural Equation Modeling*, 2nd edition. Hillsdale, NJ: Erlbaum.
- Schwab, A. and Starbuck, W. H. (2009). 'Null-hypothesis significance tests in behavioral and management research: we can do better'. In Bergh, D. D. and Ketchen, Jr., D. J. (Eds), *Research Methodology in Strategy and Management*. Vol. 5. Bingley: Emerald, 29–54.
- Schwab, D. P. (1980). 'Construct validity in organizational behavior'. In Cummings, L. L. and Staw, B. M. (Eds), *Research in Organizational Behavior*. Vol. 2. Greenwich, CT: JAI Press, 3–43.
- Semadeni, M., Withers, M. C. and Certo, S. T. (in press). 'The perils of endogeneity and instrumental variables in strategy research: understanding through simulations'. *Strategic Management Journal*, doi: 10.1002/smj.2136.
- Serlin, R. C. (1993). 'Confidence intervals and the scientific method: a case for Holm on the range'. *Journal of Experimental Education*, **61**, 350–60.
- Serlin, R. C. and Lapsley, D. K. (1985). 'Rationality in psychological research: the good-enough principle'. *American Psychologist*, **40**, 73–83.
- Shadish, W. R., Cook, T. D. and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.
- Short, J. C., Ketchen, D. J. and Palmer, P. B. (2002). 'The role of sampling in strategic management research on performance: a two-study analysis'. *Journal of Management*, **28**, 363–85.
- Sivia, D. and Skilling, J. (2006). *Data Analysis: A Bayesian Tutorial*. New York: Oxford University Press.
- Snijders, T. A. B. and Bosker, R. J. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*, 2nd edition. Los Angeles, CA: Sage.
- Spector, P. E. and Brannick, M. T. (2011). 'Methodological urban legends: the misuse of statistical control variables'. *Organizational Research Methods*, **14**, 287–305.
- Stanney, K. M. (Ed.) (2002). *Handbook of Virtual Environments*. Mahwah, NJ: Lawrence Erlbaum.
- St. John, C. H. and Harrison, J. S. (1999). 'Manufacturing-based relatedness, synergy, and coordination'. *Strategic Management Journal*, **20**, 129–45.
- Tett, R. P., Walser, B., Brown, C., Simonet, D. V. and Tonidandel, S. (2013). 'The 2011 SIOP graduate program benchmarking survey; Part 3: curriculum and competencies'. *Industrial-Organizational Psychologist*, **50**, 69–89.
- Tryon, W. W. (2001). 'Evaluating statistical difference, equivalence, and indeterminacy using inferential confidence intervals: an integrated alternative method of conducting null hypothesis statistical tests'. *Psychological Methods*, **6**, 371–86.
- Uy, M. A., Foo, M. D. and Aguinis, H. (2010). 'Using event sampling methodology to advance entrepreneurship theory and research'. *Organizational Research Methods*, **13**, 31–54.
- Van Maanen, J., Sørensen, J. B. and Mitchell, T. R. (2007). 'The interplay between theory and method'. *Academy of Management Review*, **32**, 1145–54.
- Venkatraman, N. and Grant, J. H. (1986). 'Construct measurement in organizational strategy research: a critique and proposal'. *Academy of Management Review*, **11**, 71–87.
- Vergne, J. and Durand, R. (2010). 'The missing link between the theory and empirics of path dependence: conceptual clarification, testability issue, and methodological implications'. *Journal of Management Studies*, **47**, 736–59.

- Wei, L. and Wu, L. (2013). 'What a diverse top management team means: testing an integrated model'. *Journal of Management Studies*, **50**, 389–412.
- Whetten, D. A. (1989). 'What constitutes a theoretical contribution?'. *Academy of Management Review*, **14**, 490–95.
- Whittington, R. (2006). 'Completing the practice turn in strategy research'. *Organization Studies*, **27**, 613–34.