Getting Through the Gate: Statistical and Methodological Issues Raised in the Reviewing Process

Jennifer P. Green[1], Scott Tonidandel[2] and Jose M. Cortina[1]

Abstract
This study empirically examined the statistical and methodological issues raised in the reviewing process to determine what the "gatekeepers" of the literature, the reviewers and editors, really say about methodology when making decisions to accept or reject manuscripts. Three hundred and four editors' and reviewers' letters for 69 manuscripts submitted to the *Journal of Business and Psychology* were qualitatively coded using an iterative approach. Systematic coding generated 267 codes from 1,751 statements that identified common methodological and statistical errors by authors and offered themes across these issues. We examined the relationship between the issues identified and manuscript outcomes. The most prevalent methodological and statistical topics were measurement, control variables, common method variance, factor analysis, and structural equation modeling. Common errors included the choice and comprehensiveness of analyses. This qualitative analysis of methods in reviews provides insight into how current methodological debates reveal themselves in the review process. This study offers guidance and advice for authors to improve the quality of their research and for editors and reviewers to improve the quality of their reviews.

Keywords
qualitative research, method variance, factor analysis, interactions, construct validation procedures

Organizational, management, and applied psychology research has always placed a great deal of emphasis on research methodology, with many literature reviews examining trends and frequencies of particular methods over the years (Aguinis, Pierce, Bosco, & Muslin, 2009; Austin, Scherbaum, & Mahlman, 2002; Scandura & Williams, 2000; Stone-Romero, Weaver, & Glenar, 1995). While it is essential to examine what is published in the literature, it is equally critical to understand how methods

[1]Department of Psychology, George Mason University, Fairfax, VA, USA
[2]Department of Psychology, Davidson College, Davidson, NC, USA

Corresponding Author:
Jennifer P. Green, Department of Psychology, George Mason University, 4400 University Drive, Fairfax, VA 22030, USA.
Email: jgreen24@gmu.edu

are evaluated in the manuscript review process, as editors and reviewers ultimately decide which methods are acceptable for publications. Previous opinion-based articles by editors have emphasized the importance of research methods in the manuscript review process (Bono & McNamara, 2011; Desrosiers, Sherony, Barros, Ballinger, Senol, & Campion, 2002); however, few studies have empirically examined the statistical and methodological issues raised by reviewers and the influence of these specific issues in editors' manuscript decisions. As editors and reviewers are the gatekeepers of journals, it is important to answer the following question: Are there particular methodological and statistical issues that are more prominent, and thus raised more often, in the reviewing process? The purpose of the present study is to expand our understanding of how editors and reviewers approach research methods and statistics through a quantitative and detailed qualitative analysis of peer reviews. Specifically, we used an iterative approach to code editors' and reviewers' letters for the methodological and statistical issues that they raised concerning the manuscripts and the suggestions that they provided to authors. The specific codes and themes identified provide insight into the peer review process and give authors guidance to improve the quality of their research in their own submissions.

Past Research

Relatively few studies have examined empirically the peer review process in terms of the editors' and reviewers' decisions (Fogg & Fiske, 1993; Gilliland & Cortina, 1997). Fiske and Fogg (1990) used content analysis to analyze reviewers' reports for 153 papers submitted for the first time to *American Psychological Association* journals in late 1985 and in 1986. They coded reviews of manuscripts as well as the editors' letters to the authors for every critical point of weakness. These codes were categorized into areas such as Introduction, Design and Procedures, Methods and Results, and Discussion. In addition, the authors also coded the comments into one of the two categories regarding the research process: (a) planning and execution of the research and (b) presentation of the manuscript. A follow-up study by Fogg and Fiske (1993) tested the relationship between these two categories and editors' decisions and reviewers' recommendations. They found a significant correlation between severe planning and execution criticisms and reviewers' final recommendations, demonstrating the importance of research methods and advanced planning in research design (Fogg & Fiske, 1993). However, the authors did not examine relationships between specific methodological issues, as listed in Fiske and Fogg (1990), and manuscript decisions (Fogg & Fiske, 1993).

Similar to Fogg and Fiske's (1993) finding that severe planning and execution held importance in reviewers' recommendations, Gilliland and Cortina (1997) found that reviewers' research design evaluations were the most predictive of their recommendations, followed by evaluations of conceptual or theoretical arguments, operationalization of constructs, appropriateness of topic, and adequacy of data analysis. In addition to the analyses of reviewers' evaluations (i.e., numerical ratings on general manuscript dimensions), the authors coded manuscript submissions for statistical techniques (e.g., factor analysis, analysis of variance) in order to determine what analytical factors were most predictive of manuscript decisions (Gilliland & Cortina, 1997). Manuscripts that predominantly used factor analysis or analysis of variance received less favorable recommendations than papers that relied on correlations, regression, Linear Structural Relations (LISREL), confirmatory factor analysis (CFA), path analysis, or other methods. A limitation of this paper mentioned by its authors was that the content of reviews and decision letters were not examined (Gilliland & Cortina, 1997). This limits the ability to determine if reviewers and editors raised issues pertaining to these analytical techniques and if such issues had a subsequent impact on their recommendations.

To date, a qualitative analysis by Rogelberg, Adelman, and Askay (2009) provides the most detailed account of issues raised in the peer review process. Through content analysis of 131 reviewers' comments for nearly 100 manuscripts submitted to the *Journal of Business and Psychology* over a four-month time period, the authors identified areas of concern commonly expressed by reviewers. The

general categories established by Rogelberg and colleagues were similar to that of Fiske and Fogg's (1990) categories such as Methods and Results, Analyses, and Discussion. Rogelberg et al. also provided illustrative comments for each general category. Although Fiske and Fogg (1990) included subcategories such as "a statistical error is specified, and a better statistic is suggested or implied" and "a specified statistical analysis should be added" (p. 593), they failed to provide what additional analyses were requested. In addition to stating that reviewers recommended adding statistical analyses, Rogelberg et al. provided specific examples of those analyses that reviewers requested, such as betas for a regression or using the Sobel test for mediation. Rogelberg et al.'s content analysis of reviewers' comments provides the most detailed empirical assessment of reviewers' comments to date, but it does not identify themes among these manuscript issues, nor does it compare the frequency of issues in comments to manuscript decisions to reveal how reviewers and editors evaluate these issues.

The Current Study

The current study builds on previous work by (a) expanding on the specificity of statistical and methodological issues raised in reviews while also providing unifying themes for these topics, (b) relating both these specific issues and general themes to editors' decisions, and (c) linking reviewers' and editors' comments to methodological debates in the literature. The present study used systematic procedures to qualitatively analyze editors' and reviewers' letters. This approach allowed for a more extensive and detailed qualitative analysis of peer reviews compared to previous analyses of manuscripts and reviewers' comments (e.g., Fiske & Fogg, 1990). We also expanded on the methodological issues and suggestions raised by Rogelberg et al. (2009) by not only providing common methodological issues raised in peer reviews but also offering unifying themes for these issues. We suggest that it is important to examine methodological comments with both a level of specificity and a level of generality as they provide unique information to understand what reviewers and editors care about. Specific issues can inform authors about particular errors to avoid and techniques to use in order to address these issues. Themes can help to understand reviewers' and editors' perspectives, such as the request for comprehensive analyses, that might be driving specific comments about particular issues, enabling us to provide advice to authors that can be utilized more broadly to other methods and statistical analyses.

Many researchers have analyzed and coded journal articles to examine methodological challenges acknowledged in articles (e.g., Aguinis & Lawal, 2012; Brutus, Aguinis, & Wassmer, 2013) and to see if articles' methods and statistics reflect best practices (e.g., Becker, 2005; Carlson & Wu, 2012; Cortina, 2003). To our knowledge, no study has related gatekeepers' comments to best practices. To provide authors with recommendations to improve their quality of research, and thus enhance their manuscript submissions, we not only provide findings from our iterative coding of reviewers' and editors' letters but also relate these comments to current debates and practices in the literature. Our recommendations to address common methodological and statistical issues raised in reviews should also help reviewers and editors with their reviews and decisions.

Method

Sample

The sample consisted of reviews for all relevant first submissions of manuscripts to the *Journal of Business and Psychology* (*JBP*) whose date of first decision was between July 2011 and December 2011.[1] Manuscripts that were solely literature reviews, focused exclusively on methodological procedures, or submitted for special features were excluded from the sample. The final sample from 69 manuscripts consisted of 304 letters including 138 letters from reviewers to authors, 97 letters from reviewers to editors, and 69 letters from editors to authors. Forty-one reviewers did not provide specific comments for the editor separate from their comments to the authors. Thirteen different

Table 1. Correlation Matrix of Code Families Raised in Manuscripts.

Code Family	1	2	3	4	5	6	7	8	9	10	11
1. Theory	1.00										
2. Design	.21**	1.00									
3. Measurement	.28**	.29**	1.00								
4. Control variables	.21**	.23**	.19**	1.00							
5. Common method variance	.08	.32**	.21**	.26**	1.00						
6. Correlated variables	.05	.17**	.12*	.16**	.10	1.00					
7. Moderation	.10	.20**	.19**	.11*	.14*	.19**	1.00				
8. Mediation	.02	.14**	.04	.12*	.08	.21**	.17**	1.00			
9. HLM	.01	.11*	.17**	.20**	−.05	.13*	.09	.08	1.00		
10. Factor analysis/SEM	−.02	.24**	.27**	.13*	.22**	−.01	.15**	.10	.11*	1.00	
11. Moderated regression	.10	.05	.18**	.12*	.09	.10	.61**	.14**	.06	.16**	1.00

Note: $N = 69$ manuscripts. A code family was present in a manuscript if at least one issue falling within the family was raised in an editor's and/or reviewer's letter. HLM = hierarchical linear modeling; SEM = structural equation modeling.
*Correlation is significant at the .05 level (two-tailed). **Correlation is significant at the .01 level (two-tailed).

editors and 107 reviewers reviewed the manuscripts. Of the 107 reviewers, 27 reviewed two manuscripts, 2 reviewed three manuscripts, and 78 reviewed only one manuscript in the sample.

Although our sample is drawn from the comments of editors and reviewers from one journal, their ideas likely reflect those of many other journals. It is typical for associate editors and the editorial board to also serve other journals. Therefore, the thoughts of editors and reviewers reflected in *JBP* likely reflect the thoughts of editors and reviewers from other journals. The 165 editorial board members at *JBP* during the time of the current study now comprise 24.47% ($n = 58$), 13.04% ($n = 33$), 15.07% ($n = 11$), 4.67% ($n = 14$), and 16.05% ($n = 13$) of the current editorial boards at *Journal of Applied Psychology*, *Journal of Management*, *Organizational Research Methods*, *Academy of Management Journal*, and *Personnel Psychology*, respectively. Additionally, between 20% and 30% of the current associate editors at *Journal of Applied Psychology*, *Journal of Management*, *Organizational Research Methods*, *Personnel Psychology*, and *Psychological Methods* were *JBP* board members at the time of this study. Therefore, our findings regarding editors' and reviewers' comments from *JBP* can likely be generalized to the comments of reviewers and editors at other management or applied psychology journals.

For each peer-reviewed manuscript, quantitative data were also gathered. The editor's decision (reject the article or revise and resubmit) was recorded for each of the 69 manuscripts. Of the 69 manuscripts, there were 69 manuscript decisions made by 13 editors regarding the outcome of the manuscript. We used logistic regression to see if there were differences across editors and reviewers with regard to issues raised and manuscript rejection rate. Logistic regression analyses suggest that there were no editor or reviewer effects. With 12 editor dummy codes as predictors, we ran a logistic regression with the manuscript outcome (reject or revise and resubmit) as the criterion. We also ran 11 logistic regressions with the presence or absence of each manuscript issue (for all 11 code families, see Table 1) as the criterion. For all of these analyses, we found nonsignificant regression weights, meaning that the editor reviewing the manuscript did not significantly predict whether an issue was raised, nor did it predict the manuscript outcome. Additionally, we ran similar logistic regressions for reviewers and did not find any significant regression weights for dummy-coded reviewers. Thus, for all common methodological and statistical issues, the editor or reviewer reviewing the manuscript did not significantly predict if an issue was raised or not, nor did the manuscript rejection rate depend on the editor.

Procedure

This study examined the methodological issues most noteworthy to editors and reviewers through qualitative analysis of all 304 editors' and reviewers' letters (approximately five documents for each manuscript). The inductive approach involved systematic procedures to analyze editors' and reviewers' letters and further define these gatekeepers' views on methodology.

ATLAS.ti. To code methodological issues in the 304 letters, the present study used ATLAS.ti, a computer program that assists with the analysis and coding of text documents by organizing codes into categories and superordinate categories (Pollach, 2011; Weitzman & Miles, 1995). Once the documents are uploaded into the software, the user can code phrases or words, reuse these codes later, and create links among codes. Relations for links between codes can be assigned logical properties using preprogrammed semantic relations or original relations, such as "is associated with" or "is part of" (ATLAS.ti, n.d.; Weitzman & Miles, 1995). The codes can also be grouped into categories, and these categories can be further grouped into families. Additionally, ATLAS.ti provides quotation counts for the various codes, meaning that ATLAS.ti output shows how many documents have a particular code and how many times each document has the code. In the present study, we uploaded each of the 304 reviewers' and editors' letters into ATLAS.ti as separate documents.

Qualitative coding. Through a constant comparative method (Locke, 2002), the current study coded the reviewers' and editors' letters to build conceptual categories, general themes, and overarching dimensions about research methods and statistics in the peer review process. Other examples of this iterative type of approach can be found in Clark, Gioia, Ketchen, and Thomas (2010); Pratt (2000); and Rerup and Feldman (2011). The analytic process started small, first with open coding of issues regarding research methods and then building the issues into conceptual categories, which could then be developed into themes. In first-order, or open, coding, researcher expectations do not play a role, such that the codes are specific, detailed, and directly represent the text (Gioia, Corley, & Hamilton, 2013; Strauss & Corbin, 1994). For instance, a quote from one editors' letter read, "It is not appropriate to conduct both exploratory and confirmatory factor analyses on the same sample." In the open coding stage of the analysis, this quote was coded as "do not conduct EFA [exploratory factor analysis] and CFA on the same data set." As additional letters were coded, other comments relating to CFA emerged, such as "why EFA instead of CFA," and "only need CFA not EFA for established measure." The open coding process generated 1,751 statements from the letters. The same statement could be counted more than once because statements could be coded into multiple categories. As an example, one reviewer's comment regarding a one-item control variable reflected concerns about single-item measures, scale choice, and a potential missing control variable. In order to reflect all these issues raised by the reviewer, this comment was assigned multiple open codes, including "inappropriate choice of measure" and "failure to include control variable."

This study used two coders to analyze the data. After coding 25 letters (approximately 5 manuscripts) with these specific, first-order codes directly representing the text, both coders reviewed the code list and conducted focused coding. Focused coding, or second-order coding, requires the grouping and categorization of similar codes into concepts (Lee, Mitchell, & Harman, 2011). Referring to the previous example of codes, "do not conduct EFA and CFA on the same data set," "why EFA instead of CFA," and "only need CFA not EFA for established measure," the coders categorized these codes into an overarching conceptual code called "choice of factor analysis." This focused coding enabled a comprehensive organization of the codes into conceptual categories. During this focused coding, any discrepancies between the coders regarding the open codes were discussed and resolved prior to comparing, contrasting, and subsequent organization of the codes. All categories and grouping

of codes were based on consensus between the two coders. Consensus coding is a commonly used procedure for iterative coding (e.g., Koppman & Gupta, 2014; Stigliani & Ravasi, 2012).

After focused coding was complete for a set of manuscript documents (i.e., letters), open coding would continue for 25 more letters (approximately 5 manuscripts). As the open coding continued on these new letters, the coder could use previous conceptual categories to code methodological and statistical comments within the editors' and reviewers' letters or could also create new first-order codes if none of the existing categories fit the data. ATLAS.ti software helped revise previous codes such that the renaming or combining of first-order codes in later letters would be applied to all letters that had those codes, including earlier letters. The coding of new letters helped to evaluate the current categories and reveal if there were new dimensions not yet articulated (Lee et al., 2011; Locke, 2002). After the coding of the new letters was complete, the coders would meet again to discuss new open codes and reevaluate the coding categories. Continuous comparison of codes and conceptual categories are an integral part of inductive research in order to refine the categories and ensure they reflect the data (Frost, 2011). This process continued until all of the 304 letters were analyzed and there were no new first-order, open codes or codes that did not fit into existing categories. The two coders used consensus throughout the categorization process to generate 267 second-order codes from a total of 1,751 coded statements.

After all the manuscript letters were coded, the coders evaluated the 267 final second-order codes and organized them into a broader framework using the families function in ATLAS.ti (Locke, 2002). Families allow the combination of categories into overarching groups. For instance, the "choice of factor analysis" related to other categories, such as "compare models" and "techniques to improve fit." These could be grouped into the family of "Factor Analysis/SEM." Other themes are further described in the results section but include, for example, common method variance, measurement, and control variables. Figure 1 provides an example of the reduction of codes from first-order, open codes to second-order categories and finally, to families or themes. In order to further create a comprehensive framework, comments within families regarding data analysis were grouped into three larger overarching dimensions: comprehensiveness of analyses, choice of analyses, and data analytic errors.

The families and three overarching dimensions represent the core patterns of research method comments in reviews. The following description of the findings from the coding process is organized to reflect these families and themes. Additionally, we discuss the ATLAS.ti output that provided frequency counts for the codes in the letters. These frequency counts were used to relate letter content to the editors' final decisions on the manuscripts, providing further insight about the impact of these topics on manuscript outcomes. In our discussion of these common issues and themes, we relate editors' and reviewers' comments to best practices as recommended by the literature to provide guidance and recommendations to authors for future manuscript submissions.

Results

Overview

Systematic coding for methodological and statistical comments in reviewers' and editors' letters resulted in 267 conceptual codes from 1,751 statements that were further reduced into 11 families. Reviewers' and editors' comments for topics spanning design, measurement, and analysis typically referred to perceptions of authors' mistakes or suggestions for improvement. Occasionally, these comments represented praise; however, comments of praise rarely provided specific methodological and statistical detail that could be coded. Table 1 provides a correlation matrix of the presence of issues in code families (e.g., measurement, moderation) in the manuscripts. A code family was present in a manuscript if at least one issue falling within the family was raised in an editor's and/or reviewer's letter. In general, the presence of an issue in one code family in an editor's and/or reviewer's letter for a manuscript positively related to the presence of issues in other code families for that manuscript.
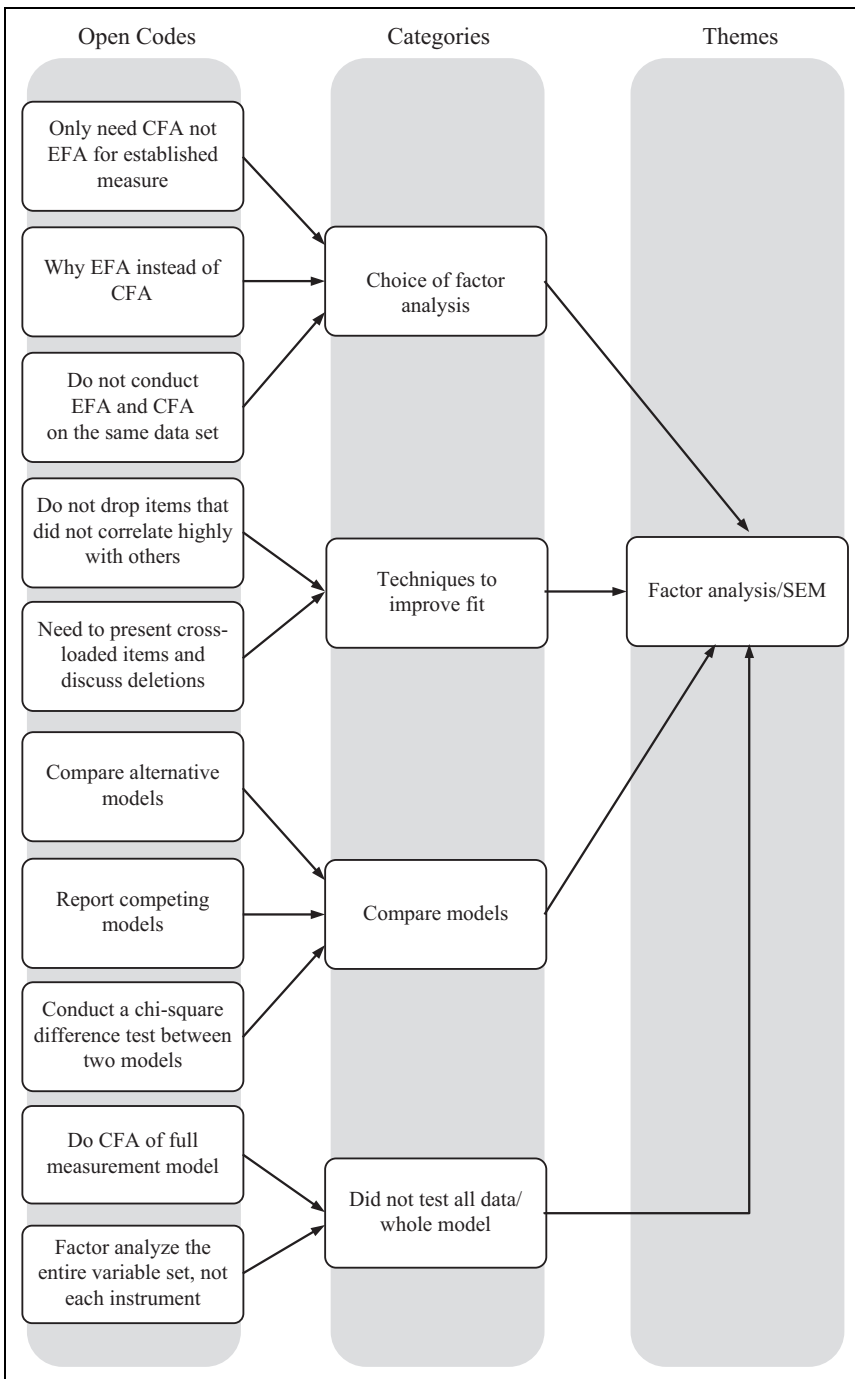
Figure 1. An example of the constant comparative qualitative coding process.

We focus our analyses on particular categories and families because these issues were most frequently raised in the sample of manuscripts. The frequencies of methodological and statistical issues raised in reviewers' and editors' letters are presented in Table 2. It is important to note that

frequency refers to whether a reviewer or editor mentioned an issue pertaining to a topic in his or her letter, not the number of times the issue was mentioned in an individual letter. Although the content of the manuscripts themselves was not coded, an issue was considered present in a manuscript if it was raised in an editor's and/or reviewer's letter. Our focus on the more common issues raised does not mean that they are necessarily the most important issues in terms of a manuscript's outcome as a rarely raised issue could actually be more critical in the outcome than these common issues. We do nevertheless link the presence of these common issues in manuscript reviews to the final manuscript decision, as summarized in Table 2.[2]

As typical in the peer review process, editors rejected more first-time manuscript submissions than they gave revise and resubmit, with editors' initial decisions reflecting a rejection rate of 75.36% of the manuscripts ($n = 52$). For comparison purposes, Table 2 not only includes the rejection rate for those manuscripts when an issue was present in an editor's and/or reviewer's letter but also the rejection rate for those manuscripts when the issue was not present. Due to the uneven frequency in manuscripts where the issue was present versus not present and the low base rates for certain issues, comparisons among these rates are difficult. As a result, we only make comparisons between these rejection rates when the number of manuscripts in which the issue was present was similar to the number of manuscripts in which the issue was not present, as was the case for common method variance (CMV) and factor analysis. To provide further information regarding these topics and manuscript decisions, Table 3 provides the percentages of the 52 rejected manuscripts and the percentages of the 17 revise and resubmit manuscripts in which issues related to these topics were present. In the Discussion section, we relate editors' and reviewers' comments to current methodological discussions in the literature in order to offer suggestions and recommendations to improve the research quality of future manuscript submissions (see Table 4).

Three prevalent families including measurement, control variables, and CMV, along with relevant categories, are described in the following. In addition to these three families, six other families of codes of frequent data analytic techniques, including factor analysis/ SEM, correlated variables, hierarchical linear modeling (HLM), moderation, mediation, and regression, are discussed as they pertain to three overarching data analytic themes, including comprehensiveness of analyses, choice of analyses, and common errors in data analysis. Other families of codes, including theory and study design, resulted from the coding process. Although our primary focus pertains to methods and statistical techniques, we will briefly discuss the common issues pertaining to study design.

Theory and Design

Comments pertaining to theory and design issues were raised in an editor's and/or reviewer's letter corresponding to 53.62% and 92.75% of the manuscripts, respectively. Typical remarks concerning theory pertained to weak theoretical bases for subsequent hypotheses. For example, an editor noted "the need for a stronger theoretical foundation for the paper as a whole . . . and for the development of specific hypotheses." Similarly, other concerns related to whether the design was capable of testing the theory behind the hypotheses. A reviewer reflected this mindset when expressing "worry that there is a bit of a disconnect between the context that is alluded to throughout the introduction and that represented by the methodology." In general, editors and reviewers were noting disconnect between theory, research questions, and research design.

Three of the most common design issues raised included threats to validity ($n = 15$ manuscripts), issues with causality ($n = 31$), and sampling ($n = 13$). Examples of threats to validity raised by editors and reviewers included history threat, regression to the mean, and selection biases. Causality issues primarily included comments that authors' designs did not allow conclusions regarding causal effects. One editor conveyed to the authors that it was necessary to test the causal link between two of their constructs but "unfortunately, no causal inferences can be drawn from your data because all measures

Table 2. Frequencies and Percentages of Issues Related to Methodological and Statistical Topics in Reviews and Editors' Decisions.

Topic	Appearance of Methodological or Statistical Issue						Rejection Rate			
	Editor's Letter		Reviewer Pair's Letters[a]		Manuscript[b]		Issue Present		Issue Not Present	
	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
Theoretical support	22	31.88	32	46.38	37	53.62	28	75.68	24	75.00
Design	34	49.28	60	86.96	64	92.75	48	75.00	4	80.00
Threats to validity	6	8.70	13	18.84	15	21.74	10	66.67	42	77.78
Causality	14	20.29	25	36.23	31	44.93	24	77.42	28	73.68
Sampling	6	8.70	11	15.94	13	18.84	11	84.62	41	73.21
Measurement	48	69.57	58	84.06	62	89.86	46	74.19	6	85.71
Theoretical justification	27	39.13	39	56.52	46	66.67	36	78.26	16	69.57
Conceptualization	19	27.54	27	39.13	29	42.03	22	75.86	30	75.00
Operationalization	28	40.58	41	59.42	46	66.67	33	71.74	19	82.61
Convergent/discriminant	26	37.68	34	49.28	39	56.52	29	74.36	23	76.67
Failure to use factor analysis	7	10.14	11	15.94	13	18.84	10	76.92	42	75.00
Choice of factor analysis	6	8.70	6	8.70	10	14.49	9	90.00	43	72.88
Control variables	16	23.19	23	33.33	29	42.03	22	75.86	30	75.00
Failure to include controls	15	21.74	21	30.43	27	39.13	20	74.07	32	76.19
Unnecessary controls	4	5.80	10	14.49	10	14.49	8	80.00	44	74.58
Common method variance	19	27.54	23	33.33	29	42.03	24	82.76	28	70.00
Correlated variables	9	13.04	15	21.74	15	21.74	8	53.33	44	81.48
Moderation	9	13.04	20	28.99	21	30.43	16	76.19	36	75.00
Test for moderators	3	4.35	7	10.14	7	10.14	6	85.71	46	74.19
Moderated mediation	2	2.90	4	5.80	4	5.80	4	100.00	48	73.85
Mediation	9	13.04	9	13.04	13	18.84	9	69.23	43	76.79
SEM to test for mediators	4	5.80	3	4.35	5	7.25	4	80.00	48	75.00
Sobel and bootstrapping	5	7.25	7	10.14	9	13.04	6	66.67	46	76.67
HLM	10	14.49	12	17.39	15	21.74	9	60.00	43	79.63
Factor analysis/SEM	22	31.88	31	44.93	37	53.62	31	83.78	21	65.63
Failure to use SEM	3	4.35	2	2.90	5	7.25	3	60.00	49	76.56
Model comparison	6	8.70	9	13.04	10	14.49	8	80.00	44	74.58
Methods to improve fit	6	8.70	6	8.70	7	10.14	6	85.71	46	74.19
Moderated regression	4	5.80	9	13.04	10	14.49	7	70.00	45	76.27
Centered variables	2	2.90	3	4.35	3	4.35	3	100.00	49	74.24
Simple slopes	3	4.35	7	10.14	8	11.59	5	62.50	47	77.05

Note. In total, editors rejected 75.36% (n = 52) of the manuscripts. HLM = hierarchical linear modeling; SEM = structural equation modeling.
[a]Reviewer pair means that the issue was raised by at least one of the two manuscript reviewers. [b]An issue was considered present in a manuscript if it was raised in an editor's and/or reviewer's letter.

Table 3. Percentage of Rejected/Revise and Resubmit Manuscripts Raising Each Issue.

Topic	Reject[a]	Revise and Resubmit[b]
Theoretical support	53.85	52.94
Design	92.31	94.12
Threats to validity	19.23	29.41
Causality	46.15	41.18
Sampling	21.15	11.76
Measurement	88.46	94.12
Theoretical justification	69.23	58.82
Conceptualization	42.31	41.18
Operationalization	63.46	76.47
Convergent/discriminant	55.77	58.82
Failure to use factor analysis	19.23	17.65
Choice of factor analysis	17.31	5.88
Control variables	42.31	41.18
Failure to include controls	38.46	41.18
Unnecessary controls	15.38	11.76
Common method variance	46.15	29.41
Correlated variables	15.38	41.18
Moderation	30.77	29.41
Test for moderators	11.54	5.88
Moderated mediation	7.69	0.00
Mediation	17.31	23.53
SEM to test for mediators	7.69	5.88
Sobel and bootstrapping	11.54	17.65
HLM	17.31	35.29
Factor analysis/SEM	59.62	35.29
Failure to use SEM	5.77	11.76
Model comparison	15.38	11.76
Methods to improve fit	11.54	5.88
Moderated regression	13.46	17.65
Centered variables	5.77	0.00
Simple slopes	9.62	17.65

Note: HLM = hierarchical linear modeling; SEM = structural equation modeling.
[a]Out of 52 rejected manuscripts. [b]Out of 17 revise and resubmit manuscripts.

were collected from a single source at one point in time." Data that cannot support causal predictions are an issue because, as one reviewer mentioned, there is "the likelihood of reverse causality." One issue raised with sampling was overall low response rate, while another reviewer expressed concern over a base-rate problem such that "it is difficult to draw conclusions about a phenomenon that is not widely manifested in the sample." Of these various design issues, the rejection rate for manuscripts that raised at least one design issue (75%) was not much different than the overall rejection rate of manuscripts in the sample (75.36%). However, regarding particular design comments, when an issue regarding sampling was raised, the manuscript rejection rate rose to 84.62%.

Measurement

The most frequently raised methodological concerns by editors and reviewers were those involving measurement. Comments for 62 of the 69 manuscripts (89.86%) discussed issues regarding construct measurement. These comments targeted choice of constructs, conceptualization and operationalization of constructs, divergent and convergent validity, and tests for measurement models.

Choice of construct. In total, reviews of 46 of 69 manuscripts (66.67%) discussed the need to justify constructs, suggesting that editors and reviewers are often unconvinced by author justifications for inclusion of constructs. Comments pertaining to the choice of construct and theoretical justification for constructs are consistent with Edwards's (2010) call to enhance theoretical progress in organizational research. In order to test meaningful propositions, Edwards and Berry (2010) suggest that authors not only use rigorous methods but also develop more precise theories. In the present study, when editors and reviewers criticized authors' choice of constructs, they typically asked that the authors theoretically justify their choices, including, specifically: why did they choose those particular constructs, why were these constructs chosen over others, and why did they use only some components of the construct. For example, one editor remarked that "Overall, a more convincing rationale is needed to justify your choice of outcome variables," and a reviewer commented that "I felt there was very little justification for why the problem investigated was important, and for why the variables selected were chosen."

Conceptualization and operationalization of constructs. Another common issue raised by reviewers and editors involved authors' conceptualization and operationalization of constructs. Issues regarding conceptualization and operationalization were mentioned in the editor's and/or reviewer's letters for 52 manuscripts (75.36%).

In terms of conceptualization, reviewers and editors expressed a desire for authors to provide a more detailed and clear definition of their constructs. Conceptualization issues were raised in reviews of 29 manuscripts (42.03%). A comment exemplifying this sentiment was, "The definition of [the construct] is not clearly defined and given this is a focal construct to your paper, both of the reviewers and I were very troubled by this."

Editors and reviewers mentioned issues with operationalization of measures for 46 manuscripts (66.67%). Issues with operationalizations included ambiguity in construct measures ($n = 5$ manuscripts), a misalignment between conceptualization and operationalization ($n = 21$), and issues with the choice of operationalization ($n = 29$). Sometimes editors and reviewers questioned the appropriateness of a construct's operationalization by noting a disconnect between the conceptualization and operationalization. A comment reflecting these concerns referred to a paper on teams: "I am not convinced that the two operationalizations of [the construct] ... used in the current manuscript are, in fact, the optimal operationalizations of [the construct].... Several plausible alternative operationalizations exist."

Convergent and discriminant validity. Another commonly critiqued aspect of measurement was a lack of convergent and discriminant evidence provided by the authors. This issue was raised in letters corresponding to a little over half of manuscripts ($n = 39$; 56.5%). Editors and reviewers questioned the convergent and discriminant validity of constructs when authors did not provide evidence of validity. For example, one reviewer's comment exemplifying this concern read, "Given the high correlations among the three forms of conflict, I would like to see evidence of discriminant validity (i.e., CFA)."

When measures were related, reviewers and editors often suggested that they might be indicators of the same construct. One reviewer reflected these thoughts for a paper on climate, "The two ... scales share a manifest level correlation of .66.... Maybe the two scales should be used as two indicators of a common construct." Some of reviewers' and editors' requests for factor analyses and structural equation modeling (as discussed further in the Data Analytic Errors section) were, in part, an attempt to make sure that the constructs were distinct, such as one reviewer's request that the authors "at least employ [an] exploratory factor analysis with all of the studies' items to provide additional evidence that the measures are distinct." Additionally, reasons provided by two reviewers to conduct a CFA were to "make sure the items load on the measures appropriately" and to "confirm that [constructs] are in fact different and appropriately measured."

Tests of measurement models. One strategy for establishing the appropriateness of measures is to test and compare measurement models. To urge authors to provide more information about the appropriateness of their measures, editors and/or reviewers for 13 manuscripts (18.94%) recommended factor analyses (e.g., CFA or EFA) when none had been conducted. In cases where analyses had been conducted, editors and reviewers raised issues regarding the authors' choice of factor analysis ($n = 10$ manuscripts). For instance, of these manuscripts, editors and reviewers specifically questioned why authors used EFAs instead of CFAs ($n = 5$) given that CFA procedures provide a stronger a priori test. As one editor noted, "EFA is used when the researcher has no prior theory regarding the proper number of factors (latent variables)." Thus, when the authors "seemed to have reasonable bases ... for a hypothesized factor structure that could be tested," an editor asked, "Why did you use an EFA as opposed to a CFA to understand the factor structure of your scale?" In particular, editors and reviewers preferred CFAs rather than EFAs for testing measurement model fit. We discuss the specific issues pertaining to the implementation of factor analyses in our review of data analytic errors.

Control Variables

In our analysis of the peer review process, the issue of control variables appeared in the editor's and/ or reviewer's letters for less than half of the manuscripts. While not as frequent as measurement topics, the issue of control variables reflected two contrary perspectives: (a) the failure to include control variables and (b) the inclusion of unnecessary controls. For 39.13% of the manuscripts ($n = 27$), reviewers and editors found it problematic that certain control variables were not included in the study, particularly ones that the editors and reviewers believed were theoretically relevant. As one reviewer wrote to the authors, "Perhaps the most concerning part of the current manuscript was the failure to control for endogenous factors that could account for the observed relationships."

Not all reviewers and editors thought it necessary to include more control variables. For 14.49% of the manuscripts ($n = 10$), editors and reviewers asked authors to justify their inclusion of control variables. Here, they were more concerned with the choice and inclusion of control variables rather than the omission of controls. For instance, one reviewer stated, "It wasn't clear to me why the control variables ... were added into the analyses. A more detailed explanation would be very helpful in this regard." Editors and reviewers appeared to be looking for theoretical justification for the inclusion of control variables that without justification may be deemed unnecessary.

Though one may think it easier to remove unnecessary controls or justify the inclusion of controls than to add control variables in a study, editors rejected relatively even percentages of manuscripts that included unnecessary controls (80%; $n = 8$ out of 10 manuscripts) compared to those manuscripts that failed to include controls (74.07%; $n = 20$ out of 27 manuscripts). Although the failure to include controls was more common in the manuscripts, these rejection rates suggest that authors should be wary of including more control variables simply to avoid omitting variables, as editors and reviewers also viewed too many controls as problematic. One reason that unnecessary control variables may be an issue is that their inclusion may prompt perceptions of weak theoretical justification for the entire study.

Common Method Variance

Common method variance can occur when similarities in measurement methods produce biased estimates of reliability and validity and result in inaccurate estimates of relationships among variables (Podsakoff, MacKenzie, & Podsakoff, 2012; Spector & Brannick, 2010). Editors and reviewers mentioned issues with CMV in their letters for 42.03% ($n = 29$) of the manuscripts. In terms of more specific CMV concerns, issues regarding self-report data were raised for 18 of these 29 manuscripts. Editors and reviewers were aware of situations in which CMV posed an issue, as exemplified by an editor's explanation to the authors:

I would have liked to see you address this problem in Study 2 by employing different procedural and/or statistical controls for CMV in order to rule out its potential biasing effects. However, the data in Study 2 were collected from a single source at one point in time using survey items with identical response scales. Such a design creates conditions that are ripe for CMV.

Some editors and reviewers deemed that the lack of validity in common method data was more of a fatal flaw than others. Two reviewers explicitly stated that research based entirely on self-report should not pass the peer review process (e.g., "I cannot envision a situation in which [self-report data] would pass muster in a peer-review process"). In contrast, others were not completely opposed to CMV. One editor wrote to the authors, "Common method bias is almost always an issue that emerges when we are examining perceptual variables simultaneously. Given the nature of what you are studying, I can live with this for sure."

There was not a significant difference between the proportion of reject and revise and resubmit decisions for manuscripts in which CMV was present (reject = 82.76%; revise and resubmit = 17.24%) compared to those manuscripts in which CMV was not present (reject = 70%; revise and resubmit = 30%). Conversely, 46.15% of all rejected manuscripts in our sample had issues with CMV, while only 29.41% of the manuscripts that received more favorable recommendations had CMV present. These results provide partial support for prior findings that editors and reviewers find issues with common method and self-report data (Chan, 2009) and may be quick to conclude that common methods inherently mean inflation of relationships between the variables.

A couple of manuscripts ($n = 2$) attempted to address CMV in their study using the Harman single factor test. Recent research has shown this technique to be largely ineffective at identifying CMV (Richardson, Simmering, & Sturman, 2009). Therefore, editors and reviewers appropriately appeared to be discouraging this practice. However, results suggest that more sophisticated statistical analyses to account for CMV may resolve the issues. Editors and reviewers recommended statistical remedies for CMV, citing Podsakoff, MacKenzie, Lee, and Podsakoff (2003; $n = 2$ manuscripts); Conway and Lance (2010; $n = 6$); and Johnson, Rosen, and Djurdjevic (2011; $n = 2$). Still, although editors and reviewers are proposing some post hoc analyses to address issues associated with common methods, results suggest that if they find issues with CMV in the study design (e.g., same source design), they will likely reject the manuscript.

Data Analytic Errors

Analysis of editors' and reviewers' comments revealed common errors in data analysis made by authors. Three major themes arose in the analyses relating to (a) the comprehensiveness of the analysis, (b) the choice of analysis, and (c) errors relating to specific procedures.

Comprehensiveness of analyses. One of the most common errors identified in the peer review process concerned analyzing one's data in a piecemeal fashion rather than using a more comprehensive approach. These comments fell into two categories: correlations between variables and testing for interactions.

Correlations between variables. For almost one quarter of the manuscripts (21.74%; $n = 15$), editors and reviewers commonly requested that authors account for the correlations between endogenous variables by conducting more comprehensive analyses. One reviewer's comment exemplified this request by saying that "rather than addressing any systematic issues ... [the author] essentially breaks what should be a complex multivariate analysis into a large set of narrow comparisons." In comparison to the overall 75.36% rejection rate for all manuscripts, editors rejected only 53.33% of the manuscripts in which this issue was raised in their and/or the reviewers' letters. Of all 69

manuscripts, 41.18% of the manuscripts that received a revise and resubmit contained this issue, while only 15.38% of the rejected manuscripts contained this issue. Because the failure to account for correlations between endogenous variables can be addressed post hoc with an alternate analysis, editors and reviewers may have been more forgiving of this error and willing to give authors a chance to confirm their original results.

Testing for interactions. Moderation was raised in editors' and reviewers' letters corresponding to 21 of the manuscripts (30.43%). Some of their comments suggested that authors test for moderators ($n = 7$ manuscripts) and look for interactions by including product terms rather than running separate additive analyses for different groups or levels of a potential moderator. Editors and reviewers also critiqued authors for taking a piecemeal approach to the testing of a model when authors considered moderation and mediation in isolation rather than testing them in a single moderated mediation ($n = 4$ manuscripts). As one editor explained to the authors, "Both reviewers point out that the analyses also lack integration that is suggested by the hypotheses, namely an analysis of moderated mediation." Reviewers and editors recognized in their comments that "testing the mediation separately from the moderation" means "taking a piecemeal approach to the testing of a model." When editors and reviewers mentioned the need to test for moderators or test for moderated mediation, these manuscripts had high rejection rates of 85.71% ($n = 6$) and 100% ($n = 4$), respectively. We propose reasons for these high rejection rates in the Discussion section.

Choice of analyses. A related yet different concern expressed frequently in the peer review process concerned choice of analyses. While these comments may be viewed by some as pointing to more comprehensive analytical approaches, we categorized these comments as reflecting a desire for more appropriate procedures.

Mediation. When editors and reviewers raised issues with manuscript authors' tests for mediation ($n = 13$ manuscripts), they often suggested more appropriate analytical procedures (69.23%; $n = 9$). Reviewers and editors criticized authors for using outdated tests for mediation, in particular the piecemeal portion of Baron and Kenny's (1986) approach ($n = 3$ manuscripts). Instead, reviewers requested advanced statistical procedures for testing for mediation including the Sobel test ($n = 5$ manuscripts), bootstrapping standard errors for indirect effects ($n = 5$), and structural equation modeling (SEM; $n = 5$). For these manuscripts where editors and/or reviewers raised issues with mediation, the rejection rate was 69.23% ($n = 9$), suggesting that editors and reviewers recognized this issue can be addressed by the authors through more sophisticated analyses.

HLM. More appropriate analyses were also suggested by editors and reviewers to account for issues with nested data and non-independence. For 15 (21.74%) manuscripts, editors and reviewers suggested HLM. For example, one editor wrote, "At the least, you need to account for organizational-level variance. A more sophisticated analysis of this type could also potentially examine some variables as a shared experience at the organizational level." Similar to tests for mediation, rejection rates for manuscripts containing this issue were relatively low, 60% ($n = 9$). Therefore, the request for authors to conduct an HLM was not necessarily a fatal flaw but rather a recommendation by editors and reviewers to account for the nested nature of the data.

SEM. In cases when dependent variables were highly related or alternative models needed to be tested ($n = 5$ manuscripts), editors and reviewers recommended that authors use SEM to estimate relationships among latent variables and simultaneously model multiple relationships. One reviewer wrote, "Authors, do not kill me, but you really need to run a SEM on this study!! You've got 3 DVs that are highly related and latent variables that have measurement issues." Another reviewer stated, "One way you could make a greater contribution [is] to take the results from the meta-analysis and use structural equation modeling to provide a comprehensive test of the … model." Consistent with

the results for mediation and HLM issues, the rejection rate for manuscripts with SEM issues was 60% ($n = 3$), although the issue arose relatively infrequently.

Errors relating to specific procedures. In addition to the two aforementioned general categories, we identified a number of common errors authors tended to make with regard to certain analytical procedures.

Factor analysis/SEM. Some of the most common comments pertaining to data analysis concerned the use of EFA, CFA, or SEM. These issues were raised in the editor's and/or reviewer's letters for over half of the manuscripts (53.62%; $n = 37$). These comments included inappropriate use of modification indices ($n = 2$ manuscripts), correlated residuals ($n = 2$), failure to test alternative models ($n = 10$), and/or improper elimination of items ($n = 4$). Of these 14 manuscripts containing such issues, editors rejected 85.71% ($n = 12$) of them.

Our analyses also identified some common errors with EFA. When authors used EFA, reviewers and editors questioned the factor analytic method being used ($n = 3$ manuscripts) and the choice of rotation method ($n = 1$). Notably, both of these issues have been identified as common errors when conducting an EFA (Bandalos & Boehm-Kaufman, 2009). Another issue raised in letters corresponding to two manuscripts had to do with authors conducting an EFA and CFA on the same data set. When manuscript authors conducted an EFA and CFA on the same sample, a reviewer explained that, "These confirmatory results do not provide evidence of confirmation, but merely indicate that two modeling approaches on the same data converge." Of the 37 manuscripts that had issues with factor analyses and/or SEM, 83.78% ($n = 31$) were rejected, which was higher than the 65.63% rejection rate ($n = 21$) of the 32 manuscripts that did not have these issues. These findings suggest that editors and reviewers are taking issues with factor analyses and SEM seriously.

Moderated regression. Editors and reviewers raised issues pertaining to moderated regression in their letters for 14.49% ($n = 10$) of the manuscripts. In 30% of the manuscripts where issues with moderated regression were raised ($n = 3$), editors and reviewers requested that authors center variables prior to testing for moderation, as is commonly recommended (Aiken & West, 1991). Additionally, authors were asked to include follow-up tests on significant interactions, as emphasized by one reviewer who wrote, "without simple slope tests we do not know which levels are actually different from one another." Simple slopes identify changes in the predicted relationship between variables at different levels of the moderator (Aiken & West, 1991). It is important to note that simple slopes are not an appropriate method if the moderator is continuous with no meaningful cutoff values because the different levels of the moderator would have no theoretical meaning (Dawson, 2014).

Discussion

This study elaborates on previous empirical studies (Fiske & Fogg, 1990; Rogelberg et al., 2009) by providing a detailed examination of methodological and statistical comments in reviewers' and editors' letters, identifying specific areas of concern, and relating them to manuscript decisions. We highlighted common errors made by authors both generally (e.g., lack of comprehensiveness of analyses) and for specific analyses (e.g., poor methods to improve model fit).

Overview of Findings

In the current study, editors and reviewers frequently commented on measurement topics (e.g., theoretical justification of constructs, conceptualization, and operationalization of measures) for manuscripts submitted for publication in *JBP* between July 2011 and November 2011. Issues with theoretical justification of constructs resulted in high manuscript rejection rates. Also common in

Table 4. Common Issues in Reviews, Corresponding Recommendations, and Supporting References.

Issue	Recommendation	References
Introduction		
Authors did not provide a detailed or clear definition of their constructs.	Constructs must be clearly defined. Specify constructs' depth (i.e., specific or general constructs for narrow or global phenomena, respectively) and dimensionality (i.e., unidimensional, multidimensional).	Bagozzi and Edwards (1998); Johnson, Rosen, Chang, Djurdjevic, and Taing (2012); Podsakoff, MacKenzie, and Podsakoff (2016)
It is unclear why these constructs were chosen over others.	Must articulate the importance of included constructs relative to viable alternatives.	Clark and Watson (1995)
It is unclear why the authors only used some components of the construct (e.g., only some facets of conscientiousness were used).	Must articulate the importance of included components of the construct relative to excluded components.	Clark and Watson (1995)
Theory does not match hypotheses.	Connections must be clear between hypotheses and the theory from which they are intended to flow.	Edwards and Berry (2010)
Design		
Study should have used a longitudinal rather than cross-sectional design.	To demonstrate change, one should collect data over multiple time points (with a minimum of three time points). If a cross-sectional design is used, language in the manuscript should reflect differences between groups/people rather than change over time.	McCleary and McDowall (2012); Ployhart and Ward (2011)
Sample is too small	It is recommended to do a statistical power analysis prior to conducting a study. If N is too small to detect the likely effect, then more data must be collected.	Cohen (1992); Kelley and Maxwell (2012); Faul, Erdfelder, Lang, and Buchner (2007); Mathieu, Aguinis, Culpepper, and Chen (2012)
Threats to validity (e.g., external validity)	Address limitations in the manuscript; but also, prior to data collection, be aware of the threats to validity posed by the study's design and try to address as many as possible.	Brutus, Aguinis, and Wassmer (2013); Highhouse (2009); Shadish, Cook, and Campbell (2002)
Operationalization		
Conceptualization and operationalization were misaligned.	Operationalization must capture as much of the conceptualization as possible (i.e., sufficiency) without also capturing extraneous factors (i.e., contamination).	Hinkin (1998); Ketchen, Ireland, and Baker (2013)
A proxy variable of dubious validity was used to measure a central construct.	For measures with questionable theoretical connection to constructs, empirical support must be especially compelling.	Hinkin (1998); Ketchen et al. (2013)
One or more items in a measure do not seem relevant to the intended construct.	Support the relevance of the item to the construct both rationally and empirically. If empirical support includes factor analysis, proper factor analysis techniques should be used (see factor analysis references in the following). Also, support retention of the item using item analysis and reliability analyses.	Bollen and Lennox (1991); Cortina (1993); Crocker and Algina (1986)

(continued)

Table 4. (continued)

Issue	Recommendation	References
Convergent and discriminant validity		
There is a lack of convergent validity evidence presented for measures.	Demonstrate convergent and discriminant validity via multitrait multimethod approaches and/or factor analyses (e.g., testing nested CFA models).	Bagozzi, Yi, and Phillips (1991); Campbell and Fiske (1959); Edwards (2003)
There is a lack of discriminant validity evidence presented for measures.		
Control variables[a]		
Authors omitted important control variables.	Include as controls any nuisance variables whose omission might adversely affect results.	Antonakis, Bendahan, Jacquart, and Lalive (2010); Antonakis and Dietz (2011)
Authors did not theoretically support the inclusion of control variables.	Explain inclusion of all control variables so that the reader can understand the negative consequences of omitting the variable.	Becker (2005); Bernerth and Aguinis (2016)
Common method variance[a]		
All of the data came from the same source. For instance, all of the data were self-report.	Potential problems with CMV should be considered and avoided in research design phase (e.g., spreading data collection over time).	Brannick, Chan, Conway, Lance, and Spector (2010); Conway and Lance (2010); Spector and Brannick (2010); Ostroff, Kinicki, and Clark (2002)
	If CMV is unavoidable after the design phase, one should choose the most appropriate statistical remedy.	Podsakoff, MacKenzie, Lee, and Podsakoff (2003), Podsakoff, MacKenzie, and Podsakoff (2012); Richardson, Simmering, and Sturman (2009)
Comprehensiveness of analyses		
Authors did not account for the correlations between endogenous variables (e.g., separate analyses for highly correlated DVs).	Correlations between endogenous variables should be accounted for with multivariate analyses such as structural equation modeling.	Huberty and Morris (1989)
Authors did not account for the correlations between exogenous variables.	Correlations between exogenous variables should be accounted for with multivariate analyses such as dominance analysis or relative weight analysis.	Azen and Budescu (2003); Johnson (2000); Tonidandel and LeBreton (2011)
Authors described moderation without seeming to realize it and thus did not test for moderators.	Recognize contingent relationships and situations where tests of moderation are appropriate and conduct those analyses using best practices for testing moderation.	Cohen (1978); Dawson (2014); Murphy and Russell (2016)
Authors proposed both mediation and moderation in the same study.	Tests for moderated mediation or mediated moderation might be necessary.	Edwards and Lambert (2007); Hayes (2013); Sardeshmukh and Vandenberg (2016)

(continued)

Table 4. (continued)

Issue	Recommendation	References
Choice of analyses		
Authors used outdated tests for mediation (i.e., Baron and Kenny).	Mediation hypotheses should be tested by focusing on the indirect effect and best practices regarding its standard error.	Cheung and Lau (2008); Hayes (2009); MacKinnon, Lockwood, Hoffmann, West, and Sheets (2002); MacKinnon, Coxe, and Baraldi (2012); James and Brett (1984); Preacher and Hayes (2004, 2008); Wood, Goodman, Beckmann, and Cook (2008)
Authors hypothesized a causal sequence, but they tested the relationships separately (i.e., piece by piece).	A causal sequence implies mediation, and it should be tested as such.	
Authors did not account for nested data and non-independence.	RCM (e.g., HLM) must be used with nested data. The traditional OLS approach should not be used with non-independent data as it underestimates standard errors.	Bliese and Ployhart (2002); Hofmann, Griffin, and Gavin (2000); Klein et al. (2000); Nezlek (2001)
Errors relating to specific procedures		
Authors used modification indices in CFA/SEM inappropriately.	Specification searches should be described clearly. Ideally, modified models should be cross validated. At the least, paths should only be added if they are theoretically defensible.	MacCallum, Roznowski, and Necowitz (1992); Kaplan (1990)
Authors correlated measurement errors in CFA/SEM.	Correlating measurement errors in CFA/SEM is almost always bad practice. The only common exceptions are when the errors are attached to indicators that represent the same variable at different time points or when some indicators are components of other indicators (e.g., X and X*Z).	Cole, Ciesla, and Steiger (2007); Cortina (2002); Landis, Edwards, and Cortina (2009); Gerbing and Anderson (1984)
Authors correlated structural residuals in SEM.	Correlating structural residuals in SEM is usually bad practice. However, there are common circumstances in which structural errors can be allowed to correlate (e.g., when they represent the same variable over time).	Kline (2005)
Authors failed to test alternative models in CFA/SEM.	It is advisable to test *plausible* alternative models with CFA/SEM.	Anderson and Gerbing (1988); MacCallum, Wegener, Uchino, and Fabrigar (1993); Vandenberg and Grelle (2009)
Authors did not do a factor analysis for their measures.	Factor analyses are an essential part of scale development.	Comrey (1988); Hinkin (1998); Reise, Waller, and Comrey (2000)
There are issues with choice of EFA or CFA (e.g., why EFA instead of CFA, or vice-versa).	EFA is more appropriate when little is known about factor structure. If an a priori factor structure exists, then CFA is more appropriate.	Bandalos and Boehm-Kaufman (2009); Floyd and Widaman (1995); Henson and Roberts (2006)
Authors conducted EFA and CFA on same data set.	Factor structure from an EFA should be confirmed with CFA on a different data set.	Henson and Roberts (2006)
Authors used questionable factor analytic methods (e.g., improperly eliminated items in CFA/ SEM).	Be aware of best practices for factor analyses, such as EFA methods to determine the number of factors to retain (e.g., parallel analysis).	O'Connor (2000); Zwick and Velicer (1986)

Table 4. (continued)

Issue	Recommendation	References
Authors used questionable rotation methods in EFA.	In EFA, oblique rotations are recommended because they permit correlations among factors. If the factors are not correlated, the oblique rotation will produce a solution close to that produced by an orthogonal rotation.	Bandalos and Boehm-Kaufman (2009); Fabrigar, Wegener, MacCallum, and Strahan (1999); Preacher and MacCallum (2003)
Authors did not center variables prior to testing for moderation.	Prior to testing for moderation, variables should be centered as this renders weights for main effect terms more interpretable. This technique will not improve fit, impact power, or change reliability of the interaction term.	Cohen, Cohen, West, and Aiken (2003); Dalal and Zickar (2012)
Authors did not include follow-up tests on simple slopes for moderated regression.	Follow-up tests on simple slopes for moderated regression should be included in order to determine the precise nature of the interaction. Simple slopes are useful to provide meaningful values for continuous variables and categorical variables but not meaningful for arbitrary cutoffs. Use the Johnson-Neyman technique or *regions of significance* as alternatives for probing significant interactions.	Aiken and West (1991); Dawson (2014); Preacher, Rucker, and Hayes (2007)

Note: CFA = confirmatory factor analysis; CMV = common method variance; DV = dependent variable; RCM = random coefficient modeling; HLM = hierarchical linear modeling; OLS = ordinary least squares; EFA = exploratory factor analysis.
[a]There tends to be widespread disagreement for issues and recommendations pertaining to common method variance and control variables.

editors' and reviewers' letters were comments on theoretical justification of study design and hypotheses. Additionally, editors and reviewers raised concern with the failure to include controls more often than the inclusion of unnecessary controls; however, the presence of either issue related to the theoretical justification of control variables and resulted in similar rejection rates.

Issues with common methods, particularly self-report data, were found to often result in manuscript rejection. Other common data analytic errors related to higher rejection rates (e.g., failure to test for moderated mediation, inappropriate factor analysis test); although results also suggested that some data analytic errors might not necessarily constitute as fatal flaws (e.g., failure to account for the correlations between endogenous variables, failure to conduct a SEM or HLM, failure to test for mediation). We discuss the apparent inconsistencies in editors' manuscript decisions for these issues that could be resolved with revisions in the following section on data analytic errors.

In order to provide recommendations to editors, reviewers, and authors concerning the methodological and statistical topics covered in the results, we relate our findings to previous research as well as current debates in the literature. Table 4 provides an overview of the common issues raised by editors and reviewers as well as recommendations and sources to remedy these issues. Authors who want to improve their current methodological practices, or who are concerned with an editor's or reviewer's comment, can refer to this table for appropriate solutions and references. Editors and reviewers can also use this table to provide comments that are consistent with recommended methodological and statistical practices. Table 4 notes that there is disagreement in the literature on the topics of CMV and control variables. We have provided sources for both sides of the debate, but in our recommendations, we identified a middle ground that does justice to both. Editors and reviewers should be aware that there does not appear to be much agreement in the literature for these topics, and authors should note these widespread disagreements when addressing issues pertaining to these topics.

Relationship to Current Methodological Practices and Recommendations

Design. Design issues were commonly raised for manuscripts within this study, with issues corresponding to threats to validity, threats to causal inference, and sampling appearing most often. Our study's findings are in line with previous findings from reviews of limitation sections in empirical articles published in *Journal of Business Venturing* between 2005 and 2010 (Aguinis & Lawal, 2012) and in empirical articles published in various top journals from 1982 to 2007 (Brutus et al., 2013). These studies found that common design limitations mentioned by empirical articles included small sample size and lack of confidence regarding causality (Aguinis & Lawal, 2012), as well as external validity issues, particularly concerning lack of generalizability (Brutus et al., 2013). This suggests that the design issues commonly raised by editors and reviewers are also reflected in the limitation sections of published manuscripts. Given the high frequency of design issues raised in the limitation section of published manuscripts, it is likely that these manuscript authors could not have mitigated these issues even if editors and reviewers raised them during the peer review process. Although manuscripts can be published with design limitations, authors should consider sampling and validity threats in the planning stage of their study to avoid as many design issues as possible.

Previous research has found that adequacy of design had the biggest influence on editorial recommendations for a manuscript (Gilliland & Cortina, 1997). Similarly, Fogg and Fiske (1993) found a significant correlation between severe planning and execution criticisms and negative reviewers' final recommendations. Although design issues were common in the current study's editors' and reviewers' letters, we did not find a higher rejection rate for manuscripts with design issues compared to the other statistical issues. Advanced planning in research design is important, but it is likely that editors and reviewers recognize that every study has some limitations. We recommend that authors follow Brutus et al.'s (2013) guidelines on how to report limitations and address the threats to validity they face with their studies.

Measurement. This study identified measurement concerns as a critical issue that authors must anticipate if they want to improve their chances of success in the publication process. Our field has a long history of emphasizing high quality measurement (SIOP, 2003). Conversely, methodologists have lamented the seeming lack of emphasis being placed on high quality measurement. This concern is exemplified in a recent exchange on the electronic mailing list of the Research Methods division of the Academy of Management (RMNET) regarding construct validity:

> This seems like it would be a great subject for a new Urban Legends piece. I can see the title now: "My measure was used in a top tier publication, so of course it's reliable and valid!" . . . The other thought this string triggered is a potentially interesting disconnect between elements of professional practice and scientific inquiry in our field. In some cases, particularly in the case of high-stakes testing or personnel selection, standards for psychometric quality of a measure seem to be different [APA's Standards and SIOP's Principles] . . . from the standards that measures are held to in journals (almost regardless of tier), For example, with few exceptions, a practicing I-O psychologist in the area of personnel selection would be hard pressed to simply cite an article in a top tier journal as an adequate defense for the psychometric properties of a selection measure s/he plans to use in a hiring process. (Putka, 2011)

Our findings that the majority of editors' and reviewers' comments raised issues pertaining to measurement suggest that editors and reviewers are taking measurement concerns seriously. The notion that measurement is of great importance coincides with the statement that "the point is not that adequate measurement is 'nice.' It is necessary, crucial, etc. Without it we have nothing" (Korman, 1974, p. 194). As a result, authors would be advised to pay particularly close attention

to the development of their measures and provide comprehensive empirical evidence in support of construct validity. It is important that authors first start with a clear concept definition (for examples and recommendations, see Podsakoff, MacKenzie, & Podsakoff, 2016). Then, authors need to conduct convergent and discriminant validation to demonstrate appropriate linkages with other measures and constructs (Rothstein & Goffin, 2000). Whereas data analytic problems may be reparable after the fact, measurement problems often are not.

While some construct validity issues could be more fatal flaws than others, we cannot say definitively that some hold more weight in the final manuscript decision. However, results from previous research have suggested that issues with operationalization of constructs are frequently mentioned in the limitations sections of empirical articles (Aguinis & Lawal, 2012) and are strongly negatively related to reviewers' manuscript evaluations and the manuscript decision (Gilliland & Cortina, 1997), suggesting that authors should pay close attention to their operationalization of constructs.

Control variables. Our analysis of comments pertaining to control variables reflected two conflicting points of view: more controls are needed and fewer controls are needed. These findings elaborate on previous research that only covered reviewers' comments when manuscript authors had no or poor controls (Fiske & Fogg, 1990). The two perspectives that emerged from our analyses mirror the debate in the statistical literature concerning the appropriate use of control variables. One side of the debate generally argues for the inclusion of more control variables (Antonakis, Bendahan, Jacquart, & Lalive, 2010), while the other typically recommends including a more limited set of control variables that can be theoretically justified (Spector & Brannick, 2011).

Those methodologists who argue for more controls warn against omitted variable bias, such that omitting variables leads to biased estimates in regression coefficients and may lead to estimations of the wrong model (Antonakis et al., 2010). In regards to the inclusion of irrelevant regressors, these researchers suggest, "It is always safer to err on the side of caution by including more than fewer control variables" (Antonakis et al., 2010, p. 1092). Carlson and Wu (2012) argue that the choice to include or not to include controls can heavily influence one's findings and that it is a common misconception that the inclusion of control variables always allows for the isolation of the effect of the independent variable on the dependent variable, holding other influences constant. Additionally, there is the concern that authors can conduct analyses with all possible subsets of a set of potential controls until they find the subset that produces the most flattering results for the constructs of interest. In order to reduce the belief that statistical controls always lead to more accurate estimates of variable relationships, Spector and Brannick (2011) suggest that "Editors and reviewers should be the first line of defense against the purification principle" by recommending that authors refrain from including too many control variables (p. 302).

Previous studies have examined journal articles for authors' use of control variables (Becker, 2005; Carlson & Wu, 2012) in order to provide authors with suggestions and guide editors' and reviewers' commentary, but to our knowledge, our study is the first to examine the peer review process itself to see if reviewers' and editors' comments reflect leading methodologists' suggestions. Becker (2005) and Carlson and Wu (2012) stressed the need for better justification for the inclusion of controls. Our findings show that reviewers and editors similarly wanted explanations as to why the authors included certain control variables. One reviewer's suggestion to "include as control variables those that make most sense on theoretical grounds," such that "a more compelling rationale should be given for the variables you have statistically controlled for in each study" exemplifies the idea that "authors should be asked to thoroughly explain and justify what they have done. They should not get away with merely saying they included a control variable just because it might affect the variables of interest" (Spector & Brannick, 2011, p. 302). Our findings fit Spector and Brannick's (2011) recommendation that when editors and reviewers request that authors include control variables, they should provide the same amount of theoretical justification that they expect

from the authors. The present study's results suggest that although editors and reviewers recognize the issues of unmodeled endogeneity (Antonakis et al., 2010), they are demonstrating vigilance by asking authors for theoretical justification of control variables and by providing justification for their suggestions to include control variables.

Both sides of the debate suggest that control variables should be included on the basis of theoretical importance; however, they differ as to whether one should be wary of including too few or too many control variables. The finding that editors and reviewers more frequently raised issues with manuscripts having too few compared to too many control variables suggests that authors need to think carefully about which variables to include and provide justification for their omission of relevant control variables. Similarly, editors and reviewers should theoretically support any control variables they recommend for manuscripts.

Common method variance. Another current methodological debate reflected in our results concerns CMV and the bias that can be associated with it. In their content analysis of reviewers' letters, Rogelberg and colleagues (2009) made mention of CMV but provided brief recommendations (e.g., CMV should be considered in the design of the study) that do not disentangle or delve into the extent of the debate on issues concerning CMV. Many researchers believe that CMV is an important issue and must be controlled, while others suggest the issues are not inherent in CMV and that these concerns are misplaced. Those who express concern regarding CMV argue that CMV can threaten construct validity, influence the structure of constructs, obscure relationships between constructs by inflating reliability estimates and convergent validity, and deflate the relationship between different constructs (e.g., Bagozzi & Yi, 1990; Baumgartner & Steenkamp, 2001; Doty & Glick, 1998; Podsakoff et al., 2003). Self-report data, the most common CMV culprit, has been characterized as creating low construct validity due in part to social desirability (Spector & Brannick, 2010) and is posited as a red flag for issues with CMV. In an editorial for the *Journal of Applied Psychology*, Kozlowski (2009) writes,

> Most desk-rejected manuscripts (aside from posing a trivial research question) are single-shot, cross-sectional, self-report survey designs. There are very rare occasions where such a design may be warranted and defensible. In such situations, authors need to make every effort to address the concerns of common source variance (see Podsakoff, MacKenzie, Podsakoff, & Lee, 2003). (p. 3)

Podsakoff et al. (2012) argue that there is an overwhelming consensus among researchers that CMV is important and should be controlled whenever possible. Our findings would suggest that on many occasions, editors and reviewers find issues with CMV. Nevertheless, the extent to which these issues manifest themselves in practice are under dispute (Chan, 2009).

The finding that editors and reviewers raised issues with CMV suggests that editors and reviewers may not agree with Spector's (2006) argument that "the assumption that method alone is sufficient to produce biases, so that everything measured with the same method shares some of the same biases" is an urban legend (p. 223). Spector's side of the CMV debate suggests that the construct and its measurement, not just the source of data, determine the amount of shared bias a measure will have with others. Conway and Lance (2010) caution against reviewers assuming "a) that relationships between self-reported variables are necessarily and routinely upwardly biased, b) other-reports (or other methods) are superior to self-reports, and c) rating sources (e.g., self, other) constitute measurement methods" (p. 235). Although they acknowledge that there is potential for CMV to bias findings, they suggest that gatekeepers should not base their criticisms purely on these assumptions. Rather, editors, reviewers, and authors should evaluate measurement within the context of the research situation (Conway & Lance, 2010).

In line with the current CMV debate, authors are encouraged not to ignore CMV but to address the threats in their justifications, procedures, and if necessary, post hoc techniques. Though having data with CMV is not necessarily a fatal flaw, authors should carefully consider in the design process if they are able to defend their choice of measures. If authors can avoid sole reliance on a single method, then they should. If authors believe that measures with common method are necessary, then they should be aware of the situations in which CMV may or may not be an issue. For instance, research suggests that CMV inflates correlations in multilevel and cross-level studies (Ostroff, Kinicki, & Clark, 2002) but that CMV does not generate artificial interaction effects (Evans, 1985).

Based on our findings, editors and reviewers seem prone to find issues with CMV, and thus, the potential for manuscript rejection is high. Authors who determine that common methods are necessary for their study should consider strategies to attenuate CMV's negative effects and be prepared to address the potential threats that CMV poses. Some techniques to address the potential threats that CMV poses include spreading data collection over time to reduce temporary affective states and memory effects (Ostroff et al., 2002), splitting the sample in half (i.e., using half the sample to measure one construct, half the other construct; Ostroff et al., 2002), and providing post hoc analyses that control method biases (e.g., Podsakoff et al., 2003). Authors should avoid the Harman single factor technique to remedy potential problems with CMV and aim to use techniques by Podsakoff et al. (2003), Conway and Lance (2010), or Johnson et al. (2011).

Data analytic errors. Data analytic errors related to three general themes in reviewers' and editors' comments: (a) the comprehensiveness of the analysis, (b) the choice of analysis, and (c) errors relating to specific procedures.

Choosing the most appropriate statistical technique mitigates the risk of errors in the comprehensiveness of data analyses. When testing for mediation, authors should use more appropriate procedures such as differences in coefficients and products of coefficients tests (e.g., Sobel's test), bootstrapping, and SEM, instead of Baron and Kenny's (1986) causal steps that do not focus on the indirect effect. Authors must also be aware of the distributional assumptions on which tests are based (e.g., the Sobel test assumes multinormality). Although additive analyses are more parsimonious, authors should conduct multiplicative analyses when appropriate. For instance, if authors propose both mediation and moderation in the same study, they should be aware that tests for moderated mediation might be necessary and should use appropriate statistical techniques to test these models (Edwards & Lambert, 2007; Hayes, 2013; Sardeshmukh & Vandenberg, 2016). However, authors should be wary of instances when multiplicative analyses are not meaningful (see e.g., Murphy & Russell, 2016).

Our findings suggest that authors should account for correlations between endogenous variables with comprehensive analyses (e.g., MANOVAs or step-down technique). Additionally, to analyze highly related dependent variables or latent variables with measurement issues, authors should consider conducting SEM analyses, although there are considerations to take into account (see recommendations and references in Table 4). In order to account for issues with nested data, such as the non-independence of data, authors are encouraged to use random coefficient modeling (RCM), as the traditional ordinary least squares (OLS) approach should not be used with non-independent data. As these analyses can be addressed in revisions, editors and reviewers did not seem to reject as many of these manuscripts compared to the overall rejection rate. However, this was not the case for all issues that could be addressed through revision.

Though other analyses could also be conducted post hoc, authors' failure to conduct these analyses may have raised questions about their study's conclusions or the manuscript itself, as demonstrated by high rejection rates for the failure to include certain analyses (e.g., moderation) and for improper implementation of factor analyses and/or SEM. For instance, when authors failed to test for interactions that would have made their hypotheses or analyses more comprehensive, the majority of these manuscripts were rejected, perhaps because authors did not include the necessary

moderators in their studies or because the interpretation of their data would change if they tested for a moderator. Similarly, high rejection rates resulted when authors improperly implemented a SEM or factor analysis, possibly because the fit or interpretation of the model would change when properly tested.

Reviewers and editors also provided particular suggestions for errors relating to factor analyses and SEM. While previous research found that reviewers and editors often requested CFAs or EFAs for measures (Rogelberg et al., 2009), the current study expands on these findings by including more specific issues and recommendations regarding the proper implementation of factor analyses and structural models. For instance, authors should be sure to appropriately use modification indices, test alternative models, explain and support the elimination of items, and avoid correlating residuals unless they can justify the choice (e.g., an endogenous variable measured at different times). When conducting an EFA, authors should avoid common errors by justifying the use of an EFA instead of a CFA, not conducting an EFA and CFA on the same data set, and supporting the choice of rotation method (see Bandalos & Boehm-Kaufman, 2009, for more on these common errors).

Regarding moderated regression, editors and reviewers suggested that authors include tests for simple slopes and center their variables. The request for simple slopes was made in several cases where issues with moderated regression were raised. This phenomenon is interesting considering the viewpoint of some researchers that tests of simple slopes may not be appropriate in most situations. As Dawson (2014) explains, tests of simple slopes often tell us little about the effect of interest and should only be used in certain circumstances (e.g., where the conditional effect at a certain value of the moderator would be particularly meaningful). Moreover, the use of arbitrary values such as one standard deviation above or below the mean to plot simple effects offers little utility. Authors should use and editors and reviewers should suggest the Johnson-Neyman technique or regions of significance as alternatives to simple slopes when moderators are continuous with no meaningful cutoffs (Aiken & West, 1991; Preacher, Rucker, & Hayes, 2007). Authors, editors, and reviewers should recognize the utility of simple slopes but also be aware of the circumstances in which this test should be used.

Editors and reviewers also occasionally requested that authors center variables prior to tests of moderation. Conversely, a recent article in the literature suggests that centering variables may be misguided (Dalal & Zickar, 2012). Authors should realize that centering variables in moderation only reduces non-essential collinearity, making the main effects easier to interpret. Authors should not misinterpret reviewers' suggestions to center variables as ones that will improve fit, impact power, or change reliability of the interaction term (Dalal & Zickar, 2012). Authors should center their variables prior to testing for moderation, but they, as well as reviewers and editors, should realize that this technique does not resolve problems such as poor model fit.

When considering the totality of issues raised, there was considerable overlap between methodology and theory, as demonstrated by the presence of theoretical concerns raised in the majority of manuscript reviews, suggesting that authors' choices in methodology should be theory driven. The emphasis on theoretical justification in editors' and reviewers' comments pertaining to control variables, measures, and analyses confirm that authors should start with theory and then select the method and analyses that best allow tests of the theory (Wilkinson, 1999). As previously mentioned, authors are encouraged to reference Table 4 for an overview of issues commonly raised by reviewers and editors, recommendations, and additional sources.

Limitations and Future Research

There are several limitations within this study that should be noted. Although we generated 267 codes from 1,751 statements in 304 letters, the letters came from only 69 manuscripts. The sample of 69 manuscripts was a relatively good size given that we incorporated both reviewers' and editors'

letters for each manuscript, allowing us to look at comments and concerns among reviewers and editors for the same manuscript. In previous studies, Fiske and Fogg (1990) coded 402 reviewers' letters, and Rogelberg et al. (2009) analyzed 131 reviewers' letters, but they did not examine all editors' and reviewers' letters for every manuscript included in their studies.

It should also be noted that the sample of editors and reviewers were a select group of people, and thus, the opinions of these editors and reviewers may not be representative of the field as a whole. Moreover, though the perspectives of multiple editors ($N = 13$) and many reviewers ($N = 107$) are represented, the sample consists of multiple letters from the same editors and from 29 of the 107 distinct reviewers. Additionally, while *JBP* is a highly respected journal, the goals of this journal may differ from those of other journals. Still, as previously mentioned, it is typical for associate editors and board members to serve other journals, and therefore, the thoughts reflected in this journal should be consistent with those reflected in other journals. Future research should expand the qualitative analysis of reviews to other journals.

An additional limitation arises from the lack of multiple independent coders. Though the initial codes were reviewed, organized, and categorized by two coders, consensus coding was used, meaning that statistics about interrater agreement and reliability cannot be computed. Because the present study is cross-sectional, the study cannot support or attest to trends in methods comments over time. Future studies should look at the peer review process over time to examine if methodological trends in reviews match those of the literature.

Conclusion

In conclusion, the present study provides prospective authors with detailed information regarding what the gatekeepers say about research methods and analysis in the peer review process. Our hope is that reviewers and editors can use these findings to improve on their own reviews and decisions. Similarly, researchers can use this information and our subsequent recommendations to enhance the quality of their current methodological practices and, most importantly, conduct effective and appropriate research. Doing so should ultimately enhance their chances of publication success.

Notes

1. There is no overlap in the current study's sample and Rogelberg, Adelman, and Askay's (2009) sample that included 100 manuscripts submitted to *Journal of Business and Psychology* between November 2008 and April 2009.
2. Table 2 considers an issue present in a manuscript if it was raised in the editor's and/or at least one of the reviewer's letters. To further link the presence of issues with manuscript outcomes, we reran analyses linking final manuscript decisions to issues raised in the editor's letter only. Editors' letters may directly relate to manuscript decisions more than reviewers' letters because editors are ultimately making the manuscript decision. The proportions of manuscripts rejected were not largely different from the proportions presented in Table 2. Thus, for the remainder of the manuscript, we refer to the results in Table 2, where we relate both editors' and reviewers' comments to manuscript outcomes. Results linking only editors' comments to manuscript outcomes are available from the first author upon request.

References

Aguinis, H., & Lawal, S. O. (2012). Conducting field experiments using eLancing's natural environment. *Journal of Business Venturing*, *27*, 493-505.

Aguinis, H., Pierce, C. A., Bosco, F. A., & Muslin, I. S. (2009). First decade of organizational research methods: Trends in design, measurement, and data-analysis topics. *Organizational Research Methods*, *12*, 69-112. doi:10.1177/1094428108322641

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: SAGE Publications, Inc.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*, 411-423.

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

Antonakis, J., & Dietz, J. (2011). More on testing for validity instead of looking for it. *Personality and Individual Differences*, *50*, 418-421.

ATLAS.ti. (n.d.). ATLAS.ti. (Version 6.2) [Computer software]. Berlin: ATLAS.ti Scientific Software Development GmbH.

Austin, J. T., Scherbaum, C. A., & Mahlman, R. A. (2002). History of research methods in industrial and organizational psychology: Measurement, design, analysis. In S. G. Rogelberg (Ed.), *Handbook of research methods in industrial and organizational psychology* (pp. 3-33). Malden, MA: Blackwell Publishing.

Azen, R., & Budescu, D.V. (2003). The dominance analysis approach for comparing predictors in multiple regression. *Psychological Methods*, *8*, 129-148.

Bagozzi, R. P., & Edwards, J. R. (1998). A general approach for representing constructs in organizational research. *Organizational Research Methods*, *1*, 45-87.

Bagozzi, R. P., & Yi, Y. (1990). Assessing method variance in multitrait-multimethod matrices: The case of self-reported affect and perceptions at work. *Journal of Applied Psychology*, *75*, 547-560.

Bagozzi, R. P., Yi, Y., & Phillips, L. W. (1991). Assessing construct validity in organizational research. *Administrative Science Quarterly*, *36*, 421-458.

Bandalos, D. L., & Boehm-Kaufman, M. R. (2009). Four common misconceptions in exploratory factor analysis. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences* (pp. 61-87). New York, NY: Routledge.

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*, 1173-1182. doi:10.1037/0022-3514.51.6.1173

Baumgartner, H., & Steenkamp, J. B. E. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, *38*, 143-156.

Becker, T. E. (2005). Potential problems in the statistical control of variables in organizational research: A qualitative analysis with recommendations. *Organizational Research Methods*, *8*, 274-289. doi:10.1177/1094428105278021

Bernerth, J., & Aguinis, H. (2016). A critical review and best-practice recommendations for control variable usage. *Personnel Psychology*, *69*, 229-283.

Bliese, P. D., & Ployhart, R. E. (2002). Growth modeling using random coefficient models: Model building, testing, and illustrations. *Organizational Research Methods*, *5*, 362-387.

Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, *110*, 305-314.

Bono, J. E., & McNamara, G. (2011). Publishing in *AMJ* part 2: Research design. *Academy of Management Journal*, *54*, 657-660. doi:10.5465/AMJ.2011.64869103.

Brannick, M. T., Chan, D., Conway, J. M., Lance, C. E., & Spector, P. E. (2010). What is method variance and how can we cope with it? A panel discussion. *Organizational Research Methods*, *13*, 407-420.

Brutus, S., Aguinis, H., & Wassmer, U. (2013). Self-reported limitations and future directions in scholarly reports analysis and recommendations. *Journal of Management*, *39*, 48-75.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81-105.

Carlson, K. D., & Wu, J. (2012). The illusion of statistical control: Control variable practice in management research. *Organizational Research Methods*, *15*, 413-435.

Chan, D. (2009). So why ask me? Are self-report data really that bad? In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends* (pp. 309-336). New York, NY: Routledge.

Cheung, G. W., & Lau, R. S. (2008). Testing mediation and suppression effects of latent variables: Bootstrapping with structural equation models. *Organizational Research Methods*, *11*, 296-325.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*, 309-319.

Clark, S. M., Gioia, D. A., Ketchen, D. J., & Thomas, J. B. (2010). Transitional identity as a facilitator of organizational identity change during a merger. *Administrative Science Quarterly*, *55*, 397-438.

Cohen, J. (1978). Partialed products *are* interactions; partialed powers *are* curve components. *Psychological Bulletin*, *85*, 858-866.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155-159.

Cohen, J., Cohen, P., West, S. G ., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.

Cole, D. A., Ciesla, J. A., & Steiger, J. H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods*, *12*, 381-398.

Comrey, A. L. (1988). Factor-analytic methods of scale development in personality and clinical psychology. *Journal of Consulting and Clinical Psychology*, *56*, 754-761.

Conway, J. M., & Lance, C. E. (2010). What reviewers should expect from authors regarding common method bias in organizational research. *Journal of Business and Psychology*, *25*, 325-334. doi:10.1007/s10869-010-9181-6

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*, 98-104.

Cortina, J. M. (2002). Big things have small beginnings: An assortment of "minor" methodological understandings. *Journal of Management*, *28*, 339-362.

Cortina, J. M. (2003). Apples and oranges (and pears, oh my!): The search for moderators in meta-analysis. *Organizational Research Methods*, *6*, 415-439.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York, NY: CBS College Publishing.

Dalal, D. K., & Zickar, M. J. (2012). Some common myths about centering predictor variables in moderated multiple regression and polynomial regression. *Organizational Research Methods*, *15*, 339-362.

Dawson, J. F. (2014). Moderation in management research: What, why, when, and how. *Journal of Business and Psychology*, *29*, 1-19. doi:10.1007/s10869-013-9308-7

Desrosiers, E. I., Sherony, K., Barros, E., Ballinger, G. A., Senol, S., & Campion, M. A. (2002). Writing research articles: Update on the article review checklist. In S. G. Rogelberg (Ed.), *Handbook of research methods in industrial and organizational psychology* (pp. 459-478). Malden, MA: Blackwell Publishing.

Doty, D. H., & Glick, W. H. (1998). Common methods bias: Does common methods variance really bias results? *Organizational Research Methods*, *1*, 374-406.

Edwards, J. R. (2003). Construct validation in organizational behavior research. In J. Greenberg (Ed.), *Organizational behavior: The state of the science* (2nd ed., pp. 327-371). Mahwah, NJ: Erlbaum.

Edwards, J. R. (2010). Reconsidering theoretical progress in organizational and management research. *Organizational Research Methods*, *13*, 615-619.

Edwards, J. R., & Berry, J. W. (2010). The presence of something or the absence of nothing: Increasing theoretical precision in management research. *Organizational Research Methods*, *13*, 668-689.

Edwards, J. R., & Lambert, L. S. (2007). Methods for integrating moderation and mediation: A general analytical framework using moderated path analysis. *Psychological Methods*, *12*, 1-22.

Evans, M. G. (1985). A Monte Carlo study of the effects of correlated method variance in moderated multiple regression analysis. *Organizational Behavior and Human Decision Processes*, *36*, 305-323.

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*, 272-299.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

Fiske, D. W., & Fogg, L. F. (1990). But the reviewers are making different criticisms of my paper! Diversity and uniqueness in reviewer comments. *American Psychologist*, *45*, 591-598. doi:10.1037/0003-066X.45.5.591

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, *7*, 286-299.

Fogg, L., & Fiske, D. W. (1993). Foretelling the judgments of reviewers and editors. *American Psychologist*, *48*, 293-294. doi:10.1037/0003-066X.48.3.293

Frost, N. (2011). *Qualitative research methods in psychology*. Berkshire, UK: Open University Press.

Gerbing, D. W., & Anderson, J. C. (1984). On the meaning of within-factor correlated measurement errors. *Journal of Consumer Research*, *11*, 572-580.

Gilliland, S. W., & Cortina, J. M. (1997). Reviewer and editor decision making in the journal review process. *Personnel Psychology*, *50*, 427-452.

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research notes on the Gioia methodology. *Organizational Research Methods*, *16*, 15-31. doi:10.1177/1094428112452151

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, *76*, 408-420.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis*. New York, NY: Guilford.

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement*, *66*, 393-416.

Highhouse, S. (2009). Designing experiments that generalize. *Organizational Research Methods*, *12*, 554-566.

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, *1*, 104-121.

Hofmann, D., Griffin, M., & Gavin, M. (2000). The application of hierarchical linear modeling to organizational research. In K. Klein & S. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations* (pp. 467-511). San Francisco, CA: Jossey-Bass.

Huberty, C. J., & Morris, J. D. (1989). Multivariate analysis versus multiple univariate analyses. *Psychological Bulletin*, *105*, 302-308.

James, L. R., & Brett, J. M. (1984). Mediators, moderators, and tests for mediation. *Journal of Applied Psychology*, *69*, 307-321.

Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research*, *35*, 1-19.

Johnson, R. E., Rosen, C. C., Chang, C. H., Djurdjevic, E., & Taing, M. U. (2012). Recommendations for improving the construct clarity of higher-order multidimensional constructs. *Human Resource Management Review*, *22*, 62-72.

Johnson, R. E., Rosen, C. C., & Djurdjevic, E. (2011). Assessing the impact of common method variance on higher order multidimensional constructs. *Journal of Applied Psychology*, *96*, 744-761. doi:10.1037/a0021504

Kaplan, D. (1990). Evaluating and modifying covariance structure models: A review and recommendation. *Multivariate Behavioral Research*, *25*, 137-155.

Kelley, K., & Maxwell, S. E. (2012). Sample size planning. In H. Cooper, M. Paul, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology: Foundations, planning, measures, and psychometrics* (Vol. *1*, pp. 181-202). Washington, DC: American Psychological Association.

Ketchen, D. J., Ireland, R. D., & Baker, L. T. (2013). The use of archival proxies in strategic management studies: Castles made of sand? *Organizational Research Methods*, *16*, 32-42.

Klein, K. J., Bliese, P. D., Kozolowski, S. W. J., Dansereau, F., Gavin, M. B., Griffin, M. A., . . . Bligh, M. C. (2000). Multilevel analytical techniques: Commonalities, differences, and continuing questions. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 512-553). San Francisco, CA: Jossey-Bass.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.) New York, NY: Guildford Press.

Koppman, S., & Gupta, A. (2014). Navigating the mutual knowledge problem: A comparative case study of distributed work. *Information Technology & People*, *27*, 83-105.

Korman, A. K. (1974). Contingency approaches to leadership: An overview. In J. G. Hunt & L. L. Larson (Eds.), *Contingency approaches to leadership*. Carbondale, IL: Southern Illinois University Press.

Kozlowski, S. W. J. (2009). Editorial. *Journal of Applied Psychology*, *94*, 1-4. doi:10.1037/a0014990

Landis, R. S., Edwards, B. D., & Cortina, J. M. (2009). On the practice of allowing correlated residuals among indicators in SEM. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends* (pp. 193-218). New York, NY: Taylor & Francis Group, LLC.

Lee, T. M., Mitchell, T. R., & Harman, W. S. (2011). Qualitative research strategies in industrial and organizational psychology. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology, Building and developing the organization* (Vol. *1*, pp. 73-83). Washington, DC: American Psychological Association.

Locke, K. (2002). The grounded theory approach to qualitative research. In F. Drasgow & N. Schmitt (Eds.), *Measuring and analyzing behavior in organizations: Advances in measurement and data analysis* (pp. 17-43). San Francisco, CA: Jossey-Bass.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modification in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, *111*, 490-504.

MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin*, *114*, 185-199.

MacKinnon, D. P., Coxe, S., & Baraldi, A. N. (2012). Guidelines for the investigation of mediating variables in business research. *Journal of Business and Psychology*, *27*, 1-14.

MacKinnon, D. P., Lockwood, C. M., Hoffmann, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test the significance of mediation and other intervening variable effects. *Psychological Methods*, *7*, 83-104.

Mathieu, J. E., Aguinis, H., Culpepper, S. A., & Chen, G. (2012). Understanding and estimating the power to detect cross-level interaction effects in multilevel modeling. *Journal of Applied Psychology*, *97*, 951-966.

McCleary, R., & McDowall, D. (2012). Time-series designs. In H. Cooper, M. Paul, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol. 2: Research designs: Quantitative, qualitative, neuropsychological, and biological* (pp. 613-627). Washington, DC: American Psychological Association.

Murphy, K. R., & Russell, C. J. (2016). Mend it or end it: Redirecting the search for interactions in the organizational sciences. *Organizational Research Methods*. doi:10.1177/1094428115625322

Nezlek, J. B. (2001). Multilevel random coefficient analyses of event- and interval-contingent data in social and personality psychology research. *Personality and Social Psychology Bulletin*, *27*, 771-785.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers*, *32*, 396-402.

Ostroff, C., Kinicki, A. J., & Clark, M. A. (2002). Substantive and operational issues of response bias across levels of analysis: an example of climate-satisfaction relationships. *Journal of Applied Psychology*, *87*, 355-368.

Ployhart, R. E., & Ward, A. K. (2011). The "quick start guide" for conducting and publishing longitudinal research. *Journal of Business and Psychology*, *26*, 413-422.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*, 879-903.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, *63*, 539-569.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2016). Recommendations for creating better concept definitions in the organizational, behavioral, and social sciences. *Organizational Research Methods*. Advance online publication. doi:10.1177/1094428115624965

Pollach, I. (2011). Taming textual data: The contribution of corpus linguistics to computer-aid text analysis. *Organizational Research Methods*, *15*, 263-287. doi:10.1177/10944281114177451

Pratt, M. G. (2000). The good, the bad, and the ambivalent: Managing identification among Amway distributors. *Administrative Science Quarterly*, *45*, 456-493.

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*, 717-731.

Preacher, K. J., & Hayes, A. F. (2008). Contemporary approaches to assessing mediation in communication research. In A. F. Hayes, M. D. Slater, & L. B. Snyder (Eds.), *The Sage sourcebook of advanced data analysis methods for communication research* (pp. 13-54). Thousand Oaks, CA: Sage Publications.

Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics: Statistical Issues in Psychology, Education, and the Social Sciences*, *2*, 13-43.

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research*, *42*, 185-227.

Putka, D. (2011, August 31). Re: Construct measurement and validity [Electronic mailing list message]. Retrieved from https://lists.unc.edu/read/?forum=rmnet

Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment*, *12*, 287-297.

Rerup, C., & Feldman, M. S. (2011). Routines as a source of change in organizational schemata: The role of trial-and-error learning. *Academy of Management Journal*, *54*, 577-610. doi:10.5465/AMJ.2011.61968107

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*, 762-800. doi:10.1177/1094428109332834

Rogelberg, S. G., Adelman, M., & Askay, D. (2009). Crafting a successful manuscript: Lessons from 131 reviews. *Journal of Business and Psychology*, *24*, 117-121. doi:10.1007/s10869-009-9116-2

Rothstein, M. G., & Goffin, R. D. (2000). The assessment of personality constructs in industrial-organizational psychology. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment* (pp. 215-248). Norwell, MA: Kluwer Academic Publishers

Sardeshmukh, S. R., & Vandenberg, R. J. (2016). Integrating moderation and mediation: A structural equation modeling approach. *Organizational Research Methods*. doi:10.1177/1094428115621609

Scandura, T. A., & Williams, E. A. (2000). Research methodology in management: Current practices, trends, and implications for future research. *The Academy of Management Journal*, *43*, 1248-1264. doi:10.2307/1556348

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

SIOP. (2003). *Principles for the validation and use of personnel section tests*. Retrieved from http://www.siop.org/_principles/principles.pdf

Spector, P. E. (2006). Method variance in organizational research truth or urban legend? *Organizational Research Methods*, *9*, 221-232. doi:10.1177/1094428105284955

Spector, P. E., & Brannick, M. T. (2010). Common method issues: An introduction to the feature topic in organizational research methods. *Organizational Research Methods*, *13*, 403-406.

Spector, P. E., & Brannick, M. T. (2011). Methodological urban legends: The misuse of statistical control variables. *Organizational Research Methods*, *14*, 287-305. doi:10.1177/1094428110369842

Stigliani, I., & Ravasi, D. (2012). Organizing thoughts and connecting brains: Material practices and the transition from individual to group-level prospective sensemaking. *Academy of Management Journal*, *55*, 1232-1259.

Stone-Romero, E., Weaver, A. E., & Glenar, J. L. (1995). Trends in research design and data analytic strategies in organizational research. *Journal of Management*, *21*, 141-157.

Strauss, A., & Corbin, J. (1994). Grounded theory methodology: An overview. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 262-272). Thousand Oaks, CA: SAGE Publications, Inc.

Tonidandel, S., & LeBreton, J. M. (2011). Relative importance analyses: A useful supplement to multiple regression analyses. *Journal of Business and Psychology*, *26*, 1-9.

Vandenberg, R. J., & Grelle, D. M. (2009). Alternative model specifications in structural equation modeling: Facts, fictions, and truth. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends* (pp. 165-191). New York, NY: Taylor & Francis Group, LLC.

Weitzman, E. A., & Miles, M. B. (1995). *Computer programs for qualitative data analysis: A software sourcebook*. Thousand Oaks, CA: Sage Publications, Inc.

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*, 594-604. doi:10.1037/0003-066X.54.8.594

Wood, R. E., Goodman, J. S., Beckmann, N., & Cook, A. (2008). Mediation testing in management research: A review and proposals. *Organizational Research Methods*, *11*, 270-295.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *99*, 432-442.

Author Biographies

Jennifer P. Green is a doctoral student in the industrial-organizational psychology program at George Mason University. Her research interests include statistical techniques, research methods, and the interaction of personality and situations.

Scott Tonidandel is the Wayne M. and Carolyn A. Watson Professor of Psychology at Davidson College and is a faculty affiliate of the organizational science program at the University of North Carolina-Charlotte. He earned his PhD in industrial-organizational psychology from Rice University. His research spans a variety of different topics, including reactions to selection procedures, leadership, mentoring, diversity, and statistical techniques.

Jose M. Cortina is a professor in the I/O psychology program at George Mason University. He is past Editor of *Organizational Research Methods* and a former Associate Editor of the *Journal of Applied Psychology*. Dr. Cortina was honored by the Society for Industrial and Organizational Psychology with the 2001 Distinguished Early Career Contributions Award, by the Research Methods Division of AOM with the 2004 Best Paper Award, and by the ORM Editorial Board with the 2012 Best Paper Award. He was honored by GMU with a 2010 Teaching Excellence Award and by SIOP with the 2011 Distinguished Teaching Award. Dr. Cortina currently serves as Past President of SIOP.