The Leadership Quarterly 25 (2014) 152–179

Contents lists available at ScienceDirect

The Leadership Quarterly

journal homepage: www.elsevier.com/locate/leaqua

# What makes articles highly cited?

John Antonakis [a,*], Nicolas Bastardoz [b,1], Yonghong Liu [c,2], Chester A. Schriesheim [c,3]

[a] University of Lausanne, Faculty of Business and Economics, Department of Organizational Behavior, Internef 618, CH-1015 Lausanne-Dorigny, Switzerland
[b] University of Lausanne, Faculty of Business and Economics, Department of Organizational Behavior, Internef 622, CH-1015 Lausanne-Dorigny, Switzerland
[c] University of Miami, School of Business Administration, Department of Management, 5250 University Drive, Coral Gables, FL 33124, USA

ARTICLE INFO

*Article history:*
Received 1 August 2013
Received in revised form 18 October 2013
Accepted 31 October 2013
Available online 4 December 2013

Editor: Francis J. Yammarino

*Keywords:*
Quantitative research
Qualitative research
Citations
Research impact
Research methods

ABSTRACT

We examined drivers of article citations using 776 articles that were published from 1990 to 2012 in a broad-based and high-impact social sciences journal, *The Leadership Quarterly*. These articles had 1191 unique authors having published and received in total (at the time of their most recent article published in our dataset) 16,817 articles and 284,777 citations, respectively. Our models explained 66.6% of the variance in citations and showed that quantitative, review, method, and theory articles were significantly more cited than were qualitative articles or agent-based simulations. As concerns quantitative articles, which constituted the majority of the sample, our model explained 80.3% of the variance in citations; some methods (e.g., use of SEM) and designs (e.g., meta-analysis), as well as theoretical approaches (e.g., use of transformational, charismatic, or visionary type-leadership theories) predicted higher article citations. Regarding statistical conclusion validity of quantitative articles, articles having endogeneity threats received significantly fewer citations than did those using a more robust design or an estimation procedure that ensured correct causal estimation. We make several general recommendations on how to improve research practice and article citations.



## 1. Introduction

A belief that pervades the sciences is that highly cited articles reflect better quality research, and this position is reasonably well supported in the scientific literature (Antonakis & Lalive, 2008; Bergh, Perry, & Hanke, 2006; Lokker, McKibbon, McKinlay, Wilczynski, & Haynes, 2008). Given that there is much variance in citation patterns within disciplines, we sought to answer the following question: Why are some articles more highly cited than others?

The answer to this question is an important one given that highly-cited authors are, by definition, trend setters and thus very influential in their fields. Moreover, research informs, or more specifically should inform, policy (Rynes, Giluk, & Brown, 2007); and, research cannot be relevant for practice if it is not rigorous (Vermeulen, 2005). This science-practice link is particularly important in the discipline of management and applied or industrial–organizational psychology, whose topics of study include leadership, motivation, job design, selection, assessment, and employment discrimination, and have real-life practical implications that can impact important societal-, firm- and individual-level outcomes.

Authors who are productive and highly cited obtain a disproportionate amount of grant monies (Ali, Bhattacharyya, & Olejniczak, 2010) and increased salaries (Judge, Cable, Colbert, & Rynes, 2007), among other benefits; thus, heavily cited articles by virtue of their visibility are a key driver of reputation effects in academia, and should probably also have a premium impact on practice. As a result, it is important to understand what distinguishes highly cited articles from those that are less known, focusing

* Corresponding author. Tel.: +41 021 692 3438.
*E-mail addresses:* John.Antonakis@unil.ch (J. Antonakis), Nicolas.Bastardoz@unil.ch (N. Bastardoz), yliu@bus.miami.edu (Y. Liu), chet@miami.edu (C.A. Schriesheim).
[1] Tel.: +41 21 692 3450.
[2] Tel.: +1 305 284 5846.
[3] Tel.: +1 305 284 3758.

1048-9843/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.leaqua.2013.10.014

specifically on the predictive effect of scholarly methods of inquiry, individual and institutional characteristics, and the validity of the article's conclusions.

Understanding the ingredients of highly cited articles has attracted the attention of some scholars (e.g., Aksnes, 2003; Persson, 2010). However, this area of research is poorly understood in management (Judge et al., 2007). In addition, the management discipline can be characterized as a "weak paradigm" (Aguinis & Edwards, in press; Glick, Miller, & Cardinal, 2007; Pfeffer, 1993) as compared to other social science disciplines such as economics or political science. Theories in management are imprecise and there is great diversity in methods used. In addition, there is no strong consensus regarding how scholars should go about practicing their scientific discipline. The consequences of having weak paradigms in a field are manifold and can affect collective reputations and access to research funds (Pfeffer, 1993) among other outcomes (e.g., publication in top general scientific journals, impact on government policy). Of course, our discipline may not be able to reach the theoretical precision of some of the sciences, particularly the "hard sciences." However, it can certainly improve by adopting the best methodological practices from foundational disciplines like psychology and economics (Antonakis, Bendahan, Jacquart, & Lalive, 2010).

Thus, in attempting to find out what drives citations, we determined what kind of research, from both theoretical and methodological perspectives, generates interest in the academic marketplace for citations. In particular, we investigated if high-quality quantitative research, which is not plagued by endogeneity bias (see Antonakis et al., 2010), is better cited. Endogeneity bias has to do with the assumption that "standard" estimators like OLS regression or ANOVA make with respect to the independent variables being exogenous (i.e., random). Variables are exogenous if they are manipulated as in an experiment, vary randomly in nature, or cannot be determined by other variables in the modeled system. Exogenous variables do not correlate with possible omitted causes (i.e., the disturbances) of the dependent variables. Violation of this assumption often occurs in observational/field research, engendering biased estimates that are confounded and therefore uninterpretable.

We will focus on one multidisciplinary and high-impact research journal that is prominent in one particular domain, leadership. Holding constant the journal and discipline brings some advantages in that fewer confounding effects will be prevalent than in attempts to predict citations across different disciplines and journals. The articles we used for our research spanned over two decades and are drawn from *The Leadership Quarterly*. This journal is cross-listed in the Web of Science in management and applied psychology, and over the last 10 years its average rank has been in the top quintile by impact factor. Furthermore, it has similar research quality standards, insofar as quantitative non-experimental articles are concerned, as do other general journals in the management and applied psychology disciplines, including the *Academy of Management Journal* and *Journal of Applied Psychology* (Antonakis et al., 2010).

Our systematic analysis of the articles in *The Leadership Quarterly* will allow us to first determine how citations are driven by the type of scientific article (e.g., quantitative, qualitative, review, theory). Next, because the majority of the articles published in the journal are quantitative and this mode of inquiry dominates the research landscape in top social sciences journals (cf. Bergh et al., 2006; Swygart-Hobaugh, 2004), we also make more fine-grained analyses to determine how citations are driven by:

1. Data gathering design (e.g., laboratory experiment, field study);
2. Statistical analysis method (e.g., structural equation modeling, ANOVA);
3. Endogeneity (e.g., omitting important causes from a model, having reverse causality).

In addition, we controlled for various confounding factors (e.g., type of leadership theory used, age of article, editor tenure, research records of coauthors) that may also determine citations. Using a broad-based research journal that publishes all types of articles, including theory building articles, quantitative and qualitative articles, reviews, as well as methodological articles allows us to see the evolution of modes of scientific inquiry and their impact on article citations. Thus, through our bibliometric analyses we contribute to the management and applied psychology discipline by bringing to light some predictive links that have not yet been studied.

Our contributions are evident for editors, who might find our recommendations useful in guiding their editorial policy; the types of articles that they publish will determine how often an average article is cited in a journal and hence have a direct effect on the journal "impact factor" and related metrics. Journal contributors will of course benefit too by focusing their efforts on producing articles that journals anticipate will become highly cited and thus can use this knowledge designing their studies and methods. Finally, educators will find our results useful in adapting teaching modules to reflect what we found to be important predictors of citations.

## 2. Determinants of citations

Citations acknowledge the impact an author has had on another's work and can be conceived of as "the currency by which we repay the intellectual debt we owe our predecessors" (Garfield, 1982, p. 624). Citations reflect the substantive relevance that the cited article has for a particular article and are not merely done in passing (Antonakis & Lalive, 2008; Baldi, 1998). Several bibliometric studies have identified specific drivers of citations. We briefly present some of these findings, which provide a basis for the types of information we coded for in the articles we analyzed. Citations have field-specific as well as individual and institutional determinants (Antonakis & Lalive, 2008; Judge et al., 2007); although journal prestige also plays an important role (Aksnes, 2003; Judge et al., 2007), this determinant as well as field differences are irrelevant for our study given that our analyses were only undertaken on one journal. Articles that have higher methodological rigor produce results that are more valid and are cited more often (Bergh et al., 2006). Thus, studies that are based on dominant and well-validated theories as well as best methodological practices should attract more attention from other scholars.

On a very basic level, year of publication is important; ceteris paribus older articles are cited more than younger ones (e.g., Bergh et al., 2006; Colquitt & Zapata-Phelan, 2007; Judge et al., 2007), though changing trends in knowledge contribution may affect citation decay (Macrae, 1969). The number of authors on an article also contributes to citation rates (Aksnes, 2003; Judge et al., 2007). This finding is interesting and could be explained by several reasons including a broader author knowledge resource base, increased possibility of international collaboration, and the distribution of results to a larger coauthors network; there is also the possibility of higher rates of self-citation stemming from more authors, though this explanation does not seem to matter for highly-cited articles (Aksnes, 2003). In addition, the number of cited references in an article (A) may predict future citations from other articles (B) because A becomes more scientifically persuasive by having more citations and because those who are cited in A may reciprocate when they publish B (Gilbert, 1977; Judge et al., 2007); that is, the citation "game" is a repeated one, which could explain the reciprocation (i.e., "you cite me, I cite you"). Of course, reciprocal citations may simply reflect the possibility that A is substantively used and cited in B given that the authors who are cited in A became aware of the paper—by virtue that they were cited in it—and its relevance to their future work.

Location plays a small role, though, it appears that the majority of highly-cited articles are U.S. based (Persson, 2010). Of course, this finding warrants teasing out the explanatory variable because a disproportional amount of highly ranked universities are based in the U.S., which is why we control for averaging ranking of university for each author team of the articles we studied (indeed, our data showed that country of origin explains 19.52% of the variance in university rankings and that out of the 52 countries present in the rankings, U.S. universities dominate the top-10 and top-20 spots and are ranked 8th on average—only universities from Singapore, Hong Kong, The Netherlands, Switzerland, Belgium, the U.K., and Denmark have a higher average rank).

Furthermore, given that citations of authors are affected by the total number of articles an author has published (Antonakis & Lalive, 2008; Egghe & Rousseau, 2006; van Raan, 2006), it is important to control for previously authored articles; the mechanism by which this effect occurs could be similar to that of having more authors in the sense that having more articles provides more visibility to one's work. Having more articles may also increase the likelihood that one or more of the articles becomes highly cited. Another aspect of author reputation could also be captured by previous citations; all things being equal, highly cited authors are more likely to be cited than less cited or unknown authors (Merton, 1968).

Because producing more articles increases citations at the individual level, it is also important to control for the number of articles a journal publishes because of the well-known (inverse) power-law (cf. Newman, 2005) concerning citation distributions across articles; that is, very few articles are highly cited and the vast majority are cited very infrequently. Journals having published more articles, therefore, may increase the likelihood that a subset of them is highly cited (e.g., see Fig. 1, which plots the frequency distribution of citations in the journal we studied, demonstrating this power law). Thus, publishing more articles in a journal may actually help maintain or even raise average citation of articles (and hence the impact factor), particularly if the journal is one of high repute and attracts solid articles already.

### *2.1. Trends in management research*

#### *2.1.1. Theory*

Our field has a "theory fetish" (Hambrick, 2007), which is ironic given the fact that we have weak theories (Pfeffer, 1993). Theories are usually very prominent in empirical articles and researchers in our field who do empirical work generally have to start with a theory; moreover, good journals generally expect articles to make a theoretical contribution. The only established journal dedicated to theory in the management field, the *Academy of Management Review*, is consistently one of the top ranked journals by impact factor. Thus, theory articles should have a premium in the marketplace for citations.

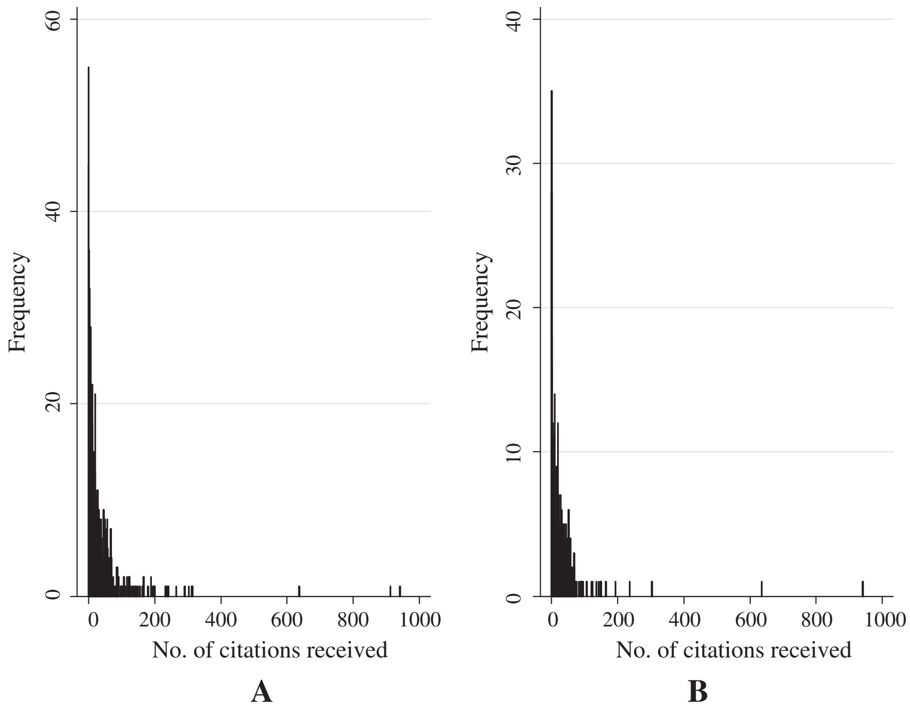
#### *2.1.2. Reviews*

Articles that review the literature tend to be cited more often than are other types of articles (Aksnes, 2003; Judge et al., 2007). Evidently, such articles can help clarify and synthesize the literature and oftentimes provide integrative knowledge on the status of a field and how it should move forward.

#### *2.1.3. Mode of inquiry*

Quantitative methods currently dominate the research landscape (Antonakis et al., 2004; Bergh et al., 2006; Swygart-Hobaugh, 2004). One possible reason for this occurrence is that only quantitative methods can test causal relations. Qualitative research is also employed in the management discipline, but to a much less degree than is quantitative research (Bergh et al., 2006; see also Swygart-Hobaugh, 2004, for parallel findings from sociology). Many reasons could contribute to this occurrence including skepticism of the results of qualitative studies due to idiosyncratic methods employed (Gephart, 2004), which has bearing on the validity and replicability of findings; thus, attempting to publish qualitative research would be relatively more risky than other methods. In addition, the management discipline is rooted in the foundational disciplines of (a) psychology, whose mainstay is experimental methodology, and (b) economics, which has the econometric techniques to model non-experimental data and very strong theory. Thus, qualitative research may not command the attention that would other types of inquiry and we would expect qualitative research to be cited less often than quantitative research (cf. Swygart-Hobaugh, 2004).

Other types of research that journals publish include methodologically-oriented articles and simulations (i.e., Monte Carlo). Journals like *Organizational Research Methods*, which for a young journal has quickly become very highly ranked both in management and applied psychology, seems to be having a growing impact on the field. Method advances help to answer research puzzles in new and robust ways and should thus attract citations from researchers; however, this is not a prolific and fast changing line of research. As for simulations, they are, unfortunately, mostly used by methodologically-oriented researchers and still not well understood by




Note: Panel A, n = 776; Panel B, n = 383; the x-axis refers to the citations an article has received and the y-axis refers to the number of articles that have received a particular citation. These distributions are typical of power law distributions. Interestingly, 7.09% of all articles (Panel A) and 9% of quantitative articles (Panel B) have never been cited; 40.85% of all articles (Panel A) and 41.41% of quantitative (Panel B) have received 10 or fewer citations.

**Fig. 1.** Frequency distribution of citations for all articles (A) and quantitative articles (B).

applied researchers (Antonakis et al., 2010). At this time, they do not seem to attract much scholarly attention. A somewhat related mode of research inquiry is agent-based modeling, which uses simulation techniques to model decisions or outcomes of agents and environments having particular attributes; these methods are also underutilized in management and applied psychology research (Harrison, Lin, Carroll, & Carley, 2007).

#### 2.1.4. “Schools” of leadership

Although there are many theoretical perspectives that attempt to model the leadership phenomenon, certain ones have stood the test of time and others are currently attracting a lot of attention. We list the most important of these schools given that we used them to estimate their impact on citations of quantitative articles. Relying on previous reviews (Bass & Bass, 2008; Gardner, Lowe, Moss, Mahoney, & Cogliser, 2010; House & Aditya, 1997; Lowe & Gardner, 2000; Van Seters & Field, 1990), Day and Antonakis (2012) classified leadership schools into the following parsimonious categories: (a) trait, focusing on stable and personal attributes (e.g., personality) of leaders (Judge, Bono, Ilies, & Gerhardt, 2002; Lord, De Vader, & Alliger, 1986); (b) behavioral, which studies behavioral styles of leaders, usually looking at social support (consideration) or task (initiating structure) orientation (Judge, Piccolo, & Ilies, 2004), or other behavioral aspects of leadership; (c) contextual, which models how context affects the leadership phenomenon (Liden & Antonakis, 2009; Osborn, Hunt, & Jauch, 2002; Porter & McLaughlin, 2006); (d) contingency, which seeks to model how situational demands affect the impact of behavioral styles on outcomes (Fiedler, 1967; House & Mitchell, 1974); (e) relational, which focuses on quality of relations between leaders and followers (Dansereau, Graen, & Haga, 1975; Graen & Uhl-bien, 1995); (f) information processing, which employs a cognitive perspective of leadership (Lord, Foti, & De Vader, 1984; Lord & Maher, 1991); (g) the “new leadership”, which focuses on visionary, values-centered, and charismatic aspects of leadership and related perspectives (Bass, 1985; House, 1977); (h) biological and evolutionary perspectives, which take a genetic, neuroscientific, “hard”-science, or evolutionary approach to leadership (Van Vugt, Hogan, & Kaiser, 2008; Waldman, Balthazard, & Peterson, 2011). They also defined the “skeptics” school, which treats leadership as a social construction (Eden & Leviatan, 1975; Meindl, Ehrlich, & Dukerich, 1985), though they have suggested that this school is mostly subsumed in the information processing perspective. Finally, they have also identified hybrid approaches as a new trend that uses a combination of the above, focuses on process perspectives, and/or does not offer a dominant theory (Lim & Ployhart, 2004; Zaccaro, 2001).

According to Day and Antonakis (2012), the most active schools, which should publish the most and attract the most citations, are currently from the “new”, relational, information processing, and trait perspectives; the contingency and behavioral are the least active. Also, they mentioned that biology and evolutionary perspectives were currently creating much interest.

### 2.2. Trends in quantitative methods

#### 2.2.1. Endogeneity and causality

Quantitative methods can test hypothesized causal relations specified by theory because these methods can establish counterfactual conditions and the researcher can assess the plausibility of estimates within some probabilistic framework (cf. Antonakis et al., 2010; Morgan & Winship, 2007; Pearl, 2000). Properties of estimators usually have known distributions and analytical properties and, when they do not, researchers can use estimators that suspend some assumptions, by using robust estimators, or use some kind of Monte Carlo technique to determine the properties (Antonakis et al., 2010; Mooney, 1997; Muthen & Muthen, 2002; Paxton, Curran, Bollen, Kirby, & Chen, 2001).

When it comes to modeling non-experimental data researchers have to face the "beast," *endogeneity*. In the presence of endogeneity bias, estimates cannot be interpreted. Any observed relationship may be partly or fully driven by omitted variables and the estimated relation could be higher, lower, or of a different sign than the actual causal relation. The confounding effect of endogeneity on estimates is one of the biggest methodological problems in management and psychology research today (Antonakis et al., 2010; Bascle, 2008; Billings & Wroten, 1978; Duncan, Magnusson, & Ludwig, 2004; Gennetian, Magnuson, & Morris, 2008; Hamilton & Nickerson, 2003; Larcker & Rusticus, 2010). It is a problem that has mostly been ignored by researchers in the past and usually plagues non-experimental research (but also experimental research too, for example, when testing mediation in the case of an endogenous mediator). If used correctly, however, causal claims can be made in non-experimental settings, whether using structural equations, two-stage least squares, regression discontinuity, propensity score, or difference-in-differences models (see Antonakis et al., 2010; Cook, Shadish, & Wong, 2008; Shadish & Cook, 1999; Shadish, Cook, & Campbell, 2002, for in-depth discussion).

#### 2.2.2. Hypothesis testing

Quantitative methods are mostly used to test theory and can model complex relations that account for contextual effects, unobserved heterogeneity, non-linear relations, configural hypotheses (i.e., latent class models), growth models, and other types of effects. Some quantitative methods can also be used in an exploratory way where an algorithm is used to find the best-fitting model; however, these algorithms capitalize on chance, making the results difficult to generalize to other settings. Exploratory methods like stepwise regression are increasingly being frowned upon and their use should start to decrease over time (Antonakis & Dietz, 2011; Derksen & Keselman, 1992; Leigh, 1988; Thompson, 1995; Whittingham, Stephens, Bradbury, & Freckleton, 2006). Even methods like exploratory factor analysis are regularly misused because most researchers do not undertake what is referred to as a "parallel analysis"—a Monte Carlo study to determine whether chance (for $k$ observed variables at a particular $n$ size) could explain how many factors are extracted (cf. Fabrigar, Wegener, MacCallum, & Strahan, 1999). Of course, the smaller the $n$–$k$ ratio the more likely the estimator will "find something" due to noise alone. Thus, nowadays, we do not expect to see these methods to be used much by researchers.

*2.2.2.1. ANOVA and regression.* In experimental settings, the method of choice is the workhorse of psychology, ANOVA (Keppel & Wickens, 2004), a special case of regression with dummy independent variables. The use of regression analysis offers many advantages over its ANOVA cousin (or its extensions, ANCOVA, MANOVA, and MANCOVA). Regression is more flexible in how it models relations (e.g., Edwards & Parry, 1993) and it can examine multivariate outcomes, complex hypotheses via testing of slopes or predicted values, as well as their differences (e.g., Lee & Antonakis, 2012). It can also accommodate a wider class of variance estimation procedures. Given that experimental research is firmly anchored in our discipline and that calls have been made to use this method more in leadership research (Brown & Lord, 1999) as well as in social sciences in general (Falk & Heckman, 2009; Shadish & Cook, 2009), we expect to see an increase in the use of both methods, particularly regression analysis, over time (cf. Scandura & Williams, 2000). As for endogeneity bias, ANOVA and regression models that have not used random independent variables will require corrective procedures. Endogeneity can affect estimates in simple models but also in more complex ones. For example, the indirect effect of manipulated factors via a mediator on $y$ cannot be estimated using the "usual" methods (i.e., Baron & Kenny, 1986; Preacher & Hayes, 2004) because the mediator is endogenous. Such models must be estimated using two-stage least squares or a SEM model that accounts for the endogeneity bias (Antonakis et al., 2010; Shaver, 2005).

*2.2.2.2. Structural equation modeling (SEM).* A method that is currently having a growing impact on organizational research is structural equation modeling (cf. Scandura & Williams, 2000). The estimation method used is usually maximum likelihood, a full information estimator. However, other types of estimators are also available like two-stage least squares, which is a limited information estimator that does not spread localized misspecification (Baltagi, 2002; Bollen, Kirby, Curran, Paxton, & Chen, 2007). SEM is particularly useful for cases of simultaneous equations models with observed or latent variables because it can explicitly model and correct for sources of endogeneity bias (Antonakis et al., 2010), in which case they are called instrumental-variable estimators. That is, by exploiting sources of variance from "instruments", or exogenous variables, the causal effects on "downstream" variables can be correctly identified. As with other methods, SEM methods are often misused or not well understood. For example, contrary to popular wisdom, Harman's one factor "test" or modeling an unmeasured latent common-method factor cannot test for such common method bias (Antonakis et al., 2010; Richardson, Simmering, & Sturman, 2009).

*2.2.2.3. PLS.* Partial-least squares analysis has been touted as a viable alternative to SEM and it almost has a cult-like following in marketing and information-system research. However, this estimator suffers from many critical weaknesses (Antonakis et al., 2010; McDonald, 1996; Rönkkö & Evermann, 2013), and we do not expect to see it much in organizational research.

### 2.2.3. Panel (hierarchical or longitudinal) models

These models include both $j$ hierarchical ($i$ leaders nested in $j$ firms) and $j$ longitudinal (observations $i$ of leader $j$ over time) models. They are popular because of their ability to model cross-level effects; that is, how variables at a higher level entity (e.g., leader personality) affect outcomes at entities nested below it (e.g., subordinates of the leader). These models are called random-effects/ coefficients models (or oftentimes HLM models, named for the popular program used to estimate them). However, these models have a major shortcoming that has been repeatedly documented in the literature (Antonakis et al., 2010; Bollen & Brand, 2010; Cameron & Trivedi, 2005; Halaby, 2004; Mundlak, 1978) and calls to correctly test these models have mostly fallen on deaf ears. As mentioned by Halaby (2004, p. 508), "Key principles that ought to routinely inform analysis are at times glossed over or ignored completely." The key problem with this method is an assumption about the panel (i.e., cluster)-specific error term, $u_j$ (which is actually called a "fixed-effect" in the statistics and econometrics vernacular) and is assumed to be orthogonal to the regressors in the random-effects specification (but not when using a fixed-effects estimator). A violation of this assumption, which is testable (Hausman, 1978), causes serious endogeneity bias in estimates.

*2.2.3.1. WABA.* Within- and between-analysis is an ANOVA-type procedure that is used to evaluate the level at which relations reside (Dansereau, Alutto, & Yamarino, 1984). It is useful to justify data aggregation and can also be used in a regression to model cross-level effects (Castro, 2002; Schriesheim, 1995; Yammarino, 1998). WABA has not been adopted as a method of choice for multilevel research, probably because of the rise of random-effects modeling methods as well as other ways to justify aggregation (Bliese, 2000; Castro, 2002). The use of WABA thus appears to have been restricted mostly to a core group of methodologists and it might not have much impact on quantitative research.

### 2.2.4. Design, mode of data gathering, and samples

We briefly look at some trends in the field of management with respect to study design, how data are gathered and what types of samples are used. In terms of the design of the study, Scandura and Williams (2000) reported that the use of data from field studies is generally on the rise and that of laboratory studies on the decline; the latter finding may be less relevant to what we may find in our study given that many researchers trained in psychology publish in *The Leadership Quarterly*, which is mostly a micro-level oriented journal (see also Brown & Lord, 1999). In addition, it is surprising to see that laboratory studies are not being used so much in the premier empirical management journal, the *Academy of Management Journal*, given that experiments are needed for clear causal claims (Falk & Heckman, 2009); in addition, laboratory studies yield similar results to those in the field (Anderson, Lindsay, & Bushman, 1999), particularly results coming from industrial–organizational psychology (Mitchell, 2012). We surmise that the driver for more field data is probably a concern for greater external validity. Indeed, it appears that data from cross-sectional studies, as well as samples from the private sector are on the increase too; moreover, single-source studies are also on the rise, whereas multiple source studies are not (Scandura & Williams, 2000). Paradoxically, cross-sectional and single-source studies are the exact type of studies that are likely to engender endogeneity bias (Antonakis et al., 2010).

## 3. Method

We gathered all bibliometric data from Scopus. We used this database because articles from *The Leadership Quarterly* were only indexed in the Web of Science from 1994, whereas the journal began publishing articles in 1990. Citation data on the journal is, however, included in Scopus from the inception of the journal. In addition, Scopus has wider coverage of journals (Meho & Yang, 2007), which means that we have a broader measure of citation impact. Also, Scopus has an efficient author finder by grouping articles attributed to a particular author name to a common affiliation (which improved the accuracy of citation and publication data at the author level). It should be noted that full cited reference data on authors is only available from 1996 onwards in Scopus, which thus affects all authors equally in terms of number of articles published and citations received. The coverage prior to 1996 is partial. Although this is a limitation, because we used author citations and number of publications as covariates, it should not affect results in a meaningful way because prolific and highly cited authors will continue to have more citations and articles after 1996 as compared to their peers who have made more modest contributions to science. Thus, for all author-level bibliometric data for the period prior to 1996, we used what data Scopus reported for the authors up to 1996. Note that we covered all types of articles except those whose content was editorial introductions, calls for articles, errata, or other announcements. Articles were the unit of analysis.

### 3.1. Dependent variables

1. *Lifetime article citations*: We downloaded data on article citations on 14 July 2013 using Scopus for all articles published between 1990 and 2012.
2. *Per year article citations*: Although we expect this measure to be strongly related to lifetime citations, we chose to also predict average article citations received per year because it is a more current measure of article influence (i.e., it factors in current decay or growth in citations).
3. *Endogeneity*: For quantitative articles, we coded whether an article had endogeneity bias based on any of the 12 threats to estimate consistency identified by Antonakis et al. (2010), which are listed in Table 1.

### 3.2. Independent variables

Below we describe the independent variables (regressors) we used, as gleaned from the literature and as also required by the exigencies of our study design; that is, we also identified coding categories after each coder read, in-depth, 30 randomly selected articles. We distinguish regressors (marked with an asterisk, *) that we used when analyzing all articles including the quantitative article subset, from those that were only relevant as regressors for quantitative articles (not marked with an asterisk). The variable marked with a plus sign (+) was only included as a regressor when the sample included all types of articles (and was not pertinent for when we analyzed quantitative articles). Regressors that we coded included the following:

1. +*Article type*. We used $k - 1$ dummy variables to model article type, including quantitative, qualitative, theory, review, commentary/discussion, methodological, and agent-based simulation articles. Note, if an article used both qualitative and quantitative data, and analyzed data quantitatively, we coded the article as quantitative (see footnote 3 in Bansal & Corley, 2011); we included meta-analyses in quantitative; if articles made contributions both to theory and empirically, either quantitatively or qualitatively, they were coded as quantitative or qualitative; if articles were both a review and also made propositions for a theoretical contribution, we coded them as theory. We coded articles that explained methods or made core methodological contributions as methodological (and those that reviewed methods as reviews).
2. **Coder*. Because we used two coders, we controlled for the possibility of coder effects using a dummy variable.
3. **Age of article*. This variable was simply the chronological age of the article (i.e., 2013 minus year of publication).
4. **Number of cited references*. We tallied the number of cited references in the article because reported in the reference list (or footnoted in some cases).
5. **Senior editor*. Because editors may have preferences for certain types of articles (or may create that expectation in submitting authors), and because editorial tenure may correlate with other variables too (e.g., citations, number of articles published), we used $k - 1$ dummy variables to capture these effects (because Robert J. House and Henry L. Tosi co-served as senior editors, we coded them together as one editor).
6. **Number of authors*. The total number of coauthors listed in the by-line of the article.
7. **Lagged author citations*. For all authors, we obtained data on the total citations they received prior to the year in which the particular article was published. For example, for an author who published an article in 2010, we measured the total citations they had received up to 2009. We averaged the citations of all coauthors for a particular article to reflect how well the authors were collectively cited; doing so is more accurate than just coding information on the first author only or the most senior author only. Given that full author citation data is only available after 1996, this variable is more accurately portrayed as a partial lag.
8. **Lagged author publications*. Similar to the above, we tallied the number of articles authors had published and averaged this number for author teams.
9. **Rank of university affiliation*. We used data from a reputed world ranking source, averaged for the periods between 2008 and 2012 (QS World University Ranking, 2008–2012); this ranking is a weighted composite consisting of academic reputation assessment, faculty to student ratio, citations per faculty member, employer reputation assessment, and proportion of international faculty and students. The rankings are very stable over time, ICC1 = .95, SE = 0.004, 95% CI from .94 to .95, ICC2 (i.e., reliability of the mean) = .98. Thus, we used the average ranking of a university. Data on different metrics are available going back to 2004, when this ranking system was combined with the THES ranking. However, given the fact that rankings over time are very stable and that it is unlikely that these rankings would change much over time (even going back to 1990), we used the average score over the 2008–2012 period. Thus, for each author, we used the average QS ranking of the author's university; we then average the scores across coauthors for each article to obtain a collective score.

**Table 1**
The 12 threats to validity of quantitative research.
Adapted from Antonakis et al. (2010).

1. Omitted regressors[a]
2. Omitted fixed effects in multilevel models[a]
3. Using random-effects without ensuring consistency vis-à-vis a fixed-effects estimation[a]
4. Independent variables not exogenous[a]
5. Comparing groups to which observations have not been randomly assigned
6. Comparing groups where selection to group is endogenous[a]
7. Participation is self-selected or sample is non-representative
8. Reverse causality[a]
9. Measurement error in the independent variables[a]
10. Common-method variance[a]
11. Misspecified mediation models[a]
12. Using a full information estimator without checking its consistency vis-à-vis a limited information estimator (when possible)[a]

Note: In addition to estimate consistency, which will be compromised by the 12 threats above, consistency of inference (i.e., the estimation of standard errors) must also be assured by using (a) robust estimates of the variance when residuals are heteroskedastic (in regression models) or data are not multivariate normally distributed (in SEM models) and (b) cluster-robust estimates of the variance for panel/multilevel data.

[a] These are threats that affected more than 70% of the articles analyzed across seven journals (Academy of Management Journal, Journal of Applied Psychology, Journal of Management, Journal of Organizational Behavior, The Leadership Quarterly, Organizational Behavior & Human Decision Processes, and Personnel Psychology)—see Appendix in Antonakis et al. (2010).

10. **Number of articles published per year*. We measured how many articles were published per year in the journal for two reasons: (a) the number of articles published per year is not constant (and there has been a general trend to publish more articles over time), and (b) we wanted to establish, as we suggested in the literature review, whether publishing more articles per year increased average journal citations (due to increasing the likelihood that one or more articles becomes highly cited). Given that fewer articles were published in the past, we also interacted number of articles published with age of the article to determine whether citation trends differ over time as a function of articles published per year.
11. **Issue number*. We used $k-1$ dummy variables to control for unobserved heterogeneity due to issue number because of two reasons: (a) articles published near the beginning of the year have increased exposure over those articles published near the end of the year, and (b) the yearly review issue is usually published in the last issue of the year.
12. *School of leadership*. We used the following nine categories (modeled as $k-1$ dummy variables): trait, behavioral, contextual, contingency, relational, information processing (we combined skeptics with information processing, given that the former only included three articles), new leadership (including transformational, charismatic, ethical, authentic, and visionary), biological and evolutionary, and hybrid (we coded articles into this category if [a] it included more than one theoretical perspective that did not dominate in the article, or [b] the theoretical perspective did not belong to one of the aforementioned theories).
13. *Statistical analysis method used for hypothesis testing*. Given than an article may have used more than one method, we used the following independent categories, each modeled separately as one dummy variable: correlation analysis, analysis of variance (including analysis of variance, covariance and their multivariate counterparts), regression, structural equation modeling (including confirmatory factor analysis, two-stage least squares and path models), exploratory factor analysis (EFA), random-effects/coefficients models, within and between analysis (WABA), partial least-squares (PLS) analysis, and others (none of the above).
14. *Number of studies*. We used the following five categories (modeled as $k-1$ dummy variables): one, two, three, four or more studies, or not applicable.
15. *Sample location*. Given that an article may have had data from several locations, we used the following independent categories, each modeled separately as one dummy variable: US, Europe, Asia, cross-national, others, or not applicable.
16. *Study design*. Given that an article may have used more than one data collection method, we used the following independent categories, each modeled separately as one dummy variable: field survey, laboratory experiment, field experiment, quasi-experiment, archival data, meta-analysis, interview or others.
17. *Data source*. Because articles may have used more than one type of data source, we used the following independent categories, each modeled separately as one dummy variable: one source, two or more subjective sources, one or more objective source, or not applicable.
18. *Temporal context of study*. Given that an article may have had more than one temporal context we used the following independent categories, each modeled separately as one dummy variable: cross-sectional, two time periods (non-repeated measurement of variables, e.g., independent variable obtained at time one and dependent variable obtained at time two), longitudinal (repeated measurement of the same variables), or not applicable.
19. *Type of scale/s used*. Given than an article may have used more than one scale, we used the following independent categories, each modeled separately as one dummy variable: new scale, original scale (including only changing the referent), modified scale (e.g., deleting items, wording changes, measuring at a different level), or not applicable.

### 3.3. Other variables

We report descriptive trends for other variables measured, which we did not use as regressors because they were not pertinent in predicting outcomes (either because of little variance in the measures or because they were not relevant to most studies). These additional categories included: the statistical method employed (not necessarily to test hypotheses), occupation of study subjects, method of data aggregation used, and level of analysis of observation units.

### 3.4. Reliability of coding

After the coding manual was developed, two coders, the second and third authors of the current article, independently coded 30 articles, which included 525 coding events. Expected agreement due to chance would have been 19.48%; however, the coders agreed on 71.62% of events and the agreement statistic, $\kappa = .65$, $SE = .02$, $z = 34.09$, $p < .001$, indicating that it was significantly better than chance. This level of agreement can be labeled as “substantial” (Landis & Koch, 1977).

After discussing coding disagreements and refining the coding manual, the coders independently coded five more articles, which included 103 coding events. Expected agreement due to chance would have produced 16.83% agreement. The coders, however, agreed on 88.35% of coding events: $\kappa = .86$, $SE = .04$, $z = 21.11$, $p < .001$, which was significantly better than chance agreement, and which can be qualified as “almost perfect” (Landis & Koch, 1977).

The coders then independently coded the rest of the articles: One coder coded articles published in even years and the other coder did the odd years. Coders discussed how to code particular categories of articles if they were unsure, or sought the senior (first and fourth) authors' advice.

### 3.5. Estimation strategy

We estimated all models with Stata version 13 (StataCorp, 2013). For the models predicting citations, our dependent variable is a count, requiring a Poisson-type modeling procedure. However, there was also a preponderance of zeros and the data were highly dispersed (see Fig. 1) with variances much higher than the means, both for total citations (mean = 33.47, SD = 66.14, variance = 4374.06) and citations per year (mean = 3.74, SD = 4.84, variance = 23.39). We therefore used the most appropriate model—a zero-inflated negative binomial regression model (Long & Freese, 2006). Because a major reason why many articles are uncited is due to their age, particularly articles that have been published very recently and have not had the time to be cited, we used article age to predict the zero inflation. To ensure correct inference, we estimated all models with robust standard errors.

For the model predicting endogeneity, we used probit regression (Aldrich & Nelson, 1984). We also explored whether endogeneity affects citations; we surmised that given the collective knowledge of the academic market (as expressed in the citations an article receives), articles that have endogeneity problems might be less cited than those that do not. Ironically, because having endogeneity might itself be endogenous—that is, there may be omitted variables that predict both citations and endogeneity, or because of measurement error (i.e., miscodings by the coders)—we estimated a structural two-stage model (Cameron & Trivedi, 2009). In the first step, we used a linear probability model (Angrist, 2001) from which we saved the residuals as a variable called "endogeneity residual" (Cameron & Trivedi, 2009). Assuming correct specification, this variable contains information from the disturbance that may be correlated with the disturbance of the citations equation. We then included the endogeneity residual as a control variable in the second stage model to predict citations. This procedure is called "augmented regression" and (a) eliminates endogeneity bias from the citations equation and (b) is an alternative way to undertake a Hausman endogeneity test (Davidson & MacKinnon, 1993). We also compared our results to an instrumental variable Poisson model using two-stage generalized method of moments estimation (which is not the most efficient estimator given our data, but which is consistent).

For our instrumental-variable specification we used instruments that reflected the collective "research wisdom" of the author team. The instruments included number of authors, author citations, author articles, and author affiliation rank, which are theoretically exogenous (i.e., any "shocks" to the dependent variables from the disturbance cannot possibly make the instruments vary, which are fixed in time). We thus surmise that collective author wisdom determines whether an article has endogeneity bias or not and that this in turn will affect whether an article is cited. We excluded the instruments from the citation equation and tested for the exclusion restriction with an over identification statistic, the Hansen J test (Hansen, 1982), as implemented in Stata's instrumental-variable Poisson model. As a check, we also computed this statistic manually by regressing the deviance residuals from a generalized linear model with a negative binomial link (McCullagh & Nelder, 1989) on the excluded instruments. If the model is misspecified and the excluded instruments correlate with the residuals of citations (i.e., if citations are driven by unobserved variance that correlates with the excluded instruments), then the model will fail the overidentification test.

## 4. Results

For descriptive statistics and correlations between key variables, see Table 2. The types of article published were distributed as follows: Quantitative = 49.48%, theory = 22.42%, review = 11.47%, qualitative = 10.70%, commentary/discussion = 3.48%, methodological = 2.06%, and agent-based modeling = .39%.

Note that for all count models reported, article age significantly predicted the inflation factor. Furthermore, the natural log of the dispersion parameter, $\ln(\alpha)$, was significant indicating that the zero inflated negative binomial model was the most appropriate model to use for these data.

### 4.1. Predicting citations of all types of articles

See Table 3 for full regression estimates.

We estimated the models in a hierarchical (nested) manner to determine how adding sets of predictors changed the pseudo $R$-square. For example, in the first model, we included number of authors, article age and the controls as regressors; the Wald test for the coefficients equaling zero was significant, indicating that these variables significantly predicted the dependent variable with an $R$-square of .60. In model 2, we added the categorical predictors indicating article type. These added a significant increment (.01) to the $R$-square.

The final model (Model 6) explained 66.58% of the variance in citations (we report results for this full model only). Given the high correlation between total citations and citations per year ($r = .88$), it was not surprising that we found similar results when predicting citations per year, as indicated by the results for Model 7. As regards predicting total citations, results indicated that quantitative, theory, review, and method articles received significantly more citations than did qualitative articles (qualitative articles were the base category to which we compared the other categories). The coefficients of the article types were significantly different from each other, Wald $\chi^2(5) = 18.74$, $p < .01$; univariate Wald tests, with a Bonferroni adjustment for multiple tests showed that the coefficient of agent-based simulations was significantly lower than the other coefficients but not significantly lower than those of qualitative and commentary. Furthermore, review articles were significantly more cited than were commentaries. See predicted mean citation rates by article type in Fig. 2A.

**Table 2**
Correlation matrix among key variables.

| | Mean | SD | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | Mean | SD |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Total cites | 26.24 | 56.19 | | .88 | .27 | .00 | −.14 | −.06 | .03 | .10 | | | | | | | | | | 33.47 | 66.14 |
| 2. Cites per year | 3.24 | 3.60 | .83 | | .05 | .06 | −.04 | .01 | .05 | .21 | | | | | | | | | | 3.74 | 4.84 |
| 3. Article age | 7.22 | 6.06 | .35 | .13 | | −.24 | −.43 | −.28 | .03 | −.31 | | | | | | | | | | 8.53 | 6.4 |
| 4. No. of authors | 3.04 | 1.38 | .02 | .02 | −.08 | | .15 | .12 | −.09 | .18 | | | | | | | | | | 2.63 | 1.39 |
| 5. Authors' cites | 292.88 | 383.23 | −.19 | −.09 | −.51 | .08 | | .69 | .12 | .11 | | | | | | | | | | 243.44 | 403.66 |
| 6. Authors' articles | 18.78 | 21.48 | −.12 | −.05 | −.33 | .02 | .65 | | .10 | .08 | | | | | | | | | | 16.74 | 20.33 |
| 7. Affiliation rank | 121.69 | 112.28 | .03 | .08 | −.02 | −.04 | .14 | .12 | | −.05 | | | | | | | | | | 114.93 | 120.48 |
| 8. Cited references | 68.59 | 25.15 | −.08 | −.02 | −.46 | .05 | .19 | .14 | −.02 | | | | | | | | | | | 75.88 | 41.7 |
| 9. Correlation analysis | .06 | .25 | .01 | .01 | .12 | .08 | .06 | .05 | −.08 | −.09 | | | | | | | | | | | |
| 10. ANOVA | .29 | .46 | −.02 | −.03 | .07 | −.02 | .03 | .04 | .05 | −.09 | .07 | | | | | | | | | | |
| 11. Regression | .41 | .49 | −.07 | −.02 | −.04 | −.07 | −.07 | −.03 | .06 | .01 | .00 | −.24 | | | | | | | | | |
| 12. Exploratory factor analysis | .06 | .23 | .03 | .03 | .09 | −.04 | .00 | −.04 | .08 | −.03 | .13 | −.03 | −.09 | | | | | | | | |
| 13. Confirmatory factor analysis | .05 | .22 | .07 | .11 | .00 | −.02 | −.01 | −.01 | .06 | .00 | .13 | −.02 | −.08 | .36 | | | | | | | |
| 14. Structural equation modeling | .17 | .37 | .11 | .09 | −.10 | .07 | .02 | −.02 | −.02 | .16 | −.09 | −.24 | −.27 | −.02 | .15 | | | | | | |
| 15. Random-effects regression | .14 | .35 | −.08 | −.01 | −.25 | −.04 | .24 | .18 | .06 | .12 | −.11 | −.23 | −.18 | .00 | −.06 | −.08 | | | | | |
| 16. Within and between analysis | .04 | .20 | −.01 | −.01 | .09 | .06 | −.05 | −.02 | −.03 | −.06 | .00 | −.04 | −.03 | −.05 | −.05 | −.09 | −.04 | | | | |
| 17. Partial least squares analysis | .04 | .19 | .05 | .11 | .02 | −.08 | −.03 | −.01 | −.04 | −.01 | −.05 | −.10 | −.14 | .01 | −.05 | −.09 | −.08 | .03 | | | |
| 18. Endogeneity | .80 | .40 | .07 | .07 | .04 | .00 | −.22 | −.13 | −.06 | −.02 | −.11 | −.46 | .14 | −.02 | .00 | .17 | .17 | .07 | .06 | | |

Note: the correlation matrix above the diagonal refers to all articles (n = 776); the correlation matrix below the diagonal refers to quantitative articles only (listwise n = 371; we have, however, observations for 384 quantitative articles, the difference being that for some articles direct endogeneity threats were irrelevant); for n-size = 371 and for $r > |.11|$, $p < .05$; $r > |.14|$, $p < .01$; $r > |.17|$, $p < .001$.

**Table 3**
Predicting total citation rates for all types of articles.

| (Model) | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| # of authors | .07* | .06* | .06* | .08** | .07** | .09** | .07* |
| | (2.46) | (2.35) | (2.30) | (2.72) | (2.59) | (3.18) | (2.57) |
| Artic. age (a) | .21** | .21** | .21** | .21** | .07 | −.05 | −.09** |
| | (8.43) | (8.66) | (8.66) | (8.44) | (1.78) | (1.38) | (2.83) |
| Cited references | .01** | .01** | .01** | .01** | .01** | .01** | .01** |
| | (5.08) | (4.49) | (4.58) | (4.64) | (4.75) | (5.47) | (4.98) |
| Quantitative | | .38** | .37** | .38** | .38** | .35** | .36** |
| | | (3.10) | (2.92) | (3.02) | (2.99) | (2.95) | (3.25) |
| Theory | | .36** | .34* | .37** | .37** | .33** | .38** |
| | | (2.64) | (2.50) | (2.73) | (2.69) | (2.59) | (3.13) |
| Review | | .55** | .52** | .54** | .56** | .52** | .60** |
| | | (2.81) | (2.68) | (2.70) | (2.74) | (2.78) | (3.25) |
| Comment/discuss. | | −.19 | −.21 | −.18 | −.11 | −.12 | −.01 |
| | | (.81) | (.89) | (.74) | (.46) | (.57) | (.05) |
| Method | | .55 | .54 | .57* | .52* | .58* | .48 |
| | | (1.83) | (1.80) | (2.00) | (2.15) | (2.30) | (1.84) |
| Agent simulation | | −.67** | −.65** | −.54** | −.48** | −.51 | −.55* |
| | | (4.53) | (4.41) | (3.66) | (2.76) | (1.70) | (2.11) |
| Author cites | | | .00 | .00 | .00 | .00 | .00 |
| | | | (.88) | (.68) | (.92) | (1.74) | (1.66) |
| Author articles | | | −.00 | −.00 | −.00 | −.00 | −.00 |
| | | | (.10) | (.16) | (.17) | (.86) | (.75) |
| Aff. rank[a] | | | | .00* | .00* | .00* | .00* |
| | | | | (2.22) | (2.31) | (2.08) | (1.97) |
| # articles pub (b)[b] | | | | | −.03** | −.04** | −.03** |
| | | | | | (4.96) | (8.62) | (5.29) |
| (a) × (b) | | | | | | .01** | .01** |
| | | | | | | (7.15) | (5.58) |
| Controls[c] | Incl. | Incl. | Incl. | Incl. | Incl. | Incl. | Incl. |
| Constant | −.35 | −.61** | −.68** | −.84** | 1.60** | 1.89** | .49 |
| | (1.68) | (2.67) | (2.93) | (3.67) | (2.83) | (4.01) | (1.07) |
| Pseudo *R*-sq | .60 | .61 | .61 | .61 | .62 | .67 | .42 |
| Δ*R*-square | | .01 | .00 | .00 | .01 | .04 | |
| Wald test[d] | 831.10** | 138.72** | 1.20 | 4.92* | 24.62** | 51.18** | 482.78** |

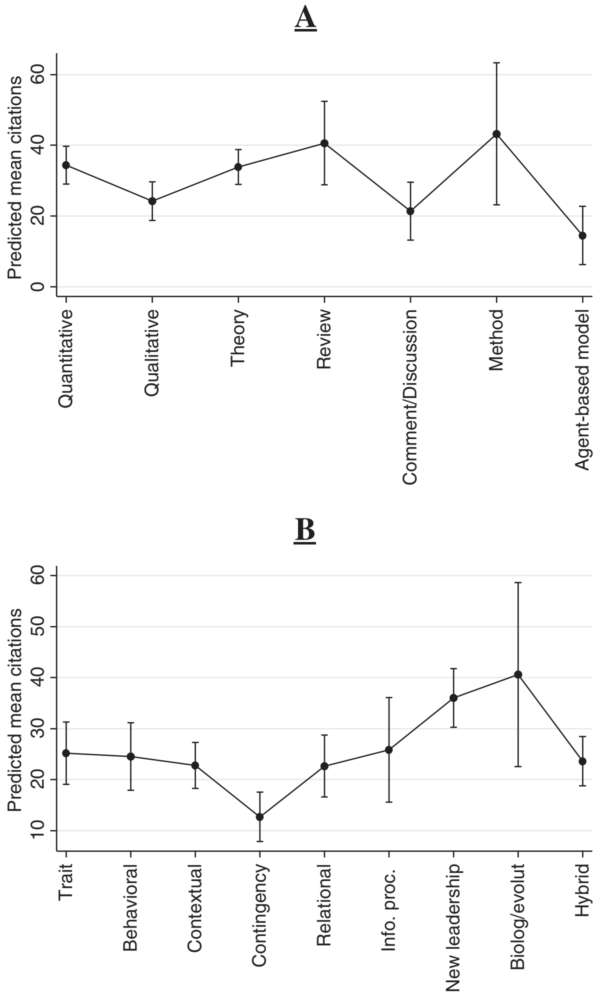
Note: $n = 776$; estimates are unstandardized; pseudo *R*-square is based on the Cox and Snell (1989) method; Models 1–6 predict total citations; Model 7 predicts citations per year; omitted category for article type is qualitative; robust *z*-statistics in parentheses.
[a] Reverse coded thus a higher number indicates higher rank.
[b] Indicates the total number of articles published in the year in which the article appeared.
[c] Includes dummy variables for issues, editor and coder.
[d] Wald test for additional regressors from previous model (for Models 1 and 7, it is the Wald test for the full equation).
** $p < .01$.
* $p < .05$.

Having more authors, more cited references, and a higher university ranking predicted more citations. As for the interaction between article age and number of articles published, it was significant. Given the non-linear nature of the regression model used, we plotted this interaction to better understand the pattern of relations (Fig. 3A). We used practically meaningful values of article age and number of articles published that were representative of mid-range values in the data.

As Fig. 3A indicates, the slopes become increasingly different at increasing values of article age. For example, the predicted value for articles that were nine years old was significantly different if the journal had published 30 articles (predicted citations 49.03) versus 40 articles (predicted citations 71.02) versus 50 articles (predicted citations 102.86): $\chi^2(2) = 122.57$, $p < .001$ (all pairwise comparisons significant). In addition, the derivative with respect to age at nine years indicated that the slopes were significantly higher as more articles were published: $\beta_{30articles} = 10.96$, $\beta_{40articles} = 22.28$, $\beta_{30articles} = 41.56$: $\chi^2(2) = 58.98$, $p < .001$ (all pairwise comparisons significant). These results suggest that journals publishing more articles per year increase the average citation of articles over time.

### 4.1.1. Robustness checks

One possible reason why certain types of articles were cited significantly less (or more) is because the journal publishes fewer (or more) articles of that type or that the research community is smaller (or larger). Whatever the case, for each type of article we tested whether the predicted percentage of total citations received differed from the observed percentage of articles published. Results indicated that predicted citations received by article type, as a percentage of citations to all articles, were (in parentheses is the percentage of articles published of that type): quantitative = 50.73% (49.48%), qualitative = 7.72% (10.70%), theory = 22.63% (22.42%), review = 13.88% (11.47%), commentaries = 2.22% (3.48%), methods = 2.66% (2.06%) and agent-based simulations = .17%




Note: For Panel A, observed mean citations by article type were: Quantitative = 28.64, Qualitative = 18.42, Theory = 42.64, Review = 57.93, Commentary/discussion = 16.59, Method = 25.00, Agent-based simulation = 7.67. For Panel B, observed mean citations by school of leadership were: Trait = 31.53, Behavioral = 22.73, Contextual = 29.28, Contingency = 16.07, Relational = 14.29, Information processing = 13.20, New leadership = 45.20, Biological/evolutionary =13, Hybrid = 20.59. Error bands are 95% confidence intervals.

**Fig. 2.** Predicted mean citation rates for all articles by article type (Panel A) and for quantitative articles by school of leadership (Panel B).

(.39%). Wald tests, with Bonferroni corrections indicated that the following three article types had a significantly different share of citations to share of articles published:

1. Qualitative articles, $\chi^2(1) = 11.10$, $p < .01$, which were undercited by 38.48% (i.e., these articles make up 10.70% of all articles, yet they received 7.72% of total citations);
2. Commentary–discussion articles, $\chi^2(1) = 8.41$, $p < .01$, which were undercited by 56.77% (i.e., these articles make up 3.48% of all articles, yet they received 2.22% of total citations);
3. Agent-based simulation articles, $\chi^2(1) = 20.65$, $p < .001$, which were undercited by 131.69% (i.e., these articles make up .39% of all articles, yet they received .17% of total citations).

Another more practically-meaningful way to understand the findings with respect to how different article types predict citations is to analyze the composition of the top-cited articles. We ranked articles on the current measure of impact (citations per year) and used the top-40 articles as the cut-off (note that we redo these results for total citations). Of course, we realize that with this exercise we are "sampling on the dependent variable" (Denrell, 2003, 2005) and that finding patterns in a skewed sample does not mean that these patterns distinguish high from low cited articles. However, in the regression analyses we predicted the full range of citations; we merely use the top-40 analysis to focus on the highly-cited side of the spectrum to see whether results here agree with the regression results (and it is possible that we would find some interesting outliers contradicting the general trends).

Results showed that the top-40 articles were composed of 35% quantitative, 32.50% theory, 30% review, and 2.5% commentary articles. Binomial probability tests (two-tailed) showed that with respect to their relative share of articles published, quantitative,

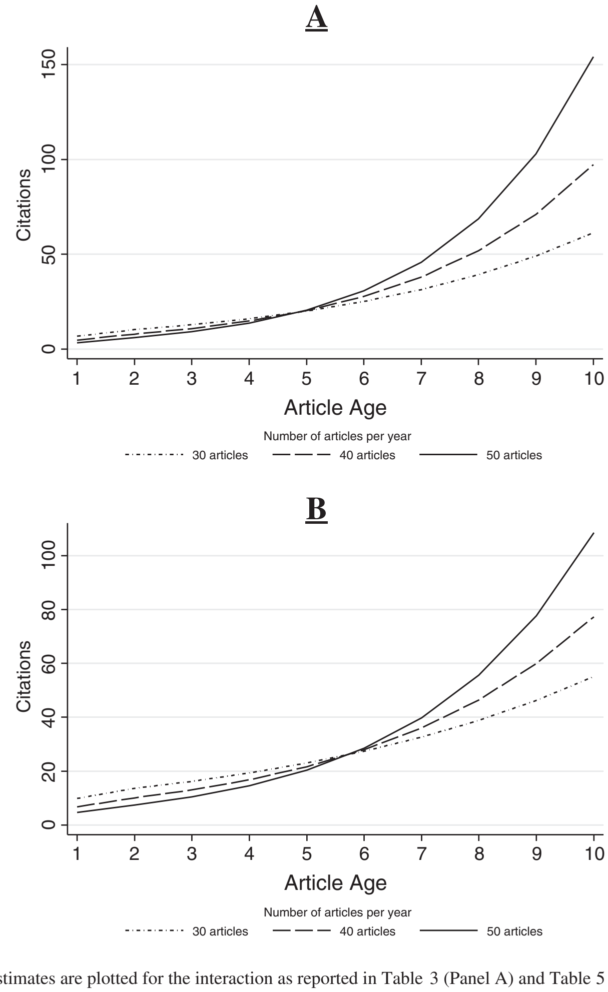



**Fig. 3.** Predicting total citation rates as function of article age and number of articles published per year for all articles (Panel A) and for quantitative articles (Panel B). Note: Estimates are plotted for the interaction as reported in Table 3 (Panel A) and Table 5 (Panel B).

theory, and commentary articles are appropriately represented in the top forty ($p$'s for binomial test >.05). However, review articles were overrepresented ($p < .01$). Qualitative articles, which constituted the next big category of articles after quantitative, theory, and reviewer articles, were significantly underrepresented ($p < .05$); that is, we would have expected to see at least four qualitative articles in the top-40 if the proportion of qualitative articles as a percentage of total articles was the same as that in the top-40. No qualitative articles appeared in the top-50; qualitative articles only began to appear in the top-60 (2 articles). Finally, methods, and agent-based simulation articles were not significantly underrepresented in the top-40.

We also compared the composition of *The Leadership Quarterly* top-40 to the top-40 of two high prestige general journals: *The Academy of Management Journal* (AMJ), which does not publish theory articles but does have a tradition of publishing qualitative research (out of the 27 best article awards given since 1986, 9 articles or 33.33%were qualitative and 18 or 66.67% were quantitative); and the *Journal of Applied Psychology* (JAP), which publishes all types of articles. For these two journals, we used data from the Web of Science (given that Scopus does not include citation information on some journals from afar back).

We first report the top-40 for *The Leadership Quarterly* (LQ) using total citations (and not citations per year)—we used this measure because it is more stable over time and to enable us to have a common comparison of journal lifetime citations that would be unaffected by current trends (which may differentially affect journals). For LQ the top-40 all-time cited articles were composed of the following article types: Quantitative = 32.5%, theory = 40%, and review = 27.5%. For AMJ the types were: Quantitative = 80% (a binomial probability test showed that this type was appropriately represented as compared to proportion of articles making up the best articles awards), qualitative = 10% (a binomial probability test showed that this type was significantly under-represented as compared to proportion of articles making up the best articles awards), review = 2.5%, commentary = 5% and method = 2.5%. Note that for JAP, the composition was: Quantitative = 72.5%, method = 17.5%, review = 7.5% and commentary = 2.5%. Without coding all articles from both JAP and AMJ to determine the proportion of all article types, we are unable to make any meaningful statistical comparisons here. The patterns though, across the three journals are telling, particularly with respect to the most and least present type of article across the three journals. That is, overall the types consisted of: Quantitative = 61.7%, theory = 13.3%, review = 12.5%, method = 6.7%, qualitative = 3.3%, and commentary = 2.5%.

### 4.2. Predicting citations of quantitative articles

We report the results of these estimations in Table 5.

The final model (Model 8) explained 80.64% of the variance in citations. Turning to school of leadership first, results indicated that with respect to the contingency school, all other schools received significantly more citations. The coefficients of the article types were significantly different from each other, the Wald $\chi^2(7) = 27.48$, $p < .001$; the univariate Wald tests, with a Bonferroni adjustment for multiple testing, showed that the coefficient of new leadership was significantly higher than the coefficients of contextual, relational leadership, and hybrid. See the predicted mean citation rates by school in Fig. 2B. Following the previous findings for all articles, the interaction between article age and number of articles published was significant (Fig. 3B), confirming that publishing more articles increases the average citation of articles over time.

Other positive predictors of citations included using random-effects models, structural equation modeling and meta-analysis; using archival data was negatively predictive of citations. From the rest of the variables we measured, cited references, and having two (rather than one) studies predicted more citations. We also found very similar results when predicting citations per year, as indicated by the results of Model 9 in Table 4.

### 4.3. Predicting endogeneity

It is alarming to note that 79.73% of articles had one or endogeneity threats (note that we did not code the meta-analyses for endogeneity because this would have required coding all articles included in each meta-analysis). A full 96.23% (51/53) of articles that used random-effects models had an endogeneity threat as did 86.36% (133/153) of articles that used regression and 95.24% (60/63) of articles using SEM.

As regards the probit model predicting endogeneity (Table 6), it fit the data very well (Hosmer–Lemeshow $\chi^2(4) = 2.39$, $p = .66$, for 6 quantiles). Notable negative predictors (that reduced endogeneity) that were significant included author citations (standardized $\beta = -.47$), number of authors (standardized $\beta = -.27$), the biological/evolutionary (standardized $\beta = -.28$), new leadership (standardized $\beta = -.34$), behavioral (standardized $\beta = -.10$) and contextual (standardized $\beta = -.16$) schools and having two, three or more studies (in contrast to 1 study), using an original scale (standardized $\beta = -.14$), using laboratory data (standardized $\beta = -.30$), in addition to other significant variables controls.

Significant positive predictors included using regression analysis (standardized $\beta = .10$), random-effects models (standardized $\beta = .24$), structural equation modeling (standardized $\beta = .36$), field data (standardized $\beta = .34$), having more authored papers (standardized $\beta = .24$), archival data (standardized $\beta = .14$), and using new scales (standardized $\beta = .11$), among other predictors. For an idea of the effect size of a change of author team citations on the probability of having endogeneity, refer to Fig. 4.

### 4.4. Does endogeneity affect citations?

In the two-stage instrumental variable model, we used number of authors, author citations, author articles, and author affiliation rank as instruments of the endogeneity dummy variable; we used all other regressors and controls as included instruments to predict both endogeneity and citations. The instruments were not as strong as we would have hoped, that is, the weak identification test of excluded instruments, $\chi^2(4) = 29.50$, $p < .001$, gave an *F*-test equivalent of 7.38 (whereas a value above 10 is considered more appropriate, see Staiger & Stock, 1997); thus, in case estimates were biased, we re-estimated all models using the two "strong" instruments, author citations and number of authors. This test of excluded instruments, $\chi^2(2) = 28.62$, $p < .001$ gave an *F*-test equivalent of 14.31. The structural estimates did not change much though this estimator was slightly more efficient (see Table 6). Given that estimates were so similar, we report results from the model using the theoretically-identified instruments (i.e., Models 2 & 3) so as to reduce the possibility of capitalizing on chance.

Results indicated that the coefficient of endogeneity was a negative predictor of total citations ($\beta = -2.09$, $p < .01$) and citations per year ($\beta = -2.24$, $p < .01$); this result was confirmed using Stata's standard instrumental-variable Poisson estimator (which is not the appropriate estimator for this data given its reduced efficiency—however, the estimated coefficient of endogeneity variable when predicting citations with this estimator was $-1.85$, $p = .05$; for citations per year it was $-1.77$, $p = .08$). Despite the endogeneity variable having a zero-order correlation that was slightly positive ($r = .07$), when instrumented, the coefficient changed sign and became significant (for interesting examples showing how a positive correlation flips when the endogenous variable is instrumented see: Antonakis et al., 2010; Levitt, 1997; Levitt, 2002). This result suggests that the endogeneity variable was indeed endogenous, and that unobserved variance correlated both with this variable and citations (as indicated in Table 6, the endogeneity residual was significant). For an idea of the effect size for articles having endogeneity, the incidence rate ratio (i.e., the ratio of predicted citations as function of endogeneity) indicated that an average article having endogeneity problems received only 12.31% (or 10.62%) of the total (or per year) citations of an article that does not suffer from endogeneity; the instrumental-variable Poisson estimator indicated that the percentages were 15.70% and 17.05% respectively.

The estimates reported for Models 8 and 9 of Table 5 were quite similar to those reported for Models 2 and 3 of Table 6 respectively. That is, what predicted citations without controlling for endogeneity did so in a similar way when correcting for endogeneity, suggesting that the endogeneity variable was mostly orthogonal to the rest of the included regressors. However, there were some differences: (e.g., the coefficient of biological/evolutionary became non-significant, and that of using regression became significant).

**Table 4**
Top 40 articles by citations per year.

| Rank (2013) | Author(s) | Pub. year | Issue/no. | Cites/year | Cites | Type of article | Rank (2010) | Rank diff |
|---|---|---|---|---|---|---|---|---|
| 1 | Graen & Uhl-Bien | 1995 | 6 (2) | 50.7 | 912 | Review | 1 | Unchanged |
| 2 | Podsakoff et al. | 1990 | 1 (2) | 40.9 | 940 | Quantitative | 3 | Up |
| 3 | Lowe et al.[a] | 1996 | 7 (3) | 37.4 | 636 | Quantitative | 2 | Down |
| 4 | Avolio & Gardner | 2005 | 16 (3) | 33.1 | 265 | Review | 8 | Up |
| 5 | Brown & Trevino | 2006 | 17 (6) | 28.4 | 199 | Review | 24 | Up |
| 6 | Wong & Law | 2002 | 13 (3) | 27.6 | 304 | Quantitative | 5 | Down |
| 7 | Gardner et al. | 2005 | 16 (3) | 23.8 | 190 | Theory | 7 | Unchanged |
| 8 | Day | 2000 | 11 (4) | 22.4 | 291 | Review | 9 | Up |
| 9 | Yukl | 1999 | 10 (2) | 22.4 | 313 | Review | 4 | Down |
| 10 | Bass & Steidlmeier | 1999 | 10 (2) | 22.3 | 312 | Theory | 10 | Unchanged |
| 11 | Gronn | 2002 | 13 (4) | 21.8 | 240 | Theory | 11 | Unchanged |
| 12 | Uhl-Bien et al. | 2007 | 18 (4) | 21.5 | 129 | Theory | 19 | Up |
| 13 | Avolio et al. | 2004 | 15 (6) | 21.1 | 190 | Theory | 15 | Up |
| 14 | Mumford et al. | 2002 | 13 (6) | 21.1 | 232 | Theory | 6 | Down |
| 15 | Bono & Ilies | 2006 | 17 (4) | 19.6 | 137 | Quantitative | 17 | Up |
| 16 | Antonakis et al. | 2003 | 14 (3) | 19.4 | 194 | Quantitative | 13 | Down |
| 17 | Uhl-Bien | 2006 | 17 (6) | 19.3 | 135 | Theory | 50 | Up |
| 18 | Burke et al.[a] | 2006 | 17 (3) | 17.7 | 124 | Quantitative | 33 | Up |
| 19 | van Knippenberg et al. | 2004 | 15 (6) | 17.3 | 156 | Review | 14 | Down |
| 20 | Shalley & Gilson | 2004 | 15 (1) | 17.1 | 154 | Review | 30 | Up |
| 21 | Den Hartog et al. | 1999 | 10 (2) | 17.0 | 238 | Quantitative | 12 | Down |
| 22 | Fry | 2003 | 14 (6) | 16.8 | 168 | Theory | 20 | Down |
| 23 | Zaccaro et al. | 2001 | 12 (4) | 16.5 | 198 | Theory | 18 | Down |
| 24 | Jung et al. | 2003 | 14 (4) | 16.5 | 165 | Quantitative | 27 | Up |
| 25 | Rafferty & Griffin | 2004 | 15 (3) | 15.9 | 143 | Quantitative | 25 | Unchanged |
| 26 | Antonakis et al. | 2010 | 21 (6) | 15.7 | 47 | Review | New | – |
| 27 | Yammarino et al. | 2005 | 16 (6) | 14.5 | 116 | Review | 16 | Down |
| 28 | Eagly & Carli | 2003 | 14 (6) | 14.0 | 140 | Review | 29 | Up |
| 29 | Amabile et al. | 2004 | 15 (1) | 13.6 | 122 | Quantitative | 35 | Up |
| 30 | Ilies et al. | 2005 | 16 (3) | 13.4 | 107 | Theory | 46 | Up |
| 31 | Shamir & Eilam | 2005 | 16 (3) | 12.5 | 100 | Theory | 40 | Up |
| 32 | Osborn et al. | 2002 | 13 (6) | 12.1 | 133 | Theory | 23 | Down |
| 33 | Ensley et al. | 2006 | 17 (3) | 12.0 | 84 | Quantitative | New | – |
| 34 | Antonakis et al. | 2009 | 20 (2) | 12.0 | 48 | Comment./Disc. | New | – |
| 35 | Conger | 1999 | 10 (2) | 11.9 | 167 | Review | 21 | Down |
| 36 | Shamir & Howell | 1999 | 10 (2) | 11.7 | 164 | Theory | 26 | Down |
| 37 | Schriesheim et al. | 1999 | 10 (1) | 11.6 | 163 | Review | 22 | Down |
| 38 | De Hoogh & Den Hartog | 2008 | 19 (3) | 11.6 | 58 | Quantitative | New | – |
| 39 | Liden et al. | 2008 | 19 (2) | 11.4 | 57 | Quantitative | New | – |
| 40 | McColl-Kennedy & Anderson | 2002 | 13 (5) | 11.4 | 125 | Quantitative | 32 | Down |

Note: Citation data is from Scopus. The 2010 rank is reported in Gardner et al. (2010), Spearman $\rho$ with our rank is .75, $p < .001$ (suggesting that successful articles continue to be successful, at least in the medium term).
[a] These articles were meta-analyses.

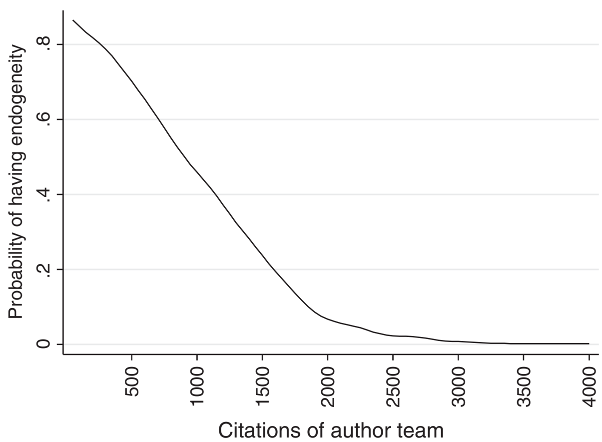



**Fig. 4.** Predicting endogeneity as function of author team citations.

The coefficients of the leadership schools were significantly different from each other, the Wald $\chi^2(7) = 14.45$, $p < .05$; the univariate Wald tests, with a Bonferroni adjustment for multiple testing, showed that only the coefficient of "new leadership" was significantly higher than that of contextual ($p = .06$); there were no other differences among the coefficients. Moreover, as before, the interaction between article age and number of articles published was significant.

### *4.5. Descriptive trends*

Below we report results on a few key trends. To determine if there was a trend in the data, we used robust regression, which weights and discounts observations according to their outlier status (Huber, 1964). We included polynomial terms if relevant and report the trend of the highest power. If the estimator failed to converge (usually because of a preponderance of zeros), we used the Poisson, zero-inflated Poisson, zero inflated negative binomial models or, as a last resort, OLS regression (all with robust standard errors). If the results from one particular estimator were uncertain, we verified them with another estimator and then decided on the trend depending on what the majority of estimators indicated. In Fig. 5, we report trends for type of article. As we note in the figure the number of quantitative, theory, and review articles published currently has a positive trend; qualitative articles, commentaries/discussions, method and agent-based simulations show a flat trend. Fig. 6 indicates that the use of ANOVA, regression, structural equation modeling, random-effects regression, and PLS is on the increase, with other methods showing a flat trend (EFA, WABA, and other methods). Fig. 7 indicates that all of the designs of the study are on the rise (except for meta-analysis).

As regards trends for the other variables we gathered, and given that the volume of articles published has increased over time, it was not surprising to see that most of the variables showed upward trends; those variables that have flat trends are, therefore, the interesting cases. For instance, for school of leadership, all are on the increase except the trait and contingency approaches. Interested readers may refer to the Appendix A, where we report trends in the rest of the data we gathered.

## 5. Discussion

At a lecture William Thomson (also known as "Lord Kelvin") once said:

> "I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a [meager] and unsatisfactory kind" (p. 792).
>
> [S. P. Thompson, 1976]

We opened our discussion with this quotation to showcase the importance of measurement and quantitative analysis for (a) undertaking research, (b) studying how research is undertaken, and (c) estimating how type of research method and statistical conclusion validity affects whether research is impactful. As our results showed, quantitative research is on the rise in our field and, along with reviews, methodological, and theoretical articles, attracts the most attention from other scholars. Qualitative articles were significantly less impactful both in terms of citations and in terms of the number of articles that are published by the journal; they were also "undercited" relative to their share of articles published. In addition, certain types of methods and theoretical approaches predicted citation rates. In particular, those schools of leadership that are currently in vogue (Day & Antonakis, 2012) tend to receive more citations, which probably reflect to some extent the fact that some of these lines of leadership (e.g., the "new leadership") have helped significantly move leadership theory forward; increased citations may also be a result of a "bandwagon" effect. The fact that we controlled for number of cited references in an article (which may reflect strategic and in-vogue citing) ensures that we have partialled out some of these effects.

Importantly, we discovered that articles having threats to the validity of results (i.e., because of endogeneity) are less cited; these results show that academic reputations, insofar as citations are concerned, will suffer if authors do not pay attention to endogeneity. We believe that learning how to correctly estimate models and using robust designs is a small price to pay as compared to the price researchers pay by not having their work cited; of course, there are other costs too, direct and indirect to researchers' institutions and society, which cannot be well informed by endogeneity-plagued research. In terms of predicting endogeneity, we found that author teams that are highly cited and have more authors predicted reduced endogeneity. Interestingly, author teams with more articles predicted increased endogeneity, which probably reflects the fact that increased output is negatively correlated with quality (cf. Antonakis & Lalive, 2008).

That the academic marketplace under-cites qualitative articles does not bode well for qualitative research; still we think that there is a place for this method of inquiry because "to better understand complex, embedded phenomena, qualitative approaches to studying leadership are also necessary" (Antonakis et al., 2004, p. 54). Top journals make exhortations to publish more qualitative work (Gephart, 2004), as does the present journal. Despite these pleas we would like to highlight that qualitative research still operates in a weak research paradigm (i.e., with respect to what constitutes commonly accepted standards of inquiry) and this in a weak paradigm field (i.e., management); more work should be done to determine specific best practices with respect to appropriateness of method used and ways to ensure that findings can be replicated.

Of course, we do not know why qualitative articles are undercited; apart from our study and another one of which we are aware (Swygart-Hobaugh, 2004), there is not much information about this topic. One interesting finding by Swygart-Hobaugh (2004) from the discipline of sociology, which has a strong qualitative tradition, is that the top two journals in sociology (i.e., American Journal of Sociology and American Sociological Review) are dominated by quantitative articles. More interestingly, in comparing citation patterns from the top two journals to two sociology journals that publish exclusively qualitative research, Swygart-Hobaugh (2004) found that qualitative researchers do cite quantitative research; however, quantitative researchers do not cite qualitative research as much.

We can only speculate as to why qualitative articles are undercited. We do however, make what we think are constructive and pragmatic suggestions for improving qualitative research and anticipate that adoption of these suggestions would make qualitative research to be viewed more positively by researchers, and quantitative researchers, in general. Researchers using qualitative methods should strive to present convincing evidence of the replicability of their results (cf. Patton, 2002). To improve the status quo we think that there should be more uniformity in the methods used. At a basic level using multiple coders and calculating coding consistency statistics, whenever possible, should increase the validity of data. It is all too easy to see patterns and inadvertently find supportive evidence (cf. Nickerson, 1998) when operating from an "interpretivist" point of view, where interpretations made can be highly idiosyncratic (Antonakis et al., 2004).

Using multiple cases should make results more generalizable and provide more data points for further (ideally quantitative) testing (cf. Ligon, Harris, & Hunter, 2012). As for sampling of cases, qualitative researchers often gather "purposive" samples by studying phenomena in very unique or restrictive settings in an attempt to glean from this setting what drives the phenomenon at hand (Yin, 1994). Although intuitively appealing, ironically, it is hard to discover valid cross-case patterns because there is no variance in the setting or the outcomes; that is, studying entities in restrictive situations for example, studying only high performing entities does not tell us much if patterns are found across the entities because it is possible that low-performing entities exhibit the same patterns (Denrell, 2003, 2005; Hastie & Dawes, 2001). Thus, there should be variance in the variables under observation, particularly in the dependent variables; that is, there should be a "control" group of cases, which can also be easily established using matching methods like propensity score analysis (D'Agostino, 1998; Li, 2012; Rubin, 2008) as is often done in the clinical sciences when comparing small groups of patients. Related to having control groups is what some qualitative researchers called using "polar opposites" cases; and ideally studies should have multiple cases too (Eisenhardt & Graebner, 2007). Even qualitative "grounded theory" approaches (Glaser & Strauss, 1968), which are supposed to discover theory from the analysis of data can benefit from the above recommendations, because discovering links between data to build a theory (that $x$ may cause $y$) for later testing rests on valid observation, classification, and pattern-matching for entities exhibiting some properties (the $x$'s) that theoretically cause some sort of outcome entity (the $y$'s) to vary.

In addition, we think that it would be best for researchers to quantify their qualitative data, to the extent that it is possible (Eagly & Antonakis, 2014; Maxwell, 2010; Simonton, 2003), and to then use these data to test hypotheses (e.g., as can be done when content analyzing data, see Study 2 in Antonakis, Fenley, & Liechti, 2011). Doing so will allow qualitative researchers to employ counterfactual conditions. Of course, we are not suggesting that the quantitative paradigm is perfect because it can also produce results that are invalid if done poorly or if one intentionally "cherry picks" findings, fiddles with, or fabricates data. Such outcomes can be very damaging in high-profile fields, particularly in the medical sciences. As a nicely reported quote by Twain (1906) notes, "There are three kinds of lies: lies, damned lies, and statistics." However an even better quotation on this matter is from the eminent statistician Frederick Mosteller: "it is easy to lie with statistics, it is even easier to lie without them" (Murray, 2005, p. 240). Thus, we hope to see more mixed-methods research that uses qualitative data in a content analytic-type framework that allows for quantification and later testing (cf. Simonton, 2003). We believe that our recommendations would not be too taxing to undertake and may help make qualitative research findings more valid and impactful, at least in our field of study.

Turning to our other results, one interesting finding that emerged both in predicting citations of all article types and of only quantitative articles was that articles tend to be more cited if the journal published more articles in that particular year. As we surmised, this result is probably explained by the fact that publishing more articles increases the likelihood that one or more of these articles will go on to be highly cited. Interestingly too, number of authors predicted citations when studying all article types; however, it did not when studying only quantitative articles (and in any case it was an excluded instrument in our two-stage procedure).

As for the results predicting citations of quantitative articles, we found several variables to be significant predictors including schools of leadership and various methodological approaches. The impact of these coefficients remained similar when controlling for the endogeneity variable. The positive effects of some of the methods on citations (e.g., the use of random-effects models, SEM, or regression) were completely undone if the articles had endogeneity threats; that is, the coefficient of the endogeneity variable was significantly higher than the average effect of these three variables or when comparing them individually to the endogeneity coefficient.

### *5.1. The future of methods*

Our results have several implications; we will briefly focus on those concerning improving the state of practice in the use of quantitative methods. As others have mentioned repeatedly, doctoral students must be better trained in statistics (Aiken, West, & Millsap, 2008; Antonakis et al., 2010; Steiger, 2001). In particular, authors as well as reviewers and editors must learn to "lift the hood" of statistical engines to better understand what assumptions and restrictions estimators make so that researchers do not violate these assumptions and hence bias estimates due to endogeneity or other factors.

For example, the use of random-effects modeling (via HLM), which is showing a clear upward trend, is worrisome because it appears that most who use the method do not know that they may have reported seriously compromised estimates by ignoring endogeneity and not having examined if their estimator is consistent (Antonakis et al., 2010; Halaby, 2004). There are very simple remedial procedures to take when the random-effects estimator is inconsistent and we hope that researchers (and journal editors) will pay more attention to these issues. There is nothing wrong with the random-effects estimator per se; it must just be used appropriately.

On another note, the use of PLS is, unfortunately, on the rise in this journal. Unlike the problem of incorrectly using estimators like random-effects, PLS itself is problematic. PLS is largely ignored in the psychometric and econometric literatures, yet it appears to

**Table 5**
Predicting citation rates for quantitative articles.

| (Model) | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| # of authors | −.00 | −.00 | .01 | .01 | .01 | .02 | .02 | .02 | .02 |
| | (.05) | (.03) | (.22) | (.41) | (.38) | (.65) | (.56) | (.60) | (.53) |
| Artic. age (a) | .21** | .19** | .20** | .20** | .20** | .19** | .04 | −.06 | −.12** |
| | (7.02) | (6.91) | (7.13) | (7.24) | (6.93) | (6.80) | (.86) | (1.51) | (2.59) |
| Cited references | .00* | .00 | .00 | .00 | .00 | .00 | .00* | .01** | .01** |
| | (2.03) | (1.32) | (1.01) | (1.47) | (1.46) | (1.51) | (2.01) | (3.03) | (3.30) |
| Trait | | .94** | .92** | .82** | .83** | .83** | .65** | .68** | .69** |
| | | (4.36) | (4.36) | (3.73) | (3.72) | (3.64) | (2.83) | (2.97) | (2.83) |
| Behavioral | | .74** | .74** | .72** | .71** | .74** | .59* | .66** | .65* |
| | | (2.94) | (2.96) | (2.92) | (2.95) | (3.04) | (2.49) | (2.71) | (2.52) |
| Contextual | | .69** | .71** | .65** | .66** | .66** | .59** | .58** | .59** |
| | | (3.53) | (3.71) | (3.32) | (3.29) | (3.22) | (2.83) | (2.67) | (2.59) |
| Relational | | .67** | .69** | .58** | .58** | .61** | .57** | .58* | .61* |
| | | (3.31) | (3.38) | (2.75) | (2.73) | (2.79) | (2.60) | (2.50) | (2.55) |
| Info. Proc. | | .69** | .65** | .65** | .66** | .68** | .60** | .71** | .70* |
| | | (2.86) | (2.86) | (2.88) | (2.87) | (2.83) | (2.58) | (2.62) | (2.42) |
| New leader | | 1.18** | 1.18** | 1.12** | 1.13** | 1.16** | 1.02** | 1.04** | 1.03** |
| | | (6.75) | (6.81) | (6.05) | (5.95) | (5.91) | (5.09) | (4.94) | (4.57) |
| Biolog/evolu | | 1.37** | 1.34** | 1.32** | 1.32** | 1.35** | 1.11** | 1.16** | 1.25** |
| | | (4.86) | (4.77) | (4.68) | (4.62) | (4.55) | (3.69) | (3.77) | (3.84) |
| Hybrid | | .90** | .90** | .83** | .84** | .84** | .62** | .62** | .63** |
| | | (4.71) | (4.67) | (4.11) | (4.10) | (4.06) | (2.90) | (2.70) | (2.68) |
| Corr. | | | .06 | −.05 | −.05 | −.00 | .01 | .00 | .11 |
| | | | (.27) | (.22) | (.21) | (.00) | (.04) | (.01) | (.65) |
| ANOVA | | | .09 | .12 | .12 | .10 | .15 | .19 | .19 |
| | | | (.72) | (1.00) | (.98) | (.81) | (1.24) | (1.64) | (1.79) |
| Reg | | | .16 | .22 | .22 | .20 | .15 | .15 | .22* |
| | | | (1.22) | (1.63) | (1.62) | (1.53) | (1.20) | (1.32) | (1.97) |
| EFA | | | .29 | .34 | .35 | .32 | .23 | .29 | .20 |
| | | | (1.28) | (1.49) | (1.51) | (1.42) | (1.12) | (1.48) | (1.00) |
| Rand. eff. (HLM) | | | .28 | .34* | .33* | .29* | .32* | .33* | .38** |
| | | | (1.87) | (2.27) | (2.26) | (2.00) | (2.20) | (2.48) | (2.77) |
| WABA | | | −.17 | −.12 | −.13 | −.12 | −.08 | −.19 | −.16 |
| | | | (.99) | (.70) | (.74) | (.71) | (.48) | (1.14) | (1.00) |
| PLS | | | .28 | .34 | .34 | .34 | .23 | .27 | .36 |
| | | | (1.32) | (1.61) | (1.60) | (1.58) | (1.12) | (1.36) | (1.81) |
| SEM | | | .27 | .28 | .27 | .24 | .34* | .32* | .36** |
| | | | (1.94) | (1.80) | (1.77) | (1.58) | (2.30) | (2.35) | (2.63) |
| Other meth. | | | .02 | −.03 | −.03 | −.06 | −.08 | −.05 | .02 |
| | | | (.09) | (.15) | (.14) | (.29) | (.38) | (.30) | (.15) |
| Field study | | | | −.25 | −.25 | −.22 | −.28 | −.26 | −.34 |
| | | | | (.78) | (.77) | (.67) | (.91) | (.89) | (1.25) |
| Lab study | | | | −.39 | −.40 | −.37 | −.40 | −.39 | −.36 |
| | | | | (1.05) | (1.07) | (1.00) | (1.13) | (1.25) | (1.36) |
| Field exp | | | | .29 | .29 | .41 | .42 | .37 | .30 |
| | | | | (.60) | (.59) | (.85) | (.93) | (.94) | (.84) |
| Quasi-exp | | | | −.42 | −.43 | −.33 | −.39 | −.35 | −.27 |
| | | | | (1.06) | (1.07) | (.84) | (1.07) | (1.07) | (.94) |
| Archiv design | | | | −.67* | −.67 | −.62 | −.63 | −.64* | −.78** |
| | | | | (1.96) | (1.94) | (1.80) | (1.93) | (2.16) | (2.73) |
| Meta-analysis | | | | 3.70** | 3.66** | 3.84** | 3.91** | 3.49** | 3.54** |
| | | | | (4.57) | (4.59) | (4.91) | (5.08) | (5.64) | (6.26) |
| Other design | | | | −.10 | −.10 | −.12 | −.12 | −.16 | −.16 |
| | | | | (.51) | (.51) | (.61) | (.66) | (.95) | (1.05) |
| Author cites | | | | | −.00 | −.00 | .00 | .00 | .00 |
| | | | | | (.16) | (.37) | (.15) | (.79) | (.40) |
| Author articles | | | | | .00 | .00 | −.00 | −.00 | −.00 |
| | | | | | (.30) | (.21) | (.21) | (.52) | (.23) |
| Aff. rank[a] | | | | | | .00 | .00* | .00 | .00* |
| | | | | | | (1.82) | (2.05) | (1.84) | (2.00) |
| # articles (b)[b] | | | | | | | −.03** | −.05** | −.03** |
| | | | | | | | (4.50) | (6.58) | (4.06) |
| (a) x (b) | | | | | | | | .01** | .01** |
| | | | | | | | | (6.45) | (4.95) |
| Cntrols[c] | Incl. | Incl. | Incl. | Incl. | Incl. | Incl. | Incl. | Incl. | Incl. |
| Constant | .97 | .01 | −.36 | −.08 | −.11 | −.15 | 2.65** | 2.67** | .98 |
| | (1.96) | (.02) | (.68) | (.16) | (.20) | (.28) | (3.29) | (3.52) | (1.32) |
| Pseudo $R$-sq | .72 | .75 | .75 | .77 | .77 | .77 | .78 | .81 | .57 |
| $\Delta R$-sq | | .03 | .01 | .02 | .00 | .00 | .01 | .02 | |

(continued on next page)

**Table 5** (*continued*)

| (Model) | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| Wald test[d] | 960.23** | 56.96*** | 11.13 | 35.10** | .09 | 3.30 | 20.26** | 41.55** | 679.34** |

Note: $n = 384$; estimates are unstandardized; pseudo $R$-square is based on the Cox and Snell (1989) method; Models 1–8 predict total citations; Model 9 predicts citations per year; omitted category for school is contingency; robust $z$-statistics in parentheses.

[a] Reverse coded thus a higher number indicates higher rank.

[b] Indicates the total number of articles published in the year in which the article appeared.

[c] Includes dummy variables for issues and editor, coder dummy variable, dummy variables for type of scale used, dummy variables for number of studies, dummy variables for location of study, dummy variables for data source, dummy variable for temporal context of study (cross-sectional, two-time periods, longitudinal).

[d] Wald test for additional regressors from previous model (for Models 1 and 7 it is the Wald test for the full equation).

** $p < .01$.

* $p < .05$.

be slowly creeping into some management journals (see Rönkkö & Evermann, 2013, for an excellent critique). The PLS estimator has many problems, primarily due to a lack of analytical undergirding regarding the properties of the estimator (Goodhue, Thompson, & Lewis, 2013): For example, given that it makes no distributional assumptions, it neither can estimate standard errors directly, nor can test a causal specification in systems of equations because it has no overidentification test (Antonakis et al., 2010). We see no use for testing complex simultaneous equation models when the veracity of the model cannot be examined.

Some researchers justify the use of PLS because it apparently has advantages over SEM in the types of models it can estimate; however, it is well known that whatever PLS can do, can also be done with SEM programs (McDonald, 1996). PLS is often hyped as being an estimator that works well even under "ridiculously small" conditions (Marcoulides & Saunders, 2006, p. iii); however, in this regard and as soberly noted by the editors of *MIS Quarterly*, "PLS is not a silver bullet"(Marcoulides & Saunders, 2006, p. vii). In fact SEM estimators have excellent properties even at very small sample sizes (Bastardoz & Antonakis, 2013; Goodhue, Lewis, & Thompson, 2012), though there will be convergence issues if the estimated ratio of parameters to sample size is extremely small (Gagne & Hancock, 2006). Issues of convergence can affect PLS too (Henseler, 2010). For converging and correctly-specified models maximum likelihood estimates will still be consistent (Bastardoz & Antonakis, 2013). In cases of small sample size to parameter ratios researchers are better off using scale indexes (aggregating/parceling indicators of latent variables) and then modeling measurement error in the parcel by constraining the disturbance to $(1 - \text{reliability}_{(\text{scale})}) * \text{variance}_{(\text{scale})}$ (cf. Bollen, 1989). This procedure removes endogeneity bias due to measurement error and, for overidentified systems of equations, researchers can test the viability of the model. For a very small sample size to parameter ratio, the chi-square statistic may require a correction because it tends to over-reject correctly specified models. One correction that has been analytically derived, the Swain correction, works particularly well and holds rejection levels of correct models precisely where Type I errors should be: at the 5% level (Antonakis & Bastardoz, 2013; Bastardoz & Antonakis, 2013; Herzog & Boomsma, 2009).

Finally, SEM has estimators with known distributions. It can correct standard errors for multivariate non-normal distributions, clustered or hierarchical samples, and has great flexibility to model all kinds of data (Flora & Curran, 2004; Muthén, 1983, 1984; Muthén, du Toit, & Spisic, in press; Muthén & Muthén, 2012; Muthén & Shedden, 1999; Rabe-Hesketh, Skrondal, & Pickles, 2004); it seems to us that this flexibility in modeling is not adequately leveraged by quantitative researchers. For example, programs like MPlus can accommodate models having latent variables with indicators that are nominal, ordinal, counts, and so forth, along with robust estimates of the variance (i.e., making no distributional assumptions) as well as robust overidentification tests. Programs like Stata can even accommodate selection models (Heckman, 1979) as well as treatment effects models (Maddala, 1983) in SEM. One can also estimate systems of equations with limited-information estimators like two-stage least squares (Baltagi, 2002; Bollen, 1996; Bollen et al., 2007); these estimators do not spread localized bias to the rest of the model. Thus, researchers using the PLS estimator are "flying blind" because they will not know if estimates are biased because of endogeneity issues. In addition the estimator is not consistent per se (Goodhue et al., 2013; Rönkkö & Evermann, 2013), which is why calls have been made to abandon its use entirely (Antonakis et al., 2010). We echo this call here.

A final concern is that researchers are not using SEM and regression models correctly; most of the articles that used these methods did not address endogeneity threats and thus reported confounded estimates. As mentioned before, this problem is particularly evident in testing mediation models via SEM or regression methods. The current "standard procedures" like those of Baron and Kenny (1986) or Preacher and Hayes (2004) make one critical assumption: that the disturbances of the mediator and the dependent variable do not covary. If there is a significant covariance, and if the researcher uses the "standard procedures," estimates will be inconsistent. This problem may even occur in experimental data, where the causal impact of manipulated factors on $y$ via a mediator that has not been manipulated, is estimated. Researchers must correctly model this source of endogeneity bias by using two-stage least squares regression or by allowing the disturbances of the endogenous variables to correlate in the context of a SEM model (refer to the following for detailed explanations on testing mediation correctly: Antonakis et al., 2010; Foster, 2010; Foster & McLanahan, 1996; Gennetian et al., 2008; Shaver, 2005).

To conclude, there are a variety of ways in which to use observational data to make clear causal inferences and interested readers should consult more technical literature on this matter (Cook et al., 2008; Foster, 2010; James, Mulaik, & Brett, 1982; Morgan & Winship, 2007; Pearl, 2000; Rubin, 1974, 2008; Shadish & Cook, 2009; Shadish et al., 2002; Shipley, 2000). Researchers must pay more attention to these important methodological issues to ensure more relevant and impactful research and because simply finding relations between variables cannot validly inform policy if those relations are confounded.

**Table 6**
Predicting endogeneity and the impact of endogeneity on citations.

| (Model) DV: | Dependent variables | | | | |
|---|---|---|---|---|---|
| | | Model with 4 instruments | | Model with 2 instruments | |
| | (1) Endogeneity | (2) Citations | (3) Cites/year | (4) Citations | (5) Cites/year |
| *Independent variables* | | | | | |
| Endogeneity | | −2.09** | −2.24** | −2.00** | −2.14** |
| | | (3.45) | (3.47) | (3.00) | (2.92) |
| Endogeneity residual[a] | | 2.27** | 2.40** | 2.17** | 2.28** |
| | | (3.70) | (3.67) | (3.22) | (3.10) |
| Artic. age (a) | .73 | −.05 | −.10* | −.05 | −.10* |
| | (1.79) | (1.06) | (2.11) | (1.16) | (2.14) |
| Cited references | −.04** | .00 | .00* | .00* | .00* |
| | (3.84) | (1.92) | (2.42) | (1.97) | (2.49) |
| Trait | −.17 | .74** | .70** | .74** | .70** |
| | (.13) | (3.33) | (2.92) | (3.31) | (2.99) |
| Behavioral | −2.81 | .62** | .62* | .65** | .64* |
| | (1.83) | (2.66) | (2.40) | (2.73) | (2.43) |
| Contextual | −3.58* | .56** | .57* | .55** | .56* |
| | (2.29) | (2.70) | (2.55) | (2.66) | (2.54) |
| Relational | −1.15 | .59** | .63** | .61** | .64** |
| | (.85) | (2.72) | (2.72) | (2.76) | (2.79) |
| Info. proc. | 1.93 | .68** | .66* | .70** | .66* |
| | (1.70) | (2.66) | (2.42) | (2.66) | (2.38) |
| New leader | −6.84** | .87** | .83** | .88** | .85** |
| | (4.04) | (4.40) | (3.81) | (4.36) | (3.84) |
| Biolog/evolu | −13.62** | .58 | .58 | .63 | .62 |
| | (4.40) | (1.81) | (1.58) | (1.87) | (1.62) |
| Hybrid | 2.02 | .77** | .79** | .77** | .78** |
| | (1.83) | (3.42) | (3.34) | (3.41) | (3.39) |
| Corr. | −.78 | −.12 | −.01 | −.09 | .01 |
| | (.97) | (.74) | (.10) | (.53) | (.07) |
| ANOVA | −.61 | .14 | .14 | .13 | .13 |
| | (.97) | (1.24) | (1.29) | (1.14) | (1.25) |
| Reg | 2.18** | .31* | .38** | .29* | .36** |
| | (3.47) | (2.44) | (3.08) | (2.23) | (2.75) |
| EFA | −3.29** | .11 | .01 | .10 | −.00 |
| | (3.89) | (.55) | (.06) | (.50) | (.01) |
| Rand. eff. (HLM) | 6.84** | .55** | .61** | .52** | .58** |
| | (4.66) | (3.81) | (4.07) | (3.36) | (3.65) |
| WABA | 3.05 | −.26 | −.26 | −.25 | −.25 |
| | (1.93) | (1.64) | (1.72) | (1.56) | (1.65) |
| PLS | 4.44** | .38* | .50** | .37 | .49* |
| | (2.60) | (2.08) | (2.68) | (1.93) | (2.51) |
| SEM | 8.44** | .56** | .59** | .54** | .56** |
| | (5.07) | (3.64) | (3.81) | (3.32) | (3.34) |
| Other meth. | 2.52** | .01 | .08 | −.01 | .05 |
| | (2.67) | (.08) | (.41) | (.05) | (.25) |
| Field study | 2.43** | −.09 | −.15 | −.07 | −.16 |
| | (2.76) | (.29) | (.54) | (.25) | (.59) |
| Lab study | −7.58** | −1.25** | −1.29** | −1.19** | −1.24** |
| | (5.07) | (3.33) | (3.76) | (3.10) | (3.40) |
| Field exp | −1.69 | −.17 | −.31 | −.06 | −.22 |
| | (1.20) | (.41) | (.85) | (.16) | (.61) |
| Quasi-exp | 2.06 | −.36 | −.31 | −.30 | −.29 |
| | (1.63) | (1.08) | (1.07) | (.91) | (.98) |
| Archiv design | 5.35** | −.21 | −.30 | −.20 | −.30 |
| | (4.29) | (.62) | (.83) | (.59) | (.85) |
| Other design | −2.75** | −.06 | −.07 | −.08 | −.08 |
| | (2.97) | (.35) | (.40) | (.47) | (.51) |
| # of authors | −1.83** | | | | |
| | (4.97) | | | | |
| Author cites | −.01** | | | | |
| | (5.27) | | | | |
| Author articles | .11** | | | −.00 | −.00 |
| | (4.21) | | | (.37) | (.31) |
| Aff. rank[b] | .01* | | | .00 | .00 |
| | (2.24) | | | (1.39) | (1.37) |
| # articles (b)[c] | .02 | −.04** | −.03** | −.05** | −.03** |
| | (.29) | (6.90) | (4.31) | (7.08) | (4.35) |

(continued on next page)

**Table 6** (*continued*)

| (Model) DV: | Dependent variables | | | | |
|---|---|---|---|---|---|
| | | Model with 4 instruments | | Model with 2 instruments | |
| | (1) Endogeneity | (2) Citations | (3) Cites/year | (4) Citations | (5) Cites/year |
| *Independent variables* | | | | | |
| (a) × (b) | −.04** | .01** | .00** | .01** | .00** |
| | (4.38) | (4.73) | (3.30) | (4.81) | (3.26) |
| All controls[d] | Incl. | Incl. | Incl. | Incl. | Incl. |
| Constant | −8.14 | 2.97** | 1.45* | 2.97** | 1.43* |
| | (1.54) | (4.09) | (2.19) | (4.12) | (2.20) |
| Pseudo *R*-sq | .56 | .80 | .57 | .81 | .57 |
| Wald test[e] | 976.16** | 1571.86** | 663.39** | 1566.00** | 603.18** |

Note: $n = 367$ (Model 1), 371 (Models 2–5); estimates are unstandardized; pseudo $R$-square is based on the Cox and Snell (1989) method; Model 1 is a probit model; Models 2 and 3 are zero-inflated negative binomial models; endogeneity coded 0 (no endogeneity) or 1 (endogeneity); omitted category for school is contingency; robust $z$-statistics in parentheses; the exclusion restriction was satisfied for all models, using IV-Poisson: Model 2: Hansen J $\chi^2(3) = 5.61$, $p = .13$; Model 3: Hansen J $\chi^2(3) = 4.63$, $p = .20$, Model 4: Hansen J $\chi^2(1) = .50$, $p = .48$; Model 5: Hansen J $\chi^2(1) = .86$, $p = .35$ (the manual structural procedure gave the following respective $p$-values, .61, .50, .34, and .23).

[a] The test for this coefficient = 0 is the Hausman (1978) endogeneity test.

[b] Reverse coded thus a higher number indicates higher rank.

[c] Indicates the total number of articles published in year j to predict citations of article i (nested in year j).

[d] Includes dummy variables for issues, coder dummy variable, dummy variables for type of scale used, dummy variables for number of studies, dummy variables for location of study, dummy variables for data source, dummy variable for temporal context of study (cross-sectional, two-time periods, longitudinal).

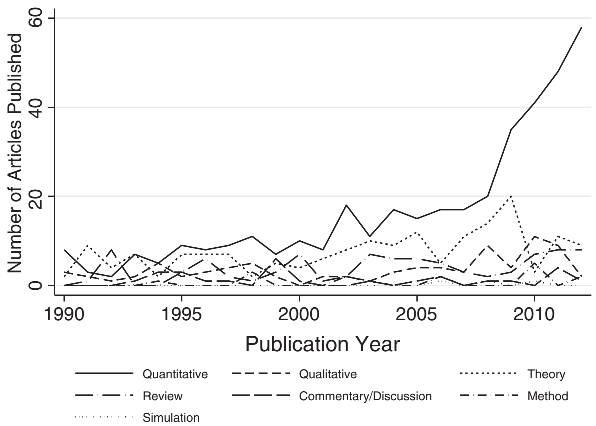
[e] Wald test for the full models.

* $p < .05$.

** $p < .01$.

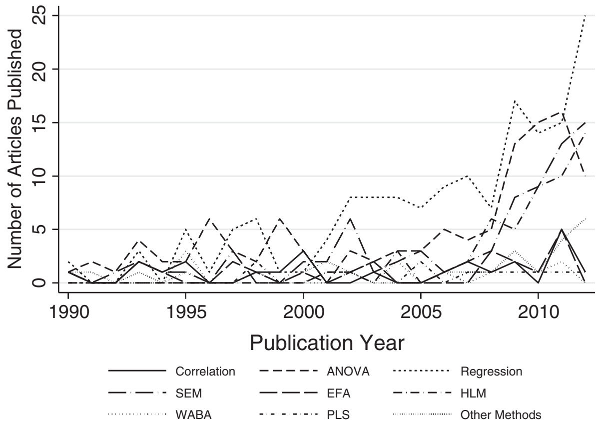
### 5.2. Limitations and suggestions

One limitation with respect to our findings concerns their generalizability to other social sciences fields. Leadership is a general social sciences topic that is studied across psychology, management, political science, economics, and sociology; in addition, similar standards have been found when comparing the methodological rigor of *The Leadership Quarterly* to more general journals in management and psychology (Antonakis et al., 2010). Furthermore, we analyzed a large sample of articles representing a wide range of authors. As reported by Scopus at the time of publication, these articles had a cumulative authorship of 2039 (1191 unique) individuals having published 36,637 (16,817 unique) articles that had received 560,692 (unique 284,777) citations. Thus, even though we have a broad spread of authors, types of articles, and methods used, it is still not clear whether our results would generalize




| | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | Total | Trend |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Quantitative | 8 | 3 | 2 | 7 | 5 | 9 | 8 | 9 | 11 | 7 | 10 | 8 | 18 | 11 | 17 | 15 | 17 | 17 | 20 | 35 | 41 | 48 | 58 | 384 | + |
| Qualitative | 3 | 2 | 1 | 2 | 5 | 2 | 3 | 4 | 5 | 2 | 0 | 2 | 2 | 1 | 3 | 4 | 4 | 3 | 9 | 4 | 11 | 9 | 2 | 83 | |
| Theory | 2 | 9 | 4 | 7 | 2 | 7 | 7 | 7 | 2 | 5 | 4 | 6 | 8 | 10 | 9 | 12 | 5 | 11 | 14 | 20 | 3 | 11 | 9 | 174 | + |
| Review | 0 | 1 | 8 | 0 | 0 | 3 | 6 | 2 | 1 | 3 | 7 | 1 | 2 | 7 | 6 | 6 | 5 | 3 | 2 | 3 | 7 | 8 | 8 | 89 | + |
| Comment/Discussion | 0 | 0 | 0 | 1 | 3 | 3 | 1 | 1 | 0 | 6 | 1 | 0 | 0 | 1 | 0 | 1 | 2 | 0 | 1 | 1 | 0 | 4 | 1 | 27 | |
| Method | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 2 | 0 | 0 | 0 | 5 | 0 | 2 | 16 | |
| Simulation | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 3 | |
| Total | 13 | 15 | 15 | 17 | 16 | 24 | 25 | 23 | 22 | 23 | 22 | 17 | 32 | 31 | 35 | 38 | 36 | 34 | 47 | 63 | 68 | 80 | 80 | 776 | |

**Fig. 5.** Descriptive trends for type of article.




| | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | Total | Trend |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Correlation | 1 | 0 | 0 | 2 | 1 | 2 | 0 | 0 | 1 | 1 | 3 | 0 | 0 | 1 | 0 | 0 | 1 | 2 | 1 | 2 | 0 | 5 | 1 | 24 | |
| ANOVA | 1 | 2 | 1 | 4 | 2 | 2 | 6 | 3 | 2 | 6 | 3 | 0 | 3 | 2 | 3 | 3 | 5 | 4 | 5 | 13 | 15 | 16 | 10 | 111 | + |
| Regression | 2 | 0 | 0 | 3 | 0 | 5 | 1 | 5 | 6 | 1 | 1 | 4 | 8 | 8 | 8 | 7 | 9 | 10 | 7 | 17 | 14 | 15 | 25 | 156 | + |
| SEM | 1 | 0 | 1 | 2 | 1 | 1 | 0 | 3 | 0 | 0 | 2 | 2 | 6 | 1 | 2 | 3 | 1 | 2 | 3 | 8 | 9 | 13 | 15 | 76 | + |
| EFA | 1 | 0 | 0 | 2 | 1 | 0 | 0 | 2 | 1 | 0 | 1 | 0 | 1 | 2 | 0 | 0 | 0 | 0 | 3 | 2 | 1 | 5 | 0 | 22 | |
| HLM | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 4 | 0 | 1 | 2 | 6 | 5 | 9 | 10 | 14 | 53 | + |
| WABA | 1 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 2 | 0 | 1 | 0 | 1 | 2 | 1 | 2 | 0 | 15 | |
| PLS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 1 | 1 | 2 | 0 | 2 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 14 | + |
| Others | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 2 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 3 | 1 | 4 | 6 | 26 | |
| Total | 8 | 3 | 2 | 14 | 5 | 14 | 7 | 13 | 14 | 8 | 11 | 10 | 22 | 16 | 19 | 15 | 19 | 22 | 28 | 53 | 51 | 71 | 72 | 497 | |

Note: Total number of statistical methods does not equal total number of quantitative articles due to non-mutually exclusive coding;ANOVA = Analysis of variance, SEM = structural equation modeling and confirmatory factor analysis, EFA = exploratory factor analysis, RE = random effects models, WABA = within and between analysis, PLS = partial least squares analysis.

**Fig. 6.** Descriptive trends for statistical methods used to test hypotheses. Note: Total number of statistical methods does not equal total number of quantitative articles due to non-mutually exclusive coding; ANOVA = analysis of variance, SEM = structural equation modeling and confirmatory factor analysis, EFA = exploratory factor analysis, RE = random effects models, WABA = within and between analysis, PLS = partial least squares analysis.

to other settings, even those that have similar research practices. We therefore recommend that future research consider journals in other disciplines as well as other variables that might predict citations.

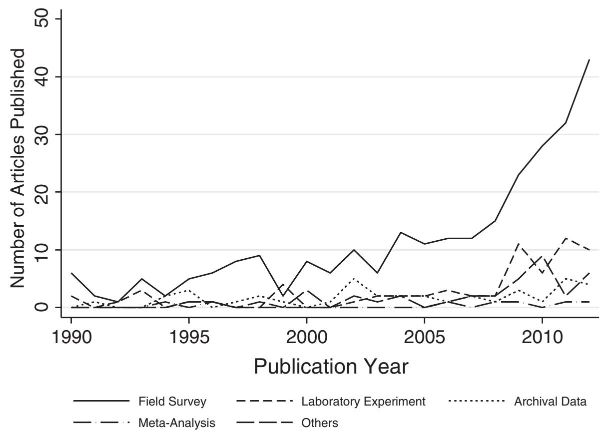
In addition, the results we obtained on the "schools of leadership" depend entirely on our coding scheme, which was rather parsimonious. For instance, because of theoretical and empirical similarities, we coded values-based "authentic" and "ethical" leadership articles under the banner of "new leadership" approaches. If these theories stand the test of time, then researchers should consider coding these approaches into unique categories. Similarly, we also included mixed-method approaches in the quantitative articles, and we did not distinguish different types of methodological articles. Thus, as more research using these approaches is undertaken, researchers should use more detailed coding protocols.

Of course, there may be a possibility that our models omit important causes of citations, though given the variance explained and our modeling specifications we believe that this threat is very limited. At the individual level, we have captured the most important drivers of citations; lagged research performance measured in previous articles published and citations received, as well as the reputation of the universities author teams, which also partials out country-level differences at the university level. Coupled with controlling for number of authors on an article, these individual-level variables are, collectively, excellent proxies for research training received, access to funding, differences in resources, and so forth. These variables would also capture networking skills of scholars as well as personality factors too; for example, better networked and extraverted scholars would more likely collaborate with a broader set of researchers. Smarter researchers or those who have received better training are probably more likely to work with more highly-cited authors or authors from higher-ranked institutions. Moreover, we controlled for unobserved heterogeneity due to time effects, unobserved heterogeneity due to editor term, journal-level effects (number of articles published) in addition to very detailed article-level controls (number of cited references, methods used and so forth).

## 6. Conclusion

As a way of facilitating the advancement of quantitative research practices and knowledge of why some published articles have more impact than others, we undertook an in-depth study of the factors that determine why scholarly articles are cited. Our results have obvious value for those interested in this journal, but also value for social science scholars in many different disciplines. Our findings also highlighted some inconvenient facts particularly regarding deficient methodological practices.

We summarize our recommendations in Table 7; these recommendations may be criticized because they may help create a self-fulfilling prophecy; that is, by encouraging those methods that breed citation success and which reduce endogeneity we may




| | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | Total | Trend |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Field Survey | 6 | 2 | 1 | 5 | 2 | 5 | 6 | 8 | 9 | 2 | 8 | 6 | 10 | 6 | 13 | 11 | 12 | 12 | 15 | 23 | 28 | 32 | 43 | 265 | + |
| Laboratory Experiment | 2 | 0 | 1 | 3 | 0 | 1 | 1 | 0 | 0 | 4 | 0 | 0 | 1 | 2 | 2 | 2 | 3 | 2 | 2 | 11 | 6 | 12 | 10 | 65 | + |
| Archival Data | 0 | 1 | 0 | 0 | 2 | 3 | 0 | 1 | 2 | 1 | 0 | 1 | 5 | 2 | 2 | 2 | 1 | 2 | 1 | 3 | 1 | 5 | 4 | 39 | + |
| Meta-Analysis | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 7 | |
| Others | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 3 | 0 | 2 | 1 | 2 | 0 | 1 | 2 | 2 | 5 | 9 | 2 | 6 | 38 | + |
| Total | 8 | 3 | 2 | 8 | 5 | 10 | 9 | 9 | 12 | 7 | 11 | 7 | 18 | 11 | 19 | 15 | 18 | 18 | 21 | 43 | 44 | 52 | 64 | 414 | |

Note: Total number of data collection techniques does not equal total number of quantitative articles due to non-mutually exclusive coding.

**Fig. 7.** Descriptive trends for type of data collection techniques. Note: Total number of data collection techniques does not equal total number of quantitative articles due to non-mutually exclusive coding.

contribute to the reduction in use of other methods. We wish to offer an alternative natural selection perspective, which suggests a more virtuous cycle. That is, we hope to see improved quality standards so that published research is more relevant and interpretable for researchers, is used more frequently in the development of new knowledge, and ultimately informs practice and policy. Such research does not have to be exclusively applied; even "hard-core" methods and Monte Carlo simulation articles can be heavily cited provided that they make clear, novel, and relevant contributions.

Considering the power law distribution of citations, which we found supported in that only a small number of articles get highly cited and many articles fail to garner much attention or citation (Meho, 2007; Seglen, 1992), we believe that our recommendations

**Table 7**
Recommendations to researchers.

| |
|---|
| Which types of articles receive the most citations? |
| 1. Methodological |
| 2. Review |
| 3. Quantitative |
| 4. Theory |
| Which types of articles receive the fewest citations? |
| 1. Agent-based simulations |
| 2. Commentary |
| 3. Qualitative |
| What are predictors of citations across all types of articles? |
| 1. Author teams from more highly ranked universities |
| 2. More cited references |
| 3. More coauthors |
| What are predictors of citations across quantitative articles? |
| 1. No endogeneity threats (depends on having a better cited author team and more coauthors) |
| 2. More cited references[a] |
| 3. Using "new leadership," trait, behavioral, contextual, relational, information-processing, or hybrid and not contingency models |
| 4. Regression, random-effects, or structural equation modeling[b] |
| 5. Not basing data on a laboratory study |
| 6. Data from cross-national contexts |
| 7. Meta-analysis[c] |

Note: the above recommendations report the conclusions from Tables 3 and 6 for models predicting total article citations (we include here substantive results from control variables not reported in the tables).

[a] This recommendation follows from the model using two instruments.

[b] We do not recommend using PLS for the reasons presented in the discussion (also the coefficient for PLS was not significant for the model using two instruments).

[c] This recommendation is based on the results of Table 5 (that does not include possible endogeneity threats to the meta-analyses).

## Appendix A. Descriptive trends in variables measured for quantitative articles

| Category | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | Total | Trend |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *School of leadership* | | | | | | | | | | | | | | | | | | | | | | | | | |
| Trait | 0 | 2 | 0 | 1 | 0 | 1 | 0 | 0 | 3 | 1 | 5 | 2 | 5 | 1 | 2 | 1 | 2 | 2 | 0 | 2 | 1 | 10 | 2 | 43 | |
| Behavioral | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 2 | 1 | 0 | 0 | 2 | 1 | 1 | 2 | 1 | 2 | 4 | 3 | 1 | 26 | + |
| Contextual | 0 | 1 | 0 | 2 | 1 | 1 | 2 | 5 | 0 | 1 | 1 | 0 | 7 | 2 | 1 | 1 | 3 | 3 | 3 | 6 | 12 | 2 | 10 | 64 | + |
| Contingency | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 3 | 14 | |
| Relational | 2 | 0 | 0 | 1 | 0 | 2 | 1 | 2 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 2 | 0 | 5 | 6 | 6 | 4 | 14 | 48 | + |
| Information processing | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 2 | 0 | 0 | 2 | 2 | 4 | 5 | 20 | + |
| New leadership | 3 | 0 | 0 | 1 | 1 | 2 | 2 | 0 | 4 | 5 | 1 | 4 | 4 | 4 | 7 | 7 | 2 | 5 | 5 | 8 | 14 | 15 | 13 | 107 | + |
| Biological/evolutionary | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 2 | 0 | 3 | 5 | 15 | + |
| Hybrid | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 2 | 2 | 3 | 3 | 5 | 5 | 6 | 2 | 7 | 5 | 47 | + |
| Total | 8 | 3 | 2 | 7 | 5 | 9 | 8 | 9 | 11 | 7 | 10 | 8 | 18 | 11 | 17 | 15 | 17 | 17 | 20 | 35 | 41 | 48 | 58 | 384 | |
| *Location of study* | | | | | | | | | | | | | | | | | | | | | | | | | |
| North America | 6 | 2 | 1 | 5 | 3 | 8 | 6 | 3 | 9 | 6 | 7 | 6 | 11 | 6 | 12 | 10 | 9 | 11 | 12 | 23 | 30 | 27 | 32 | 245 | + |
| Europe | 0 | 0 | 1 | 2 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 1 | 1 | 4 | 2 | 1 | 2 | 5 | 5 | 9 | 19 | 56 | + |
| Asia | 1 | 0 | 0 | 0 | 0 | 2 | 1 | 2 | 1 | 0 | 2 | 0 | 1 | 2 | 1 | 1 | 2 | 2 | 3 | 5 | 5 | 6 | 8 | 45 | + |
| Others | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 1 | 3 | 0 | 2 | 0 | 1 | 0 | 2 | 0 | 0 | 2 | 1 | 14 | |
| Cross-national | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 2 | 2 | 0 | 2 | 0 | 4 | 1 | 18 | |
| N/A | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 2 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 11 | |
| Total | 8 | 3 | 2 | 7 | 5 | 10 | 8 | 9 | 11 | 8 | 10 | 8 | 18 | 11 | 17 | 15 | 17 | 17 | 20 | 35 | 41 | 48 | 61 | 389 | |
| *Sources of data* | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 subjective source | 3 | 1 | 1 | 2 | 2 | 0 | 3 | 4 | 3 | 3 | 3 | 4 | 9 | 4 | 5 | 6 | 7 | 6 | 9 | 7 | 9 | 16 | 16 | 123 | + |
| 2 or more subj. sources | 2 | 0 | 0 | 2 | 1 | 5 | 2 | 3 | 6 | 1 | 7 | 1 | 6 | 4 | 5 | 3 | 5 | 3 | 9 | 14 | 15 | 21 | 27 | 142 | + |
| 1or more obj. source(s) | 3 | 2 | 1 | 5 | 1 | 5 | 2 | 2 | 2 | 4 | 0 | 3 | 5 | 3 | 7 | 6 | 5 | 8 | 4 | 16 | 20 | 18 | 17 | 139 | + |
| N/A | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 4 | 11 | |
| Total | 8 | 3 | 2 | 9 | 5 | 10 | 8 | 9 | 11 | 8 | 11 | 8 | 20 | 11 | 17 | 15 | 18 | 17 | 23 | 38 | 44 | 56 | 64 | 415 | |
| *Temporal context* | | | | | | | | | | | | | | | | | | | | | | | | | |
| Cross-sectional | 5 | 1 | 1 | 3 | 3 | 7 | 4 | 7 | 6 | 2 | 7 | 3 | 11 | 5 | 10 | 8 | 13 | 11 | 15 | 17 | 21 | 19 | 34 | 213 | + |
| Two time periods | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 3 | 0 | 2 | 3 | 5 | 2 | 2 | 2 | 1 | 2 | 1 | 6 | 10 | 10 | 11 | 63 | + |
| Longitudinal | 1 | 2 | 0 | 3 | 0 | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 3 | 2 | 0 | 1 | 0 | 1 | 1 | 8 | 2 | 28 | |
| Non-survey studies | 2 | 0 | 1 | 2 | 1 | 2 | 3 | 1 | 1 | 5 | 1 | 2 | 3 | 3 | 2 | 3 | 5 | 3 | 4 | 13 | 12 | 16 | 16 | 101 | + |
| Total | 8 | 3 | 2 | 8 | 5 | 11 | 8 | 9 | 11 | 7 | 10 | 8 | 19 | 11 | 17 | 15 | 19 | 17 | 20 | 37 | 44 | 53 | 63 | 405 | |
| *Scale used* | | | | | | | | | | | | | | | | | | | | | | | | | |
| New scale | 2 | 0 | 1 | 3 | 2 | 1 | 4 | 2 | 4 | 4 | 4 | 1 | 8 | 6 | 3 | 2 | 5 | 3 | 7 | 9 | 15 | 14 | 15 | 115 | + |
| Original scale | 5 | 0 | 1 | 5 | 2 | 4 | 3 | 6 | 7 | 2 | 6 | 6 | 14 | 10 | 14 | 13 | 12 | 15 | 13 | 29 | 33 | 39 | 46 | 285 | + |
| Modified scale | 0 | 0 | 0 | 4 | 0 | 2 | 1 | 6 | 1 | 2 | 3 | 2 | 2 | 5 | 0 | 3 | 1 | 7 | 5 | 16 | 6 | 19 | 7 | 92 | + |
| No scale used | 1 | 3 | 1 | 1 | 1 | 4 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 2 | 2 | 2 | 2 | 3 | 4 | 1 | 4 | 4 | 40 | + |
| Total | 8 | 3 | 3 | 13 | 5 | 11 | 9 | 14 | 13 | 9 | 14 | 10 | 24 | 21 | 19 | 20 | 20 | 27 | 28 | 58 | 55 | 76 | 72 | 532 | |
| *No. of studies* | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 study | 8 | 3 | 1 | 5 | 5 | 8 | 8 | 8 | 10 | 4 | 7 | 6 | 16 | 10 | 17 | 13 | 12 | 17 | 12 | 25 | 33 | 37 | 46 | 311 | + |
| 2 studies | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 3 | 1 | 0 | 1 | 1 | 0 | 2 | 4 | 0 | 6 | 6 | 6 | 6 | 9 | 48 | + |
| 3 studies | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 3 | 1 | 5 | 3 | 17 | + |
| 4 or more studies | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 8 | |
| Total | 8 | 3 | 2 | 7 | 5 | 9 | 8 | 9 | 11 | 7 | 10 | 8 | 18 | 11 | 17 | 15 | 17 | 17 | 20 | 35 | 41 | 48 | 58 | 384 | |

(continued on next page)

**Appendix A** (*continued*)

| Category | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | Total | Trend |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Agg. technique* | | | | | | | | | | | | | | | | | | | | | | | | | |
| ICC | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 2 | 1 | 0 | 0 | 9 | 1 | 6 | 1 | 8 | 6 | 6 | 8 | 10 | 16 | 26 | 104 | + |
| RWG | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 2 | 1 | 0 | 0 | 6 | 4 | 7 | 2 | 8 | 4 | 6 | 5 | 8 | 10 | 11 | 77 | + |
| RE | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 0 | 1 | 2 | 6 | 0 | 6 | 3 | 9 | 30 | + |
| FE | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 3 | |
| LGM | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | |
| Other | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 2 | 1 | 6 | |
| Total | 0 | 0 | 0 | 0 | 1 | 2 | 1 | 4 | 4 | 2 | 0 | 0 | 16 | 6 | 16 | 3 | 18 | 12 | 18 | 13 | 25 | 31 | 49 | 221 | |
| *Level of analysis* | | | | | | | | | | | | | | | | | | | | | | | | | |
| Individual | 6 | 2 | 1 | 7 | 3 | 2 | 5 | 8 | 6 | 7 | 10 | 8 | 9 | 5 | 4 | 11 | 7 | 11 | 12 | 20 | 24 | 36 | 31 | 235 | + |
| Dyad | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 2 | 0 | 1 | 1 | 0 | 1 | 2 | 4 | 3 | 18 | + |
| Group | 1 | 0 | 1 | 0 | 0 | 1 | 2 | 1 | 4 | 0 | 0 | 0 | 5 | 2 | 3 | 3 | 7 | 1 | 1 | 2 | 1 | 5 | 7 | 47 | + |
| Organization/country | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 3 | 1 | 4 | 0 | 1 | 3 | 6 | 2 | 1 | 26 | + |
| Multilevel | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 1 | 0 | 0 | 0 | 4 | 2 | 5 | 0 | 1 | 4 | 6 | 9 | 9 | 4 | 16 | 66 | + |
| Total | 8 | 3 | 2 | 7 | 4 | 9 | 7 | 9 | 11 | 8 | 10 | 8 | 20 | 11 | 17 | 15 | 20 | 17 | 20 | 35 | 42 | 51 | 58 | 392 | |
| *Sample occupation* | | | | | | | | | | | | | | | | | | | | | | | | | |
| Private companies | 4 | 3 | 1 | 4 | 2 | 7 | 5 | 6 | 4 | 2 | 3 | 3 | 9 | 3 | 13 | 8 | 6 | 7 | 12 | 17 | 19 | 22 | 35 | 195 | + |
| Public companies | 0 | 0 | 0 | 1 | 2 | 3 | 0 | 4 | 2 | 1 | 2 | 4 | 1 | 1 | 3 | 3 | 2 | 7 | 4 | 10 | 7 | 19 | 8 | 84 | + |
| Teens & undergrads | 3 | 0 | 1 | 4 | 1 | 2 | 1 | 1 | 3 | 4 | 1 | 1 | 4 | 3 | 2 | 3 | 4 | 2 | 4 | 12 | 11 | 13 | 17 | 97 | + |
| Graduate students | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 2 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 2 | 14 | + |
| Executives | 2 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 2 | 1 | 4 | 4 | 4 | 4 | 2 | 28 | + |
| Military members | 1 | 0 | 1 | 2 | 0 | 0 | 0 | 1 | 1 | 0 | 7 | 2 | 4 | 3 | 0 | 1 | 2 | 4 | 2 | 2 | 2 | 2 | 1 | 38 | + |
| Total | 10 | 3 | 3 | 11 | 6 | 12 | 7 | 12 | 11 | 9 | 13 | 11 | 21 | 12 | 19 | 15 | 16 | 21 | 27 | 46 | 44 | 62 | 65 | 456 | |

ICC = intraclass correlation; RWG = within-group interrater reliability; RE = random-effects; FE = fixed-effects; LGM = latent-growth model.

should help improve article quality and citation patterns across-the-board. Thus, by recommending that only the fittest methods and approaches survive, we anticipate that more research that is valued in the academic marketplace will be produced in the future.

## Acknowledgments

We are grateful to Pauline Rattaz and Petrix Kafexolli for their help in gathering the data. The participants of the Organizational Behavior departmental seminar at the University of Lausanne, and in particular Christian Thoeni, Christian Zehnder, and Joerg Dietz provided helpful comments.

## References


Aguinis, H., & Edwards, J. R. (2013). Methodological wishes for the next decade and how to make wishes come true. *Journal of Management Studies* (in press).

Aiken, L. S., West, S. G., & Millsap, R. E. (2008). Doctoral training in statistics, measurement, and methodology in psychology — Replication and extension of Aiken, West, Sechrest, and Reno's (1990) survey of PhD programs in North America. *American Psychologist*, *63*(1), 32–50.

Aksnes, D. W. (2003). Characteristics of highly cited papers. *Research Evaluation*, *12*(3), 159–170.

Aldrich, J. H., & Nelson, F. D. (1984). *Linear probability, logit, and probit models.* Beverly Hills: Sage Publications.

Ali, M. M., Bhattacharyya, P., & Olejniczak, A. J. (2010). The effects of scholarly productivity and institutional characteristics on the distribution of federal research grants. *Journal of Higher Education*, *81*(2), 164–187.

Anderson, C. A., Lindsay, J. J., & Bushman, B. J. (1999). Research in psychological laboratory: truth or triviality? *Current Directions in Psychological Science*, *8*(1), 3–9.

Angrist, J. D. (2001). Estimation of limited dependent variable models with dummy endogenous regressors: Simple strategies for empirical practice. *Journal of Business & Economic Statistics*, *19*(1), 2–16.

Antonakis, J., & Bastardoz, N. (2013). Swain: Stata module to correct the SEM chi-square overidentification test in small sample sizes or complex models. http://econpapers.repec.org/software/bocbocode/s457617.htm

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly*, *21*(6), 1086–1120.

Antonakis, J., & Dietz, J. (2011). Looking for validity or testing it? The perils of stepwise regression, extreme-scores analysis, heteroscedasticity, and measurement error. *Personality and Individual Differences*, *50*(3), 409–415.

Antonakis, J., Fenley, M., & Liechti, S. (2011). Can charisma can be taught? Tests of two interventions. *Academy of Management Learning & Education*, *10*(3), 374–396.

Antonakis, J., & Lalive, R. (2008). Quantifying scholarly impact: IQp versus the Hirsch h. *Journal of the American Society for Information Science and Technology*, *59*(6), 956–969.

Antonakis, J., Schriesheim, C. A., Donovan, J. A., Gopalakrishna-Pillai, K., Pellegrini, E., & Rossomme, J. L. (2004). Methods for studying leadership. In J. Antonakis, A. T. Cianciolo, & R. J. Sternberg (Eds.), *The nature of leadership* (pp. 48–70). Thousand Oaks: Sage.

Baldi, S. (1998). Normative versus social constructivist processes in the allocation of citations: A network-analytic model. *American Sociological Review*, *63*(6), 829–846.

Baltagi, B. H. (2002). *Econometrics.* New York: Springer.

Bansal, P., & Corley, K. (2011). The coming of age for qualitative research: Embracing the diversity of qualitative methods. *Academy of Management Journal*, *54*(2), 233–237.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182.

Bascle, G. (2008). Controlling for endogeneity with instrumental variables in strategic management research. *Strategic Organization*, *6*(3), 285–327.

Bass, B. M. (1985). *Leadership and performance beyond expectations.* New York: The Free Press.

Bass, B. M., & Bass, R. (2008). *The Bass handbook of leadership: Theory, research, and managerial applications* (4th ed.)New York: Free Press.

Bastardoz, N., & Antonakis, J. (2013). How small can a sample size be for a structural equation model? *Paper presented at the Congress of the Swiss Society of Psychology: Crossing Borders. Basel, Switzerland*.

Bergh, D. D., Perry, J., & Hanke, R. (2006). Some predictors of SMJ article impact. *Strategic Management Journal*, *27*(1), 81–100.

Billings, R. S., & Wroten, S. P. (1978). Use of path-analysis in industrial–organizational psychology: Criticisms and suggestions. *Journal of Applied Psychology*, *63*(6), 677–688.

Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for aggregation and analysis. In S. W. J. Kozlowski, & K. J. Klein (Eds.), *Multilevel theory, research, and methods in organizations* (pp. 349–381). San Francisco, CA: Jossey-Bass.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Bollen, K. A. (1996). An alternative two stage least squares (2SLS) estimator for latent variable equations. *Psychometrika*, *61*, 109–121.

Bollen, K. A., & Brand, J. E. (2010). A general panel model with random and fixed effects: A structural equations approach. *Social Forces*, *89*(1), 1–34.

Bollen, K. A., Kirby, J. B., Curran, P. J., Paxton, P. M., & Chen, F. N. (2007). Latent variable models under misspecification — Two-stage least squares (2SLS) and maximum likelihood (ML) estimators. *Sociological Methods & Research*, *36*(1), 48–86.

Brown, D. J., & Lord, R. G. (1999). The utility of experimental research in the study of transformational/charismatic leadership. *The Leadership Quarterly*, *10*(4), 531–539.

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications.* New York: Cambridge University Press.

Cameron, A. C., & Trivedi, P. K. (2009). *Microeconometrics using Stata.* College Station, Tex.: Stata Press.

Castro, S. L. (2002). Data analytic methods for analysis of multilevel question: A comparison of intraclass correlation coefficients, rwg(j), hierarchical linear modeling, within- and between-analysis, and random group resampling. *The Leadership Quarterly*, *13*, 69–93.

Colquitt, J. A., & Zapata-Phelan, C. P. (2007). Trends in theory building and theory testing: A five-decade study of the Academy of Management Journal. *Academy of Management Journal*, *50*(6), 1281–1303.

Cook, T. D., Shadish, W. R., & Wong, V. C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management*, *27*(4), 724–750.

Cox, D. R., & Snell, E. J. (1989). *Analysis of binary data* (2nd ed.)London, New York: Chapman and Hall.

D'Agostino, R. B. (1998). Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine*, *17*(19), 2265–2281.

Dansereau, F., Alutto, J. A., & Yamarino, F. J. (1984). *Theory testing in organizational behavior: The varient approach.* Englewood Cliffs, NJ: Prentice-Hall.

Dansereau, F., Graen, G. B., & Haga, W. J. (1975). A vertical dyad linkage approach to leadership within formal organizations: A longitudinal investigation of the role making process. *Organizational Behavior and Human Performance*, *13*, 46–78.

Davidson, R., & MacKinnon, J. G. (1993). *Estimation and inference in econometrics.* New York, Oxford: Oxford University Press.

Day, D. V., & Antonakis, J. (2012). Leadership: Past, present, future. In D. V. Day, & J. Antonakis (Eds.), *The Nature of Leadership* (pp. 3–25) (2nd ed.). Thousand Oaks: Sage.

Denrell, J. (2003). Vicarious learning, undersampling of failure, and the myths of management. *Organization Science*, *14*(3), 227–243.

Denrell, J. (2005). Selection bias and the perils of benchmarking. *Harvard Business Review*, *83*(4), 114–199.

Derksen, S., & Keselman, H. J. (1992). Backward, forward and stepwise automated subset-selection algorithms — frequency of obtaining authentic and noise variables. *British Journal of Mathematical & Statistical Psychology*, *45*, 265–282.

Duncan, G. J., Magnusson, K. A., & Ludwig, J. (2004). The endogeneity problem in developmental studies. *Research in Human Development*, *1*(1&2), 59–80.
Eagly, A. H., & Antonakis, J. (2014). Leadership. In G. Borgida, & J. Bargh (Eds.), *APA Handbook of Personality and Social Psychology: Attitudes and Social Cognition* (Vol. 1, pp. 571-592). Washington: APA Books, http://dx.doi.org/10.1037/14341-018.
Eden, D., & Leviatan, U. (1975). Implicit leadership theory as a determinant of the factor structure underlying supervisory behavior scales. *Journal of Applied Psychology*, *60*(6), 736–741.
Edwards, J. R., & Parry, M. E. (1993). On the use of polynomial regression equations as an alternative to difference scores in organizational research. *Academy of Management Journal*, *36*(6), 1577–1613.
Egghe, L., & Rousseau, R. (2006). Theory and practise of the g-index. *Scientometrics*, *69*(1), 131–152.
Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of Management Journal*, *50*(1), 25–32.
Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299.
Falk, A., & Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science*, *326*(5952), 535–538.
Fiedler, F. E. (1967). *A theory of leadership effectiveness.* New York: McGraw-Hill.
Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*(4), 466–491.
Foster, E. M. (2010). Causal inference and developmental psychology. *Developmental Psychology*, *46*(6), 1454–1480.
Foster, E. M., & McLanahan, S. (1996). An illustration of the use of instrumental variables: Do neighborhood conditions affect a young person's change of finishing high school? *Psychological Methods*, *1*(3), 249–260.
Gagne, P., & Hancock, G. R. (2006). Measurement model quality, sample size, and solution propriety in confirmatory factor models. *Multivariate Behavioral Research*, *41*(1), 65–83.
Gardner, W. L., Lowe, K. B., Moss, T. W., Mahoney, K. T., & Cogliser, C. C. (2010). Scholarly leadership of the study of leadership: A review of the leadership quarterly's second decade, 2000–2009. *The Leadership Quarterly*, *12*(6), 922–958.
Garfield, E. (1982). The ethics of scientific publication: Authorship attribution and citation amnesia. *Current Contents*, *30*, 622–626.
Gennetian, L. A., Magnuson, K., & Morris, P. A. (2008). From statistical associations to causation: What developmentalists can learn from instrumental variables techniques coupled with experimental data. *Developmental Psychology*, *44*(2), 381–394.
Gephart, R. P. (2004). Qualitative research and the Academy of Management Journal. *Academy of Management Journal*, *47*(4), 454–462.
Gilbert, G. N. (1977). Referencing as persuasion. *Social Studies of Science*, *7*(1), 113–122.
Glaser, B. G., & Strauss, A. L. (1968). *The discovery of grounded theory: Strategies for qualitative research.* London: Weidenfeld and Nicolson.
Glick, W. H., Miller, C. C., & Cardinal, L. B. (2007). Making a life in the field of organization science. *Journal of Organizational Behavior*, *28*(7), 817–835.
Goodhue, D. L., Lewis, W., & Thompson, R. (2012). Does PLS have advantages for small sample size or non-normal data? *MIS Quarterly*, *36*(3), 981–1001.
Goodhue, D. L., Thompson, R., & Lewis, W. (2013). Why you shouldn't use PLS: Four reasons to be uneasy about using PLS in analyzing path models. *2013 46th Hawaii International Conference on System Sciences* (pp. 4739–4748).
Graen, G. B., & Uhl-bien, M. (1995). Relationship-based approach to leadership: Development of leader-member exchange (LMX) theory of leadership over 25 years: Applying a multi-level multi-domain perspective. *The Leadership Quarterly*, *6*(2), 219–247.
Halaby, C. N. (2004). Panel models in sociological research: Theory into practice. *Annual Review of Sociology*, *30*, 507–544.
Hambrick, D. C. (2007). The field of management's devotion to theory: Too much of a good thing? *Academy of Management Journal*, *50*(6), 1346–1352.
Hamilton, B. H., & Nickerson, J. A. (2003). Correcting for endogeneity in strategic management research. *Strategic Organization*, *1*(1), 51–78.
Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, *50*, 1029–1054.
Harrison, J. R., Lin, Z., Carroll, G. R., & Carley, K. M. (2007). Simulation modeling in organizational and management research. *Academy of Management Review*, *32*(4), 1229–1245.
Hastie, R., & Dawes, R. M. (2001). *Rational choice in an uncertain world: the psychology of judgment and decision making.* Thousand Oaks, Calif.: Sage.
Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, *46*(6), 1251–1271.
Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*(1), 153–161.
Henseler, J. (2010). On the convergence of the partial least squares path modeling algorithm. *Computational Statistics*, *25*(1), 107–120.
Herzog, W., & Boomsma, A. (2009). Small-sample robust estimators of noncentrality-based and incremental model fit. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(1), 1–27.
House, R. J. (1977). A 1976 theory of charismatic leadership. In J. G. Hunt, & L. L. Larson (Eds.), *The Cutting Edge*. Carbondale: Southern Illinois: University Press.
House, R. J., & Aditya, R. N. (1997). The social scientific study of leadership: Quo vadis? *Journal of Management*, *23*(3), 409–473.
House, R. J., & Mitchell, T. R. (1974). Path–goal theory of leadership. *Journal of Contemporary Business*, *3*(4), 81–97.
Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, *35*, 73–101.
James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis: Assumptions, models, and data.* Beverly Hills: Sage Publications.
Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology*, *87*(4), 765–780.
Judge, T. A., Cable, D. M., Colbert, A. E., & Rynes, S. L. (2007). What causes a management article to be cited — Article, author, or journal? *Academy of Management Journal*, *50*(3), 491–506.
Judge, T. A., Piccolo, R. F., & Ilies, R. (2004). The forgotten ones? The validity of consideration and initiating structure in leadership research. *Journal of Applied Psychology*, *89*(1), 36–51.
Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook.* Upper Saddle River, NJ: Pearson.
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.
Larcker, D. F., & Rusticus, T. O. (2010). On the use of instrumental variables in accounting research. *Journal of Accounting and Economics*, *49*(3), 186–205.
Lee, Y. -t, & Antonakis, J. (2012). When preference is not satisfied but the individual is: How power distance moderates person–job fit. *Journal of Management*.
Leigh, J. P. (1988). Assessing the importance of an independent variable in multiple-regression—Is stepwise unwise? *Journal of Clinical Epidemiology*, *41*(7), 669–677.
Levitt, S. D. (1997). Using electoral cycles in police hiring to estimate the effects of police on crime. *American Economic Review*, *87*(3), 270–290.
Levitt, S. D. (2002). Using electoral cycles in police hiring to estimate the effects of police on crime: Reply. *American Economic Review*, *92*(4), 1244–1250.
Li, M. (2012). Using the propensity score method to estimate causal effects: A review and practical guide. *Organizational Research Methods*.
Liden, R. C., & Antonakis, J. (2009). Considering context in psychological leadership research. *Human Relations*, *62*(11), 1587–1605.
Ligon, G. S., Harris, D. J., & Hunter, S. T. (2012). Quantifying leader lives: What historiometric approaches can tell us. *The Leadership Quarterly*, *23*(6), 1104–1133.
Lim, B. C., & Ployhart, R. E. (2004). Transformational leadership: Relations to the five-factor model and team performance in typical and maximum contexts. *Journal of Applied Psychology*, *89*(4), 610–621.
Lokker, C., McKibbon, K. A., McKinlay, R. J., Wilczynski, N. L., & Haynes, R. B. (2008). Prediction of citation counts for clinical articles at two years using data available within three weeks of publication: Retrospective cohort study. *BMJ*, *336*(7645), 655–657.
Long, J. S., & Freese, J. (2006). *Regression models for categorical dependent variables using Stata* (2nd ed.)College Station, TX: StataCorp LP.
Lord, R. G., De Vader, C. L., & Alliger, G. M. (1986). A Meta-analysis of the relation between personality traits and leadership perceptions: An application of validity generalization procedures. *Journal of Applied Psychology*, *71*(3), 402–410.
Lord, R. G., Foti, R. J., & De Vader, C. L. (1984). A test of leadership categorization theory: Internal structure, information processing, and leadership perceptions. *Organizational Behavior and Human Performance*, *34*, 343–378.
Lord, R. G., & Maher, K. J. (1991). *Leadership and information processing: Linking perceptions and performance.* Boston: Unwin Hyman.
Lowe, K. B., & Gardner, W. L. (2000). Ten years of the leadership quarterly: Contributions and challenges for the future. *The Leadership Quarterly*, *11*(4), 459–514.
Macrae, D. (1969). Growth and decay curves in scientific citations. *American Sociological Review*, *34*(5), 631–635.

Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics.* Cambridge: Cambridge University Press.
Marcoulides, G. A., & Saunders, C. (2006). PLS: A silver bullet? *MIS Quarterly*, *30*(2), III–IX.
Maxwell, J. A. (2010). Using numbers in qualitative research. *Qualitative Inquiry*, *16*(6), 475–482.
McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.)London, New York: Chapman and Hall.
McDonald, R. P. (1996). Path analysis with composite variables. *Multivariate Behavioral Research*, *31*(2), 239–270.
Meho, L. I. (2007). The rise and rise of citation analysis. *Physics World*, *20*(1), 32–36.
Meho, L. I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus Scopus and Google scholar. *Journal of the American Society for Information Science and Technology*, *58*(13), 2105–2125.
Meindl, J. R., Ehrlich, S. B., & Dukerich, J. M. (1985). The romance of leadership. *Administrative Science Quarterly*, *30*(1), 78–102.
Merton, R. K. (1968). Matthew effect in science. *Science*, *159*(3810), 56-&.
Mitchell, G. (2012). Revisiting truth or triviality. *Perspectives on Psychological Science*, *7*(2), 109–117.
Mooney, C. Z. (1997). *Monte Carlo simulation.* Thousand Oaks, Calif.: Sage Publications.
Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles for social research.* New York: Cambridge University Press.
Mundlak, Y. (1978). Pooling of time-series and cross-section data. *Econometrica*, *46*(1), 69–85.
Murray, C. (2005). How to accuse the other guy of lying with statistics. *Statistical Science*, *20*(3), 239–241.
Muthén, B. O. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics*, *22*(1–2), 43–65.
Muthén, B. O. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*(1), 115–132.
Muthén, B. O., du Toit, S. H. C., & Spisic, D. (2013). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. *Psychometrika* (in press).
Muthen, L. K., & Muthen, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, *9*(4), 599–620.
Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.)Los Angeles, CA: Muthén & Muthén.
Muthén, B. O., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, *55*(2), 463–469.
Newman, M. E. J. (2005). Power laws, Pareto distributions, and Zipf's law. *Contemporary Physics*, *46*(5), 323–351.
Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220.
Osborn, R. N., Hunt, J. G., & Jauch, L. R. (2002). Toward a contextual theory of leadership. *The Leadership Quarterly*, *13*(6), 797–837.
Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.)Thousand Oaks, Calif: Sage Publications.
Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen, F. N. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling: A Multidisciplinary Journal*, *8*(2), 287–312.
Pearl, J. (2000). *Causality: Models, reasoning, and inference.* New York: Cambridge University Press.
Persson, O. (2010). Are highly cited papers more international? *Scientometrics*, *83*(2), 397–401.
Pfeffer, J. (1993). Barriers to the advance of organizational science: Paradigm development as a dependent variable. *The Academy of Management Review*, *18*(4), 599–620.
Porter, L. W., & McLaughlin, G. B. (2006). Leadership and the organizational context: Like the weather? *The Leadership Quarterly*, *17*(6), 559–576.
Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 717–731.
QS World University Ranking (2008-2012). http://www.topuniversities.com/qs-world-university-rankings (accessed May 2013)
Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika*, *69*, 167–190.
Richardson, H. A., Simmering, M. J., & Sturman, M. C. (2009). A tale of three perspectives: Examining post hoc statistical techniques for detection and correction of common method variance. *Organizational Research Methods*, *12*(4), 762–800.
Rönkkö, M., & Evermann, J. (2013). A critical examination of common beliefs about partial least squares path modeling. *Organizational Research Methods*, *16*, 425–448.
Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701.
Rubin, D. B. (2008). For objective causal inference, design trumps analysis. *Annals of Applied Statistics*, *2*(3), 808–840.
Rynes, S. L., Giluk, T. L., & Brown, K. G. (2007). The very separate worlds of academic and practitioner periodicals in human resource management: Implications for evidence-based management. *Academy of Management Journal*, *50*(5), 987–1008.
Scandura, T. A., & Williams, E. A. (2000). Research methodology in management: Current practices, trends, and implications for future research. *Academy of Management Journal*, *43*(6), 1248–1264.
Schriesheim, C. A. (1995). Multivariate and moderated within- and between-entity analysis (WABA) using hierarchical linear multiple regression. *The Leadership Quarterly*, *6*(1).
Seglen, P. O. (1992). The skewness of science. *Journal of the American Society for Information Science*, *43*(9), 628–638.
Shadish, W. R., & Cook, T. D. (1999). Comment-design rules: More steps toward a complete theory of quasi-experimentation. *Statistical Science*, *14*(3), 294–300.
Shadish, W. R., & Cook, T. D. (2009). The renaissance of field experimentation in evaluating interventions. *Annual Review of Psychology*, *60*, 607–629.
Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton Mifflin.
Shaver, J. M. (2005). Testing for mediating variables in management research: Concerns, implications, and alternative strategies. *Journal of Management*, *31*(3), 330–353.
Shipley, B. (2000). *Cause and correlation in biology: A user's guide to path analysis, structural equations, and causal inference.* : Cambridge University Press (Cambridge, UK; New York, NY. USA).
Simonton, D. K. (2003). Qualitative and quantitative analyses of historical data. *Annual Review of Psychology*, *54*, 617–640.
Staiger, D., & Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, *65*(3), 557–586.
StataCorp (2013). *Stata Statistical Software: Release 13.* College Station: StataCorp LP.
Steiger, J. H. (2001). Driving fast in reverse — The relationship between software development, theory, and education in structural equation modeling. *Journal of the American Statistical Association*, *96*(453), 331–338.
Swygart-Hobaugh, A. J. (2004). A citation analysis of the quantitative/qualitative methods debate's reflection in sociology research: Implications for library collection development. *Library Collections Acquisitions & Technical Services*, *28*(2), 180–195.
Thompson, S. P. (1976). *The life of Lord Kelvin* (2nd ed.)New York, N.Y.: Chelsea Pub. Co.
Thompson, B. (1995). Stepwise regression and stepwise discriminant-analysis need not apply here — A guidelines editorial. *Educational and Psychological Measurement*, *55*(4), 525–534.
Twain, M. (1906). *Chapters from my autobiography. The Project Gutenberg EBook*.
van Raan, A. F. J. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics*, *67*(3), 491–502.
Van Seters, D. A., & Field, R. H. G. (1990). The evolution of leadership theory. *Journal of Organizational Change Management*, *3*(3).
Van Vugt, M., Hogan, R., & Kaiser, R. B. (2008). Leadership, followership, and evolution — Some lessons from the past. *American Psychologist*, *63*(3), 182–196.
Vermeulen, F. (2005). On rigor and relevance: Fostering dialectic progress in management research. *Academy of Management Journal*, *48*(6), 978–982.
Waldman, D. A., Balthazard, P. A., & Peterson, S. J. (2011). Leadership and neuroscience: Can we revolutionize the way that inspirational leaders are identified and developed? *Academy of Management Perspectives*, *25*(1), 60–74.
Whittingham, M. J., Stephens, P. A., Bradbury, R. B., & Freckleton, R. P. (2006). Why do we still use stepwise modelling in ecology and behaviour? *Journal of Animal Ecology*, *75*(5), 1182–1189.
Yammarino, F. J. (1998). Multivariate aspects of the varient/WABA approach: A discussion and leadership illustration. *The Leadership Quarterly*, *9*(2).
Yin, R. K. (1994). *Case study research: Design and methods.* Thousand Oaks, CA: Sage Publications.
Zaccaro, S. J. (2001). *The nature of executive leadership: A conceptual and empirical analysis of success.* Washington, D.C.: American Psychological Association.