January 1, 2008

Mediation Testing in Management Research: A Review and Proposals

R. E. Wood
J. S. Goodman
N. Beckmann
Alison Cook, *Utah State University*

Mediation Testing in Management Research: A Review and Proposals

Robert E. Wood
Australian Graduate School of Management
University of New South Wales
Sydney NSW 2042 Australia
Telephone: 61 2 9931 9190
Facsimile: 61 2 9931 9199
Email: rwood@agsm.edu.au

Jodi S. Goodman
University of Connecticut
School of Business
Department of Management
2100 Hillside Road Unit 1041
Storrs, CT 06269-1041
Telephone: 1 860 486 0938
Facsimile: 1 860 486 6415
Email: jodi.goodman@business.uconn.edu

Nadin Beckmann
Australian Graduate School of Management
University of New South Wales
Sydney NSW 2042 Australia
Telephone: 61 2 9931 9187
Facsimile: 61 2 9931 9199
Email: nadinb@agsm.edu.au

Alison Cook
Department of Management and Human Resources
College of Business
Utah State University
3555 Old Main Hill
Logan, UT 84322 - 3555
Telephone: 1 435 797 7654
Facsimile: 1 435 797 1091
Email: alison.cook@usu.edu

Topic area: Mediation

Mediation Testing in Management Research: A Review and Proposals

ABSTRACT

We reviewed and critiqued the conduct and reporting of mediation analyses in 409 studies published in five leading organization studies journals over the past 25 years. The aim of our study was to learn from past practice and to use that knowledge to signal to researchers the importance of correctly applying mediation tests, and to facilitate the valid testing of mediation models and the reporting of mediation results in future studies. We content coded our sample for a wide range of characteristics and found that the majority of inferences of full and partial mediation were based on testing procedures that deviated significantly from procedures recommended by statisticians. In addition, the reporting of results was often incomplete and inefficient. We discussed and evaluated the findings of our study and made recommendations for future testing and reporting of results for mediation models.

As organizational behavior theorists have sought to move beyond descriptions and predictions of phenomena to explanations for how situational and personal factors influence organizational outcomes, statistical tests of mediation processes have become increasingly important to the scientific status of the field. While there are differences in terminology relative to mediation (i.e., indirect effects, intervening variables, mediation), multiple methods for testing mediation (see MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002), and some differences in the criteria for claims of mediation (e.g., Baron & Kenny, 1986; James & Brett, 1984); there is general agreement that mediation occurs when the effects of one variable on another can be explained by a third, intervening variable (Baron & Kenny, 1986; James & Brett, 1984; MacKinnon et al., 2002; Shrout & Bolger, 2002).

In this paper, we reviewed and critiqued the conduct and reporting of mediation analyses in Journal of Applied Psychology (JAP), Organizational Behavior and Human Decision Processes (OBHDP), Academy of Management Journal (AMJ), Personnel Psychology (PPsych) and Administrative Science Quarterly (ASQ) over the past 25 years and provided recommendations to support the correct and consistent application of existing approaches for testing mediation and for reporting results. There are two major reasons for being concerned about the testing and reporting of mediation. First, if the mediation procedures are either incorrectly applied or the results misinterpreted, the validity of explanations for observed outcomes is called into question. Second, inconsistencies in the testing and reporting of mediation across studies obstructs the accumulation of knowledge about organizational phenomena, which is the primary aim of organizational research.

The overall aim of our paper is to help researchers avoid the mistakes of the past and improve the testing of mediation models and the reporting of mediation results in organizational

research. With this in mind, we reviewed authors' choices among existing frameworks and methods of testing for mediation and then evaluated the extent to which they correctly applied methods and how appropriately and effectively they reported their mediation results. We used our critique plus information on the approaches for testing for mediation to develop a set of recommendations intended to assist researchers in choosing among mediation methods, improve the accuracy of the application of those methods and increase the consistency of use and reporting of the methods.

Our analyses and recommendations do not address the relative statistical merits of the different approaches currently accepted by journal editors and reviewers, except as they are germane to our analyses of the applications of the methods. We leave the more detailed analyses of the existing approaches to statisticians (e.g., MacKinnon et al., 2002).

We begin with a review of existing approaches for the testing of mediation. This is followed by a description of our sample of mediation studies published in five leading journals over the past 25 years and the results of our content coding of the sample and analyses of the data. We conclude with a summary of recommendations, drawn from the analyses, for improving the testing and reporting of mediation results.

Overview of Existing Approaches for Testing for Mediation

Following the work of Judd and Kenny (1981), organizational researchers began to include mediation tests in their analyses. In their 1984 article, James and Brett (1984) noted an increasing emphasis on the study of mediation models in industrial psychology and organizational behavior, citing job design, turnover, and leadership attribution research as examples. The impetus for their article was an observation that researchers often predicted linear, additive mediation models, but departed from their arguments by testing and interpreting

findings as if they had predicted moderation. Thus, James and Brett (1984) sought to clarify the meanings of mediation, moderation, and moderated mediation and the roles of variables in these effects.

James and Brett (1984) also pressed upon researchers to more seriously consider the extent to which their data meet the conditions necessary for confirmatory analysis (e.g., specification of causal order, specification of direction of causality, unmeasured variables) and cautioned against the use of causal language when the conditions for inferring causality are not met in a study. The assumption of causality is implicit in the definition of mediation, as a mediator is defined as an explanatory mechanism through which one variable impacts another (Baron & Kenny, 1986; Kenny, Kashy, & Bolger, 1998; MacKinnon, et al., 2002). However, including a mediator in a study does not guarantee that the commonly accepted conditions for inferring causality are met. These include: (1) correlations among variables presumed to be causally linked, (2) temporal precedence of variables that appear earlier in the causal chain (i.e., X precedes M, which precedes Y in time), (3) non-spurious relationships between the variables, (4) properly specified causal order, and (5) a strong theoretical rationale for the explanatory mechanisms among the variables (Cohen, Cohen, West, & Aiken, 2003). When these conditions are not met, James and Brett recommend interpreting results supporting mediation in an exploratory or correlational manner; for example, "the covariation between x and y vanishes if m is controlled" (1984, p. 318).

The motivation for Baron and Kenny's (1986) article similarly had to do with observations about how researchers addressed mediation and moderation. They observed that these terms were frequently (and inappropriately) used interchangeably and sought to distinguish conceptually and statistically between mediation and moderation. With regard to mediation, they

outlined four conditions for deciding whether a variable likely operates as a mediator and a set of regression procedures to test those conditions, based on Judd and Kenny (1981).

MacKinnon and colleagues (MacKinnon et al., 2002) recently identified and compared 14 methods of testing for mediation, intervening variables, and indirect effects. They categorized the methods into three general frameworks: (1) the causal steps approach, (2) differences in coefficients, and (3) products of coefficients. The causal steps approach includes a series of conditions or "rules" for inferring mediation, which vary somewhat across developers (Baron & Kenny, 1986; James & Brett, 1984). Regression analyses and structural equation modeling (SEM) have been recommended to test the conditions of the causal steps framework (Baron & Kenny, 1986; James & Brett, 1984). The Sobel (1982) test, one of the statistical methods associated with the products of coefficients approach, has also been recommended as a supplemental test in the causal steps approach (Baron & Kenny, 1986). Table 1 includes summary outlines of these three approaches and associated statistical tests and some descriptive information from the data we collected, which we will describe in later sections.

INSERT TABLE 1 HERE

The causal steps approach, initially developed in the 1980's by Kenny and colleagues (Baron & Kenny, 1986; Judd & Kenny, 1981), is based on four conditions that need to be met before an inference of mediation can be made. Other authors (see Table 1), most notably James and Brett (1984), have proposed similar approaches but with variations in the conditions required for inferences of mediation. The causal steps approach has low Type-I error rates and low statistical power for detecting mediation effects for small and moderate effect sizes and for large

effect sizes with samples of less than 100 (MacKinnon et al., 2002). The most contentious issue with Baron and Kenny's (1986) recommended analyses for their causal steps approach relates to testing condition 4 listed in Table 1 with comparisons of the sizes of regression coefficients before (b_{yx}) and after the mediator is included in the analysis ($b_{yx.m}$). Baron and Kenny (1986) did recommend that the Sobel (1982) test be used to test the significance of the change in the coefficient due to the introduction of the mediator. However, the widely adopted recommendation is the interpretation of a change in the significance of the regression coefficient (i.e., b_{yx} is significant and $b_{yx.m}$ is non-significant) as grounds for inferring full mediation and a reduction (i.e., $b_{yx.m}$ is smaller than b_{yx} but still significant) as grounds for inferring partial mediation. This is problematic because inferences of mediation are made without any assessment of the statistical significance of the mediation effect (MacKinnon et al., 2002).

James and Brett (1984) specified similar conditions to Baron and Kenny (1986) for the bivariate relationships between the independent variable and the mediator and mediator and the dependent variable, conditions 1 and 2 as listed for each approach in Table 1. The condition that the independent and dependent variables are no longer related when the mediator is controlled is also similar for the two causal steps approaches; that is, condition 4 for Baron and Kenny (1986) and condition 3 for James and Brett (1984) as noted in Table 1. However, the evidence required to satisfy this condition differs between the two approaches. Instead of the changes in regression coefficients recommended by Baron and Kenny (1986), James and Brett (1984) recommend that once conditions 1 and 2 are met, inferences of mediation be based on R^2. Specifically, James and Brett (1984) require that the independent variable add nothing to the prediction of the dependent variable over that already explained by the mediator (i.e., $R^2_{y.mx}$ is not significantly greater than $R^2_{y.m}$). This rule of evidence poses the same problem as Baron and Kenny's (1986)

recommendation about changes in regression coefficients in that the inference of mediation is not based on a statistical test of the indirect effect. James and Brett explicitly note that there is no analog to test for the indirect effect in OLS regression (1984, p. 319).

In addition, James and Brett's (1984) condition 4 suggests that the mediator should add uniquely to the prediction of the dependent variable in relation to the independent variable (i.e., $R^2_{y.mx}$ is significantly greater than $R^2_{y.x}$), although it is not entirely clear whether they intended to require this as a condition for mediation. It is germane to our later discussion to point out that, by itself, this rule can lead to a misattribution of mediation as this change in R^2 is not due to the mediation of the independent – dependent variable relationship, *per se*. Rather, as James and Brett (1984) stated, it relates to the additive effects that the proposed mediator has on the dependent variable, over and above the effects of the independent variable on the dependent variable. If the independent – dependent variable relationship is fully mediated through the independent variable's effects on the mediator and the mediator's effects on the dependent variable, in its role as a mediator, the variable should not add to the variance in the dependent variable explained by the independent variable; it should replace it. A mediator is a mechanism that accounts for the impact of the independent variable on the dependent variable (Cohen et al, 2003). Any additional variance explained by the mediator does not preclude its role as mediator, but it is evidence of an additive effect, not evidence of mediation.

Originators of the causal steps approach recommended that structural equation modeling (SEM) be used as an alternative to regression in tests for mediation when multiple indicators are collected for variables to address measurement unreliability (Baron & Kenny, 1986), when the conditions for confirmatory analysis have been met (i.e., accurate specification of causal order and direction, no unmeasured variables problem, relationships are truly linear, relationships are

stationary; James & Brett, 1984), and when a model includes latent constructs (Kenny et al., 1998). MacKinnon (2000) addressed the use of SEM for complex models that include multiple mediators and/or dependent variables. He reported that standard SEM packages compute only total mediated and direct effects and their standard errors, and refers readers to Bollen (1987, cited in MacKinnon, 2000) for matrix routines used to test the effects of individual mediators. SEM enables the tests of more complex mediation models than the simple X->M->Y model discussed in James and Brett (1984) and Baron and Kenny (1986), including the simultaneous testing of multiple paths, with full statistical controls for the relationships between variables within sets of independent, mediator and dependent variables. This reduces the risks of incorrect inferences for effects that may be due to multicolinearity within sets of variables or a chance finding among multiple tests, which increase when complex models are broken down into simple models and tested separately. If used properly, there are also statistical benefits to using SEM (Baron & Kenny, 1986; Cohen et al., 2003; Kenny, et al., 1998; Shaver, 2005), which are beyond the scope of this paper. Precise guidelines for how to test for mediation using SEM have been proposed only recently (James, Mulaik, & Brett, 2006; see also Mathieu & Taylor, 2006), and, as we will discuss later, a variety of criteria have been used to make inferences about mediation from SEM analyses.

When a complex model is proposed, but the study design does not satisfy the conditions for SEM, researchers can turn to MacKinnon (2000) and Cohen and colleagues (Cohen & Cohen, 1983; Cohen et al., 2003) for variations of the causal steps regression approach to mediation. MacKinnon (2000) described how to extend the causal steps approach to multiple mediator models using a regression approach, and presented a procedure for computing individual mediator effects and their standard errors for multiple mediator models (MacKinnon, 2000).

Cohen and colleagues (Cohen & Cohen, 1983; Cohen, et al., 2003) discussed using hierarchical regression with sets of independent variables and mediators, estimating the "net" (Cohen et al., 2003, p. 467) mediation effect across the set of mediators, and decomposing the effects of each variable into direct, mediated, and spurious components. They did not, however, address significance testing for mediation effects. Instead, their examples were based on visual inspection of relative sizes of and changes in effects. They noted that the significance of the two direct effects multiplied to produce the indirect effect is sufficient for the indirect effect to be significant (Cohen & Cohen, 1983).

The differences in coefficients approach was developed in various fields in the 1980's and 1990's and involves statistically comparing coefficients before and after adjustment for the mediator (MacKinnon et al., 2002). The products of coefficients approach was developed in sociology starting in the 1940's (Aroian, 1944, cited in MacKinnon et al., 2002). It includes the Sobel (1982) test and its variants, among other tests, which involve testing for indirect effects using a path model. Both the differences in coefficients and products of coefficients tests provide estimates of the standard error and assessment of the statistical significance of the mediation effect (MacKinnon et al., 2002). The product of coefficients is algebraically equivalent to tests of the change in the regression coefficient following the introduction of the mediator (i.e., b_{yx} - $b_{yx \cdot m}$) (MacKinnon et al., 1995), which, we presume, was the reason Baron and Kenny (1986) suggested the Sobel test as a possible test of the fourth condition in their causal steps approach. MacKinnon et al. (2002) reported several conclusions regarding the Type 1 errors and statistical power of the different types of tests for the differences in coefficients and products in coefficients approaches. Overall, they report accurate or low Type 1 error rates and higher power to detect mediation effects compared to the causal steps approach.

A criticism of the Sobel test is that it tends to be conservative when effect and sample sizes are small because non-normal effect size distributions associated with small sample sizes violate the normality assumption associated with the test statistic (Bollen & Stine, 1990; MacKinnon, et al, 2002; Preacher & Hayes, 2004; Shrout & Bolger, 2002). Bootstrapping can correct this problem and can increase statistical power to detect mediation effects. It involves estimating the standard errors used in the calculation of p-values and confidence intervals from a distribution created through a process of repeated re-sampling with replacement from the data (Shrout & Bolger, 2002). Shrout and Bolger (2002) report that bootstrapping is well known to statisticians, is an option in popular SEM programs, and began to appear in the psychology literature in the late 1990's. The purpose of their 2002 article was to encourage the widespread use of bootstrapping.

Study: Review and Critique of Mediation Testing in Organizational Research

Sample

Published articles that reported mediation tests in JAP, OBHDP, AMJ, PPsych and ASQ for the 25 years from January,1981 to August, 2005 were identified and coded. The selected years were chosen to provide an assessment of the testing and reporting of mediation up to the present, including changes between the years before and following publication of the James and Brett (1984) and Baron and Kenny (1986) articles. The five journals sampled were chosen to provide breadth of coverage of the different types of empirical organizational studies published in leading journals. Across the five journals, there is variety in the study designs (cross sectional surveys, laboratory and field experiments, meta-analyses and longitudinal studies) that include tests of mediation. The five journals also include studies at different levels of analysis (individual, group, work unit and organization). While the focus in this review was on the testing

and reporting of mediation in published studies, we acknowledge that we were unable to address differences between papers rejected and accepted for publication.

We searched for variations of the word "mediate" and its synonyms (indirect effect, intervening variable) in article titles, abstracts, and keywords using the PsychInfo database (American Psychological Association). We excluded studies for which no statistical mediation test was reported, studies using qualitative methods and theory articles. We then supplemented our search with full text searches of all articles from the years 1981, 1986, 1991, 1996 and 2000 in the five journals. The full text searches yielded an additional 63 studies across those five years because authors of those studies used our search terms in the bodies of their articles, but not in the titles, abstracts or article keywords. Together the two searches yielded 409 studies, published in 368 articles. The studies reported tests for 709 different mediation models and 787 mediation effects. The number of effects reported was greater than the number of models tested because authors often reported mediation effects for individual variables when they tested multiple mediators in a single model.

We content coded the studies identified from our searches for a variety of mediation-related characteristics. Some key categories of variables included: mediation framework used and sources cited, study design, mediation conditions or "rules" tested, types of statistical analyses used, time ordering of variables, and bases of claims of full and partial mediation. The coding sheet was developed and then tested through the coding of 100 articles drawn from the journals at five year intervals across the 25 years covered by the sample. As a result of this test coding, categories, definitions and instructions on the coding sheets were refined. The 100 test articles were then recoded using the refined coding sheet. Four graduate students were trained as coders by one of the authors, who acted as the lead coder. The lead coder recoded and cross

checked 10 articles coded by each of the graduate students and provided additional training as needed. Throughout the process, coding questions were discussed with the lead coder and resolved according to coding rules agreed upon by the authors.

Results

We organized the results into two sections. In the first section we provided an overview of the data and reported on the major descriptive results from our coding of the published studies. These included summary data on trends in the frequency of testing and reporting of mediation effects over time for the five journals in our sample, the types of study designs used in our sample, the distribution of claims for full and partial mediation, the sources cited by authors for their approaches to testing for mediation over time, and the general data analysis methods used to test for mediation. In the second section we presented results of evaluative analyses of: (1) variations between the statistical analyses used to test for mediation and the procedures recommended by statisticians, (2) the bases of inferences of full and partial mediation, (3) the testing of complex mediation models, (4) the appropriateness of claims of causality, and (5) the quality of reporting of results. We also included an analysis of predictors of claims for full and partial mediation. The results of these analyses were used to derive our recommendations for the future testing and reporting of mediation effects.

Overview of the sample and major descriptive statistics

Across the 25 years, almost half of the mediation studies in the sample were reported in JAP (48%). OBHDP (22%) and AMJ (18%) were the other main sources of studies. PPsych (8%) and ASQ (5%) published far fewer mediation studies. Figure 1a shows the cumulative frequencies of mediation articles published in each of the five journals in our sample over the 25 years. The dominance of JAP as a source of mediation studies in our sample was only partly due

to the larger number of papers published in that journal. Across the 25 years reviewed, the percentage of papers that reported mediation studies (and the total number of papers published) for each of the journals sampled were: JAP 7.3% (179 of 2457 papers), OBHDP 5.8% (71 of 1218 papers), AMJ 6.2% (70 of 1126 papers), PPsych 2.8% (30 of 1080 papers), and ASQ 5.1% (18 of 356 papers).

INSERT FIGURE 1A HERE

There was a trend of increased reporting of mediation tests over the two and one half decades covered by the studies sampled. Across all five journals, there was one study reporting mediation tests published in 1983, and 39 through August in 2005. Most of this increase was due to increasing numbers of mediation studies published in JAP and OBHDP and, to a lesser degree, AMJ. In the period 2001 to 2005, approximately one in seven (15%) of the papers published in JAP reported mediation models, up from one in 50 (2%) for the period 1985 to 1989. The equivalent proportions of mediation studies for OBHDP were one in five (19%) published papers from 2001 to 2005, up from one in 33 (3%) from 1985 to 1989, while AMJ increased from one in 33 (3%) to one in eight (12%). Papers reporting mediation studies in PPsych and ASQ ranged between zero and five per year for the full period of the sample. Beyond the longer-term trend across all 25 years, there was no increase in the number of mediation papers published in the years immediately following the publications of James and Brett (1984) and Baron and Kenny (1986), compared to the sample years preceding the publication of those articles. An additional trend we noted is that the tendency to confuse

moderation and mediation has largely disappeared, but not totally. In the journals we examined, two papers published since 2000, one in 2005, confused the terms.

There was a fairly even split between experimental or quasi-experimental designs (50%) and field studies (49%). Four studies (1%) reported mediation tests within a meta-analysis. The experimental (39%) and quasi-experimental (11%) designs included time ordering of the independent, mediator and dependent variables, consistent with one of the requirements for causal inferences (Cohen et al., 2003), and some form of control or comparison group for the assessment of the independent variable effects. The ratio of experiments to field studies varied from year to year and averaged 1:1.26 between 1980 and 2001. Since 2002, there has been an increase in the number of experiments relative to field studies that report mediation tests, with a ratio of 1:0.87 from 2002 to 2005.

There was a strong bias for significant results. Inferences of mediation effects were reported for 595 (75.6%) of the 787 effects in our sample, including 422 (54%) that were used to infer full mediation and 173 (22%) used to infer partial mediation. This is interesting given the low power of many of the methods used to test for mediation (MacKinnon et al., 2002). A search inclusive of unpublished studies would almost certainly yield a higher number of non-significant results.

Figure 1b shows the cumulative frequencies of studies by the sources cited for the four most commonly cited sources of methods for testing for mediation, which together accounted for 63% of the total citations in our sample. Ninety-four (23%) of the studies cited no source for their tests of mediation. Across the 25 year period, the Baron and Kenny (1986) causal steps approach was cited in 52% of the papers, and their method was applied in 51% of the studies in our sample. The James and Brett (1984) approach was cited in 12% of the papers, and their

method was used in 10% of the studies. The Cohen and colleagues (Cohen & Cohen, 1983; Cohen et al., 2003) approach was cited in 10% of the papers, and their method was applied in 6% of the studies. Considering that some studies used multiple approaches, a total of 399 studies (98% of the sample) employed some form of the causal steps approach, including the 94 studies that did not cite any source for their approach.

In our sample, products of coefficients tests were reported only since 2000 and 14 (50%) of the 28 studies that included such tests were published in 2005. The Sobel (1982) test was the most popular of the products of coefficients approaches and was cited and applied in all 28 studies (7% of studies in the sample). These 28 studies included three studies that cited and tested a combination of Sobel (1982) and Goodman (1960), MacKinnon and Lockwood (2001) or MacKinnon et al. (1998). Three studies (1%) used the differences in coefficients approach (Clogg, Petkova, & Shihadeh, 1992; Olkin & Finn, 1995), including one in combination with a products of coefficients test. The majority of the products of coefficients tests (19 out of 28 studies) were used in combination with the causal steps approach. Table 1 includes the number of citations and applications of each of the various methods of testing for mediation.

INSERT FIGURE 1B HERE

The distribution of the 409 mediation studies in the sample by statistical analyses used and sources cited are shown in Table 2. Across all types of study designs, the most common type of statistical analysis employed in the testing of mediation effects was regression, which was used in 63% of the studies reported. Four percent of the studies used partial correlation to infer mediation and 8% used analysis of covariance. Neither of these methods is explicitly

recommended by statisticians for testing for mediation, but both are comparable to regression. The use of SEM to test for mediation effects has grown significantly in papers published since 1990, and a total of 102 studies (25%) have employed that statistical technique in the past 25 years, including 92 studies since 1990.

Many studies used a combination of types of analyses to test for mediation to either supplement one another or to check for similarity in results. Most commonly, a correlation was used to test the independent-dependent variable relationship and the other conditions in the causal steps approach were then tested using regression (52 studies, 13%), analysis of covariance (7 studies, 2%), SEM (14 studies 3%), or partial correlations (7 studies, 2%). Six studies relied solely on bivariate correlations between the independent, mediator and dependent variables to make inferences about mediation. This is not consistent with strategies recommended by statisticians because it does not assess the extent to which the mediator accounts for the relationship between the independent and dependent variables.

INSERT TABLE 2 HERE

Conclusions that can be drawn from our initial look at the 25-year sample of articles are as follows. (1) Causal steps has been the most common approach to mediation, supplemented in recent years with the products of coefficients approach, primarily in the form of the Sobel (1982) test. (2) Regression has been the most common statistical method used. (3) Baron and Kenny (1986) has been the most frequently cited source of guidance for mediation tests in organizational psychology and behavior. (4) Other available sources that either cover issues not addressed by Baron and Kenny (e.g., McKinnon et al., 2002; Shrout & Bolger, 2002) or suggest

supplemental, alternative, and potentially more appropriate approaches for the hypothesized models (e.g., Cohen et al., 2003; MacKinnon, 2000; MacKinnon, Warsi, & Dwyer, 1995; Shrout & Bolger, 2002) so far have been rarely cited. Another feature to note is that the number of experiments, which emphasize conditions for causal inferences, has increased as a proportion of total mediation studies in recent years. Experiments have not been replaced by field studies, whose effects are often assumed to be more likely to generalize to applied contexts.

Evaluative analyses

The following analyses address characteristics of the studies and mediation tests in our sample that are suggestive of potential threats to the validity of inferences of full and partial mediation in many of the studies.

Adherence to recommended testing procedures. Table 3 shows the numbers of studies, plus the models and effects included in those studies, that adopted the statistical procedures for the products of coefficients, differences in coefficients and causal steps approaches to mediation (MacKinnon, et al, 2002), and the numbers of claims of full, partial and no mediation for the effects under each approach. The studies in the causal steps category are further broken down into those that used SEM and those in which regression or some equivalent method (ANCOVA, partial correlations, etc.) was used. This final group excludes all studies in which products of coefficients, differences of coefficients or SEM were used in combination with regression or equivalent methods. Those studies are included in the products of coefficients, differences of coefficients or SEM categories.

INSERT TABLE 3 HERE

Because the causal steps approach specifies multiple conditions and requires a series of analyses when regression or equivalent methods are used, the effects in the studies that used these methods alone were further divided into those for which all causal steps conditions were tested and those for which an incomplete set of conditions was tested. The testing of an effect using regression or an equivalent method was categorized as incomplete when either one or more of the causal step conditions was not tested or an inappropriate analysis was used. For example, tests of the independent – dependent variable relationship with a bivariate correlation is analogous to using simple regression and therefore was considered correctly tested, whereas, using a bivariate correlation to test Baron and Kenny's (1986) condition 4 is not equivalent to the procedure they recommended because the mediator must be controlled. This would be considered incorrectly tested and categorized as incomplete. Alternatively, computing a partial correlation $r_{yx.m}$ to test condition 4 is appropriate because it controls for the mediator in the same way as multiple regression.

We were unable to report equivalent information for the other procedures reported in Table 3. The products of coefficients and differences in coefficients procedures each require a single test, and while statistical assumptions may be violated in the use of the procedures, we were not able to detect violations from what was reported in the studies. In addition, we were unable to reliably identify the exact procedures followed when SEM was used to test for mediation in many of the studies.

The data in Table 3 revealed that 375 (65%) of the 574 effects that were tested using regression or equivalent methods were based on testing an incomplete set of the casual steps conditions specified by James and Brett (1984) or Baron and Kenny (1986). This represents 48% of the 787 effects in the total sample and includes 40% of the 422 inferences of full

mediation, 53% of the 173 inferences of partial mediation, and 61% of the 192 inferences of no mediation effect.

Recently, some statisticians have suggested that the Baron and Kenny (1986) condition for the significant relationship between the independent and dependent variables should not be required when small effect sizes are predicted or when the mediator acts as a suppressor of the independent-dependent variable relationship (e.g., Shrout & Bolger, 2002). For an additional view of the data, we removed from the incomplete procedure category the effects for which only the independent-dependent relationship was either not reported or was insignificant and considered those complete tests for the moment. We found that 32% of all claims of a full mediation effect and 44% of all claims for a partial mediation effect were based on an incomplete set of procedures. This alternative analysis assumes, of course, that authors had legitimate reasons for not requiring the independent-dependent variable relationship.

Thus, even under quite liberal interpretations, based on the untested assumption that all studies that used products of coefficients, differences of coefficients and SEM reached valid conclusions regarding mediation effects and that the independent-dependent variable relationship was not relevant for the model tested, 40% of all claims regarding mediation (full, partial or no effects) were based on tests of an incomplete set of causal steps conditions.

It is worth noting that complete testing of the causal steps conditions was unrelated to conclusions of full, partial or no mediation, at least for our sample of published studies. This suggests that authors are not omitting tests of conditions so they can report significant and potentially more publishable results, but that possibility cannot be completely ruled out. In addition, at least some of the incomplete procedures used when inferences of no mediation effects were made may be due to authors concluding that there is ‘no mediation effect’ after

finding a non-significant relationship for one of the first three conditions of the causal steps approach, and then not conducting or reporting tests for subsequent conditions. This is an appropriate strategy. However, for the majority of effects categorized as incomplete, inferences of full or partial mediation were made. We should also emphasize that those effects categorized as complete did not include a test for the significance of the mediation effect. We address this issue in the next section.

Bases of claims for full and partial mediation. The majority of inferences for full mediation and partial mediation effects in our sample were based on changes in coefficients, as recommended by Baron and Kenny (1986). As we previously discussed, this strategy does not assess the magnitude of the change and, therefore, can lead to invalid conclusions. We also found a great deal of variance in authors' bases for their claims of full mediation in SEM and for claims of partial mediation across studies that used SEM, regression or other related procedures. First we discuss bases for claims of full mediation, then partial mediation.

As we indicated earlier, full mediation was inferred for 422 of the 787 reported effects in our sample. Fifty percent (210 effects) of the full mediation claims were based on changes from significance to non-significance in the coefficient for the independent - dependent variable relationship following the inclusion of the mediator (i.e., $b_{xy} > 0$; $b_{xy.m} = 0$). This is consistent with one of Baron and Kenny's recommendations, but ignores their recommendation to follow up with the Sobel (1982) test of significance for the indirect effect. This could be because Baron and Kenny (1986) discuss the Sobel (1982) test, but do not directly state that it must be done or discuss how to interpret it relative to making inferences for full versus partial mediation.

Only 41 (9.7%) of the claims for full mediation effects were based on a test of whether the magnitude of the change in the coefficient for the independent variable following the

introduction of the mediator was significant. These 41 effects came from 28 studies (7%) that tested the product of coefficients (Goodman, 1960; MacKinnon & Lockwood, 2001; MacKinnon, Lockwood, & Hoffman, 1998; Sobel, 1982). Only one of the 28 studies that used a Sobel (1982) test reported the use of bootstrapping, which can be useful for addressing low power when sample sizes are small to moderate (Shrout & Bolger, 2002). Studies in our sample reporting a Sobel test had a median sample size of N = 183 (Quartiles: 116; 269; Range: 57; 16,466). Seventy-five percent of the studies that reported a Sobel test had samples of 116 or more, which is sufficient to detect moderate and large effects, but not small effects (MacKinnon et al., 2002).

One hundred and one (24%) of the inferences of full mediation effects in our sample were based on SEM analyses, and the claims were based on a range of criteria. Most authors inferred full mediation when a model excluding direct effects of the independent variable on the dependent variable exhibited better fit than a model including both direct and indirect effects. Other authors based their conclusions of full mediation on one or more of the following findings: (1) significant linkages between the different variables in the hypothesized mediation model, (2) good model fit of the proposed meditation model, (3) better model fit of the proposed model compared to one or more alternative models, and (4) statistically significant indirect paths(s) and/or insignificant direct paths(s) in the proposed model.

Few researchers, and the number has been decreasing over time, have taken account of the James and Brett (1984) recommendation regarding the impact of mediators on R^2. Forty-five (11%) inferences of full mediation effects in our sample were based on changes in R^2 following the introduction of the mediator to the regression model (whether $R^2_{y.mx}$ was significantly greater than $R^2_{y.x}$). Conversely, a lack of change in $R^2_{y.mx}$ was interpreted, typically along with other

evidence of lack of support, to infer no mediation effect in 17 cases. As previously noted, this change in R^2 says nothing about whether the proposed mediator explains (i.e., mediates) the relationship between the independent and dependent variables. Nevertheless, it has been used as a basic rule for establishing mediation by the authors of some of the studies in our sample.

Another 9 (2%) claims of full mediation effects were based solely on significant bivariate correlations between the independent, mediator and dependent variables. This strategy is incorrect because the correlations among the set of variables show only that they are related, but says nothing about whether the mediator accounts for the relationship between the independent and dependent variable. For the remaining 16 claims of full mediation, it was not possible to determine the grounds for the claims from the articles.

The types of evidence used for the 173 inferences of partial mediation (22% of the 787 total effects) were many and varied. The products of coefficients approach was used for 15 of these claims. Inferences were based primarily on a significant Sobel test plus a significant $b_{xy.m}$ (11 effects); however, it was not possible to determine the grounds for four of the inferences. One claim for partial mediation was based on the difference of coefficients approach. The authors of this study inferred partial mediation from a significant difference of coefficient test and a significant reduction in the variance explained in the dependent variable by the independent variable after controlling for the mediator. The direct relationship between the independent and dependent variables remained significant after controlling for the mediator.

Regression or equivalent analyses were used to test the causal steps approach for 134 inferences of partial mediation. Among the more common grounds for these partial mediation inferences were: (1) “marginal” significance of the regression coefficient for the effect of the independent variable on the dependent variable after adding the mediator to the regression

model; (2) a decreased, but still significant, coefficient for the effect of the independent variable after introducing the mediator, which was also significant, as recommended by Baron & Kenny (1986); and (3) a percent decrease in the coefficient for the effect of the independent variable, which remained significant after the mediator was introduced into the regression equation. None of these included a test of the significance for the change in coefficient, so the grounds for inferring partial mediation are questionable.

SEM was the chosen statistical analysis method for 23 inferences of partial mediation. The basis for these inferences was typically the finding that a model including direct and indirect paths between the independent and dependent variables exhibited better model fit than the proposed full mediation model, which excluded the direct path from the independent to the dependent variable.

In summary, studies that included some significance test of the change in coefficients and correctly interpreted the results of that test or used SEM, assuming SEM was used appropriately, were responsible for 142 (34%) of the claims of full mediation and 39 (23%) of the claims for partial mediation. Conversely, 280 (66%) and 134 (77%) of the claims for full and partial mediation effects, respectively, were based on questionable grounds and therefore are potentially invalid.

Issues regarding testing complex mediation models. Fifteen percent (109) of the 709 models in our sample were hypothesized in the simple form, involving one independent variable, one mediator and one dependent variable (X->M->Y). This is the type of model addressed in the sources most commonly cited for the causal steps approach (Baron & Kenny, 1986; James & Brett, 1984). The other 85% (600 models) were from studies that hypothesized more complex

mediation models, including multiple independent variables, multiple mediators, multiple dependent variables or some combination of these.

SEM was used to test 91 of the complex models hypothesized in our sample. SEM can accommodate multiple mediators as well as other sources of complexity, such as multiple independent variables and multiple dependent variables in a mediation model (MacKinnon, 2000). Of the other complex models, 319 were analyzed by including multiple mediators, multiple predictors or both in a single regression model. This approach allows for the interpretation of coefficients similar to a simple model but the effect of the loss of degrees of freedom should be considered. In one illustrative case, seven mediators were included in a single model. The elimination of the significant correlation between the independent variable and dependent variable ($r = .26$, $p = .046$) when the seven proposed mediators and the independent variable were added to the regression equation ($b_{yx.mi} = .23$, $p = .052$) was used to infer a full mediation effect, attributed to the one mediator that had satisfied other conditions of the causal steps approach. However, mediation is only one possible explanation for the results. The change in significance of the coefficients from $p = .046$ to $p = .052$ also could have been due to the lower degrees of freedom associated with including the seven proposed mediators or could be due to chance.

The remaining 190 complex models were broken down and tested as a series of simple X->M->Y models. One study included Bonferroni corrections for the family-wise Type-I error rate that can occur when conducting multiple, related tests, suggesting that some of the claims of mediation may have capitalized on chance effects.

As we mentioned earlier, there are existing methods for analyzing multiple mediator models in addition to SEM. However, neither MacKinnon's (2000) recommendations for

extending the causal steps approach to multiple mediator models nor Cohen and colleagues (Cohen & Cohen, 1983; Cohen et al., 2003) hierarchical regression approach for testing sets of mediators was used in our sample.

Issues regarding causality. As James and Brett (1984) discuss, and as we found in our sample of studies, conditions for causality are often not met, especially in field studies. However, authors still make causal claims when reporting results rather than following James and Brett's (1984) recommendation to use non-causal language and discuss effects in terms of covariation. In our sample, claims of causality were made or implied for 471 (66%) of all mediation models tested. Just over half of these claims of causality (52% or 247 claims) were for models tested in non-experimental designs, including 233 models in which measures of the independent, mediator, and dependent variables were all collected at the same time. For 10 studies (17 models) in our sample, we were unable to determine whether there was a lag in the measurement of the variables, and in 9 studies (15 models), the order of measurement was different from the proposed causal order (e.g., Y was measured before M). We acknowledge that the temporal order of the phenomena is vital for causal inference, not the order of measurement, *per se*. However, we were unable to ascertain the temporal order of the phenomena, and, given the nature of the constructs in organizational research, the study authors may not have been able to either.

In addition, for 204 (29%) of the models reported in our sample, measures of all variables in the model were self-reported perceptual data from a single source. When single methods are used, nuisances, such as response sets of participants or another bias in the chosen method that similarly impacts the measurement of all variables can account for or partially account for the relationships among variables. This common method variance should, therefore, be tested and

ruled out or reported as an alternative explanation for the results of mediation tests based on single source, self-report data (Spector, 2006).

Reporting of mediation results. The reporting of statistical data is an important part of the development of knowledge in any discipline. Incomplete reporting makes it difficult for readers of published studies to assess the validity of the claims made and for researchers to systematically synthesize results and analyze the cumulative results for particular effects and models. Of the 409 studies in our sample, 84 (21%) provided no separate tables for reporting the mediation results and only referenced the mediation results in the text. These reports were typically incomplete and did not present enough information for the reader to be clear about the grounds for inferring support for either full or partial mediation. Another common form of reporting was the use of path diagrams, with coefficients recorded along the paths, which was done in 137 (34%) of the studies, usually in combination with tables. When it was used alone in 34 studies (8%), it provided incomplete information about tests of mediation. Most common was the reporting of the results for mediation in several different tables, which were then drawn together in the text. This was done in 236 (58%) of the studies in the sample. While separate tables avoid the repetition of information, they may lead to certain details being excluded or not being immediately obvious to the reader. For example, we found differences in sample sizes across different tables, possibly due to different ways of handling missing data in different analyses (e.g., pairwise in correlation tables, listwise in regression tables). Fifty-five (13%) of the studies in our sample reported mediation results in a single table, which is a more accessible and, typically, a more complete reporting format. Use of a single table has been increasing since 1995.

Inadequate description of the conditions for inferences of mediation tested and steps followed in the analyses can also make it difficult to ascertain what was done. When using the causal steps approach, many authors simply referenced a source and do not specify the conditions they were testing. Just over half of the studies (53%, or 110 of 207 studies) that cited Baron and Kenny (1986) as their authority for mediation testing, specified which of the conditions they followed. This dropped to 39% (16 out of 41) for the studies that cited James and Brett (1984) and to 16% (4 out of 25) of the studies that cited Cohen and colleagues (Cohen & Cohen, 1983; Cohen et al., 2003).

Across all journals, the lack of detail and the variations in the reporting of mediation results limit the feasibility of summarizing reported effects, such as might be required for meta-analyses. Fortunately, the most comprehensive reporting occurred in the two journals, JAP and OBHDP, that also published the most mediation studies and, in recent years, papers published in AMJ and PPsych have also included more complete information.

Predictors of inferences. As a final step, we conducted a series of multiple regression analyses to see if any of the coded variables in our study predicted the claims made for mediations effects. These analyses all produced the same finding, that the use of SEM predicted the likelihood of a significant claim of mediation, so we report two illustrative analyses. In the first regression analysis, the dependent variable was the claim of full mediation for the first model tested in each of the 409 studies; dummy coded as 1= full mediation and 0 = partial or no mediation. In the second regression analysis the number of mediation effects (full or partial) reported across all 709 models tested in the sample was the dependent variable.

For the first model tested in each of the 409 studies, the likelihood of a full mediation claim was greater when SEM was used ($Chi2 = 17.01$, $df = 7$, $p < .05$, $N = 409$, $B = 1.04$, $SE =$

.37, Wald = 7.82 (df=1), $p < .01$). No other variable predicted claims of full mediation. Across all 709 models in the sample, the frequency of inferences of partial or full mediation effects were similarly predicted by use of SEM ($R^2 = .35$, $F_{8/400} = 28.61$, $p < .001$, $N = 409$, beta = .15, $t = 2.77$, $p < .01$) and also by the number of models tested in each study (beta = .60, $t = 14.06$, $p < .001$), which was included as a control variable.

The findings that the use of SEM predicts reporting of significant effects may be due to the better statistical controls for complex models and power of the technique. The lack of significant predictors among all other coded variables indicates that there are no obvious systematic biases in testing procedures and mediation reporting that are related to claims of significant effects for the published studies in our sample. A study of published and unpublished studies may reveal a different picture.

Summary of evaluative analyses. Our evaluative analyses show that the main threats to the validity of the inferences of mediation in the studies reviewed arose from: (1) incomplete or inaccurate application of existing approaches for testing mediation; (2) basing claims of full or partial mediation on changes in the magnitude of coefficients, without testing the significance of that change; (3) using procedures developed for simple mediation models (X->M->Y) to test complex models, instead of using available procedures for the analyses of complex models; and (4) making causal claims when conditions for causality are not met. In addition, the incomplete descriptions of the conditions for inferences of mediation followed and the incomplete or inefficient reporting of results of mediation tests make it difficult for readers to judge the validity of mediation inferences and impede the accumulation of knowledge across studies. Overall, while we cannot say with certainty that the majority of mediation inferences in the sample are invalid, our analyses revealed significant sources of potential threats to their validity.

Summary of Recommendations

Based on our review and analyses, we put forth a set of recommendations for improving the testing and reporting of mediation results. Our recommendations do not break new ground in this arena. Nevertheless, we believe it is important to state them because our analysis of 25 years of mediation research shows that a majority of conclusions were based on incomplete or inappropriate testing procedures, a lack of direct statistical evidence for the inferences of full and partial mediation, overstatement of claims for causality and insufficient reporting of results. Overall, high standards have not been applied in the testing and reporting of mediation. If our recommendations are adopted by researchers, reviewers, and editors, we believe they will significantly advance the quality of inferences and facilitate the accumulation of knowledge about mediation mechanisms in organizational psychology and behavior.

We organized our recommendations around the first five subsections of the preceding evaluative analysis section. We present the recommendations as a set of succinct and direct statements, partially to limit redundancies with our results section, but primarily in an attempt to avoid the apparent ambiguity in how original sources for mediation testing have been interpreted by authors.

1. As a general rule, authors should abide by all of the conditions specified by the statisticians whose work they cite. For instance, when Baron and Kenny's (1986) causal steps approach is used, all four conditions should be examined, and tests of the conditions should be supplemented with a test of differences in coefficients or products of coefficients, such as the Sobel (1982) test. Readers should consult MacKinnon et al. (2002) for available options and a statistical comparison of the options and Shrout and Bolger (2002) for information on using bootstrapping to estimate standard errors. When authors choose to skip a condition or test, the

approach should be explicitly acknowledged and justified.

2. Inferences of full, partial and no mediation should be grounded in sound statistical testing. As we discussed earlier, an observed decrease in $b_{yx.m}$ compared to $b_{y.x}$ is insufficient grounds for inferring mediation effects. The change in regression coefficient needs to be tested for statistical significance. This can be tested using a difference of coefficients test or a product of coefficients test. Product of coefficients tests, such as the Sobel (1982) test will be significant only when the change in coefficients is also significant, and therefore provide grounds for an inference of significant change. An inference of partial mediation requires that there be a significant change in coefficients plus a remaining significant direct relationship between the independent and dependent variables after controlling for the mediator. In addition, finding that $R^2_{y.mx}$ is significantly greater than $R^2_{y.x}$ is indicative of an additive effect, and, while it certainly does not preclude a mediation effect, it is not diagnostic of mediation effects, one way or the other.

We direct readers to recent methodological articles that advocate *a priori* specification of hypotheses for full mediation, partial mediation (James et al., 2006; Mathieu & Taylor, 2006) or indirect effects (i.e., where no independent-dependent variable relationship is anticipated; Mathieu & Taylor, 2006). James et al. (2006) and Mathieu and Taylor (2006) discuss the focal statistical models for these different hypotheses and outline strategies for testing the models using SEM. Their specification of strategies for testing for and drawing conclusions about mediation with SEM should be particularly helpful given the variety of grounds for inferring full and partial mediation we found in our sample.

3. Appropriate tests of mediation should be chosen to accommodate complex models, such as those that include multiple mediators and/or dependent variables. Available options include:

(1) the hierarchical regression approach described by Cohen and colleagues (Cohen & Cohen, 1984; Cohen et al., 2003), supplemented with a statistical test for the size and significance of the mediation effect; (2) SEM, if the sample size is sufficiently large and requisite statistical assumptions are met; (3) MacKinnon's (2000) extension of the regression approach for testing multiple mediator models and the methods for estimating individual mediator effects in complex models.

4. As much as possible, authors should design their studies to meet the conditions for causal inference. Those conditions that are not met should be explicitly acknowledged, and care should be taken to use "non-causal" language when interpreting the results of mediation tests, as suggested by James and Brett (1984) and discussed earlier. We acknowledge that the very definition of mediation implies causality, and some statisticians argue that conditions for causality are required for valid inferences of mediation (Stone-Romero & Rosopa, 2004). We refer readers to the "Preconditions for Mediation Tests" section of Mathieu and Taylor's (pp. 2-9; 2006) article for a thoughtful treatment of this issue.

5. Complete descriptions of the conditions for mediation authors used (e.g., Baron & Kenny's (1986) four causal steps conditions) and the associated steps followed in the analyses (e.g., the three regression equations, plus the Sobel test) should be provided in a Method section, and results of analyses should be reported in one, complete mediation table. We included Table 4 as a suggested format for reporting tests of mediation when the causal steps approach is used and is supplemented by the Sobel test (1982).

INSERT TABLE 4 HERE

Conclusion

The accumulation of results for selected mediators has great potential for advancing knowledge in organization studies. For example, many of the individual-level mediation studies in JAP, OBHDP, AMJ and PPsych include mediators such as self-efficacy, emotional reactions and self-set goals, and group-efficacy and group affect are tested as mediators in many studies of group processes. Establishing the validity and magnitude of mediation effects would contribute greatly to our understanding of a common set of explanatory mechanisms that account for the effects of a range of individual and situational factors on performance, mental health, ethical behavior and other outcomes of interest.

Our review and critique of mediation testing in five leading journals in the past 25 years leads us to question the validity of many of the inferences of mediation effects reported . We hope that our review and the recommendations derived from it will signal to researchers the importance of correctly applying mediation approaches and tests and facilitate the valid testing of mediation models and reporting of mediation results in future studies.

References

Aroian, L. A. (1944). The probability function of the product of two normally distributed variables. Annals of Mathematical Statistics, 18, 265-271.

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.

Bentler, P. M. (1980). Multivariate analysis with latent variables: Causal modeling. Annual Review of Psychology, 31, 419-456.

Bollen, K.A. (1987). Total direct and indirect effects in structural equation models. In C.C. Clogg (Ed.), Sociological Methodology (pp. 37-69). Washington D.C.: American Psychological Association.

Bollen, K.A. & Stine, R. (1990). Direct and indirect effects: Classical and bootstrap estimates of variability. Sociological Methodology, 20, 115-140.

Bollen, K. A. (1989). Structural equations with latent variables. Oxford, England: Wiley.

Clogg, C. C., Petkova, E., & Shihadeh, E. S. (1992). Statistical methods for analyzing collapsibility in regression models. Journal of Educational Statistics, 17, 51-74.

Cohen, J. & Cohen, P. (1983). Applied multiple regression/correlational analysis for the behavioral sciences (2 ed.). Hillsdale, NJ: Erlbaum.

Cohen, J., Cohen, P., West, S.G., & Aiken, L.S. (2003). Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences (3 ed.). NJ: Erlbaum.

Goodman, L. A. (1960). On the exact variance of products. Journal of the American Statistical Association, 55, 708-713.

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

James, L. R., Mulaik, S. A., & Brett, J. M. (2006). A tale of two methods. Organizational Research Methods, 9, 233-244.

James, L. R., Mulaik, S. A., & Brett, J. M. (1982). Causal Analysis: Assumptions, models, and data. Beverly Hills, CA: Sage.

Judd, C. M. & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. Evaluation Review, 5, 602-619.

Jöreskog, K. G. & Sörbom, D. (1993). LISREL 8: Structural equation modeling with the SIMPLIS command language. Hillsdale, NJ, England: Erlbaum.

Kenny, D.A., Kashy, D.A., & Bolger, N. (1998). Data analysis in social psychology, (pp. 233-265). In D.T. Gilbert, S.T. Fiske, & G. Lindzey (Eds.) Handbook of Social Psychology, (4th Ed). New York: McGraw Hill.

MacKinnon, D.A. (2000). Contrasts in multiple mediator models. In J.S. Rose, L. Chassin, C.C. Presson, & S.J. Sherman (Eds.), Multivariate Applications in Substance Use Research: New Methods for New Questions (pp. 141-160). Mahwah, N.J.: Erlbaum.

MacKinnon, D.A. & Lockwood, C.M. (2001). Distribution of products test for the mediation effect. Unpublished manuscript.

MacKinnon, D.A., Lockwood, C., & Hoffman, J. (1998, June). A new method to test for mediation. Paper presented at the annual meeting of the Society for Prevention Research, Park City, UT.

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

MacKinnon, D.A., Warsi, G., & Dwyer, J.H. (1995). A simulation study of mediated effect measures. Multivariate Behavioral Research, 30, 41-62.

Mathieu, J.E. & Taylor, S.R. (2006). Clarifying conditions and decision points for mediational type inferences in organizational behavior. Journal of Organizational Behavior, 27, 1-26.

Olkin, I. & Finn, J.D. (1995). Correlation redux. Psychological Bulletin, 118, 155-164.

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.

Shaver, J.M. (2005). Testing for mediating variable in management research: Concerns, implications, and alternative strategies. Journal of Management, 31, 330-353.

Shrout, P.E. & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. Psychological Methods, 7, 422-445.

Sobel, M.E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), Sociological methodology (pp. 290-312). Washington, DC: American Sociological Association.

Spector, P.E. (2006). Method Variance in Organizational Research: Truth or Urban Legend? Organizational Research Methods, 9, 221-232.

Stone-Romero, E.F. & Rosopa, P. (2004). Inference problems with hierarchical multiple regression-based tests of mediating effects. In J. Martocchio (Ed.), Research in Personnel and Human Resources Management (Vol. 23, pp. 249-290). Greenwich, CT: Elsevier.

Table 1

Mediation Approaches, Numbers of Citations and Applications of Different Approaches, and Common Problems Observed in the Applications of the Approaches in the Data Set

Authors	Description	No of citations	No of applications	Problems Observed in Data Set
	Causal steps			
Baron & Kenny (1986, p. 1176)	Conditions for mediation: 1) "Variations in the levels of the independent variable significantly account for variations in the presumed mediator" (path a). 2) "Variations in the mediator significantly account for variations in the dependent variable" (path b). 3) A significant relationship between the independent and dependent variable. This condition is not separately listed by Baron and Kenny (1986), but it is implied in their statement of the following condition. 4) "When Paths a and b are controlled, a previously significant relation between the independent and dependent variables is no longer significant, with the strongest demonstration of mediation occurring when Path c is zero." This is consistent with full mediation, whereas a reduction in Path c is consistent with partial mediation. Recommended analyses: Three multiple regression models are estimated to test the conditions above. If multiple indicators of each construct are available, mediation paths can be estimated by latent variable structural equation modeling (SEM). In addition, the Sobel's (1982) test is suggested as a significance test for the indirect effect. Regression equations: 1) Regression of the mediator on the independent variable	212 (52 %)	207 (51 %)	Assuming independence of variables in multivariate models Ignoring or not providing indictors of significance of the regression models, not commenting on very small and/or non-significant models (e.g., $R^2 = .02$) Combining both mediation and moderation tests (e.g. focus on interaction terms as predictors in regression models) and ignoring special issues with this approach. Interpreting a lack of change in the coefficient as partial mediation Not testing for significance of the mediation effect Interpreting partial mediation as direct and indirect effects of the IV on the DV rather than as effects of unmeasured

	(equation 1). The independent variable must significantly affect the mediator. 2) Regression of the dependent variable on the independent variable (equation 2). The independent variable must significantly affect the dependent variable. 3) Regression of the dependent variable on both the independent variable and the mediator (equation 3). The mediator must significantly affect the dependent variable.			variables or even measurement error Applying multiple (e.g., up to 14) regression analyses (consisting of 3 regression equations each), and not adjusting for family-wise Type 1 error
Kenny, Kashy & Bolger (1998)	Conditions for mediation: Same as Baron & Kenny (1986), but relationship between independent and dependent variable is implied if path a and path b are significant. Recommended analyses: When a mediation model includes latent constructs, the estimation of structural equation models is suggested. If the mediation model only involves measured variables, multiple regression models should be estimated.	17 (4%)	10 (2%)	Same observed problems as listed above for Baron & Kenny (1986)
James & Brett (1984)	Conditions for mediation: 1) The independent variable has a direct effect on the mediator. 2) The mediator has a direct effect on the dependent variable. 3) The independent variable is not related to the dependent variable when the mediator is held constant (complete mediation). Statistically, inclusion of the independent variable adds nothing to the prediction of the dependent variable over that already explained by the mediator ($R^2_{y.mx}$ is not significantly greater than $R^2_{y.m}$). 4) The inclusion of the mediator in the model serves to enhance the explanatory power of the model, because the mediator explains how the independent variable is related to or influences the dependent variable. Statistically, the mediator adds uniquely to the prediction of the dependent variable in relation to the independent variable ($R^2_{y.mx}$ is significantly greater than $R^2_{y.x}$). It is not clear whether James and Brett (1984) intended to propose that this condition be central for establishing mediation, since they discuss it in a section on	47 (12 %)	41 (10 %)	Interpreting significant R^2 increment due to the addition of the proposed mediator as an indicator of mediation rather than as an additive, direct effect

	specification errors. However, it has been interpreted as a basic rule for establishing mediation by authors of studies in our sample. Recommended analyses: In the presence of a serious unmeasured variables problem hierarchical OLS (based on covariation) should be employed as an exploratory test of mediation. If no major misspecification of the model is likely then confirmatory analytic techniques (SEM, path analysis) should be used.			
Cohen & Cohen (1983, p. 366)	Intervening variable effects are inferred when the independent variable is significantly related to the intervening variable and the intervening variable is significantly related to the dependent variable; that is, when separate tests of path a and path b are jointly significant. Cohen et al., (2003, p 79) take a slightly different position: Mediation is described as a situation in which the partial coefficient for X predicting Y (when controlling for M) approaches zero, indicating no direct effect of X on Y and an indirect effect that takes place entirely via M.	41 (10 %)	25 (6 %)	Focusing on the bivariate relationships between X, M and Y and not considering the X to Y relationship while controlling for M
e.g., Bentler (1980) Bollen (1989) Jöreskog & Sörbom (1993) James et al.,(2006)	Structural Equation Modeling: Estimation of the indirect, direct vs. total effect, Empirical estimation of the model fit, and Comparison to alternative models SEM analytic method recommended by James et al. (2006): If complete mediation is hypothesized, use SEM to test the following equations: $m = b_{mx}x + e$; $y = b_{ym}m + e$; $\hat{r}_{xy} = b_{mx}\, b_{ym}$. Note that the latter equation for the indirect effect differs from the Baron and Kenny (1986) approach, where $b_{mx}\, b_{ym.x}$ is used. Results are consistent with full mediation when b_{mx} and b_{ym} are significant and $\hat{r}_{xy}$ is not significantly different from the observed r_{xy}. If partial mediation is hypothesized, use SEM to test the following	102 (25%)	102 (25 %)	Testing direct and indirect effects in one model and ignoring that links represent partial effects, e.g. MY.x and XY.m Not testing the mediation model against alternative models (e.g., a model with only direct effects) Estimating SEMs with less than 3 manifest variables per one latent variable Not testing for significance of indirect effects (e.g., Sobel test is included in SEM packages)

	equations: $m = b_{mx}x + e$; $y = b_{yx.m}x + b_{ym.x}m + e$. Results are consistent with partial mediation when all three parameters are statistically significant. Also recommend testing alternative causal models.		
	<u>Difference in coefficients</u>		
Clogg et al. (1992)	The difference between the coefficients when regressing the dependent variable on the independent variable before and after adjusting for the mediator variable is divided by its standard error and compared to the <u>t</u> distribution for a test of significance. $t_{N-3} = \frac{\tau - \tau'}{\lvert \rho_{XI} \sigma_{\tau'} \rvert}$	2 (0.5%)	2 (0.5 %)
Olkin & Finn (1995)	The difference between a simple correlation and the same correlation partialed for the mediator variable is divided by the standard error and compared to the standard normal distribution for a test of significance. $z = \frac{\rho_{XY} - \rho_{XY.I}}{\sigma_{\text{Olkin \& Finn}}}$	1 (0.3%)	1 (0.3 %)
	<u>Product of coefficients</u>		
Sobel (1982)	The estimate of the mediation effect (product of both path coefficients a and b: the independent variable to the mediator variable, path a, the mediator variable to the dependent variable, path b) is divided by its standard error and compared to the standard normal distribution to test for significance. $z = \frac{\alpha\beta}{\sqrt{\alpha^2\sigma_\beta^2 + \beta^2\sigma_\alpha^2}}$	28 (7%)	28 (7 %)
Goodman (1960)	The estimate of the mediation effect (product of both path coefficients a and b) is divided by its standard error and compared to the standard normal distribution to test for significance. $z = \frac{\alpha\beta}{\sqrt{\alpha^2\sigma_\beta^2 + \beta^2\sigma_\alpha^2 - \sigma_\alpha^2\sigma_\beta^2}}$	1 (0.2%)	1 (0.2%)

MacKinnon & Lockwood (2001) MacKinnon et al. (1998)	The estimate of the mediation effect (product of both path coefficients a and b) is divided by its standard error and compared to the theoretical distribution of two normal random variables. $z' = \frac{\alpha\beta}{\sqrt{\alpha^2\sigma_\beta^2 + \beta^2\sigma_\alpha^2}}$ The estimate of the mediation effect (product of both path coefficients a and b) is divided by its standard error and compared to the non-normal distribution of the product of random variables (asymmetric confidence limits are provided). $\alpha\beta \pm CL\sqrt{\alpha^2\sigma_\beta^2 + \beta^2\sigma_\alpha^2}$	11 (3 %)	2 (0.5 %)

Note: N = 409

Table 2

Numbers and percentages of studies that used different statistical methods to test for mediation by the sources cited in support of the methods used.

		Statistical Method				
Source	Number of studies that refer to the source[1]	Correlation	Partial correlation	Regression	ANCOVA	SEM
Baron & Kenny (1986)	207 (51%)	46 (11%)	6 (2%)	170 (42%)	19 (5%)	16 (4%)
James & Brett (1984)	41 (10%)	11 (3%)	1 (0.2%)	30 (7%)	3 (1%)	9 (2%)
Cohen & Cohen[2]	25 (6%)	7 (2%)	1 (0.2%)	21 (5%)	1 (0.2%)	4 (1%)
Product of coefficients	28 (7%)	5 (1%)	1 (0.2%)	21 (5%)	3 (1%)	5 (1%)
Difference in coefficients	3 (1%)	1 (0.2%)	1 (0.2%)	3 (1%)	0	0
No source cited	94 (23%)	10 (2%)	5 (1%)	29 (7%)	10 (2%)	39 (10%)
Other[3]	56 (14%)	12 (3%)	3 (1%)	20 (5%)	0	33 (8%)
Total	409 (100%)	76 (19%)	16 (4%)	257 (63%)	34 (8%)	102 (25%)

Notes:

1. Total in the first column sums to more than 409 due to multiple statistical methods in single studies. For example, authors might have cited Baron and Kenny (1986) and employed correlation as well as regression to test for mediation. Also, multiple sources were cited in support of a single statistical method in some articles.
2. Includes citations of all editions of the book; Cohen & Cohen (1983) and Cohen, et al. (2003)
3. "Other" sources cited more than once are: James, Mulaik, & Brett (1982), twice; Kenny et al. (1998), 9 times. Entries included in the "Other" row represent authoritative sources that were cited once in our sample of studies.

Table 3

Numbers and percentages of studies, models and effects reporting different test procedures by claims of mediation across all effects[1]

Mediation Frameworks and Testing Procedures	No. of studies	No. of models	No. of effects	Types of inferences for effects: Full mediation	Partial mediation	No mediation
1. Product of coefficients	28 (7%)	60 (8%)	63 (8%)	37 (9%)	15 (9%)	11 (6%)
2. Differences in coefficients	2 (0.5%)	6 (1%)	6 (1%)	4 (1%)	1 (1%)	1 (0.5%)
3. Causal steps (total)[2]	379 (93%)	643 (91%)	718 (91%)	381 (90%)	157 (91%)	180 (94%)
3.1 SEM	97 (24%)	129 (18%)	144 (18%)	101 (24%)	23 (13%)	20 (10%)
3.2 Regression or equivalent method (total)[3]	282 (69%)	514 (72%)	574 (73%)	280 (66%)	134 (77%)	160 (83%)
3.2.1 Complete set of conditions tested	88 (22%)	172 (24%)	199 (25%)	113 (27%)	43 (25%)	43 (22%)
3.2.2 Incomplete sets of conditions tested[4]	194 (47%)	342 (48%)	375 (48%)	167 (40%)	91 (53%)	117 (61%)
3.2.2.1 Incomplete sets that exclude XY	29 (7%)	63 (9%)	64 (8%)	31 (7%)	15 (9%)	18 (9%)
3.2.2.2 Other incomplete sets	165 (40%)	279 (39%)	311 (40%)	136 (32%)	76 (44%)	99 (52%)
Total	409 (100%)	709 (100%)	787 (100%)	422 (100%)	173 (100%)	192 (100%)

Notes:
1. All percentages are based on the total for the corresponding column.
2. Sum of SEM and regression or equivalent method used
3. Sum of complete sets of tests and incomplete sets of tests
4. Sum of incomplete sets without XY and "other incomplete sets"

Table 4

Recommended reporting format for presenting causal steps mediation results, using regression and the Sobel (1982) test as example analyses

Predictors	b (s.e.)	t	F	df	ΔR^2	Total R^2	Sobel Z
Model 1							
X	b_{yx} (s.e.)	t_{byx}	$F_{model\ 1}$	(_, _)	$R^2_{y.x}$	$R^2_{y.x}$	
Model 2							
M	$b_{ym.x}$ (s.e.)	$t_{bym.x}$			$R^2_{y.m}$		
X	$b_{yx.m}$ (s.e.)	$t_{byx.m}$					
			$F_{model\ 2}$	(_, _)	$R^2_{y.xm-y.x}$	$R^2_{y.xm}$	Z

Note: N = ___. * p <.05; ** p <.01; *** p <.001

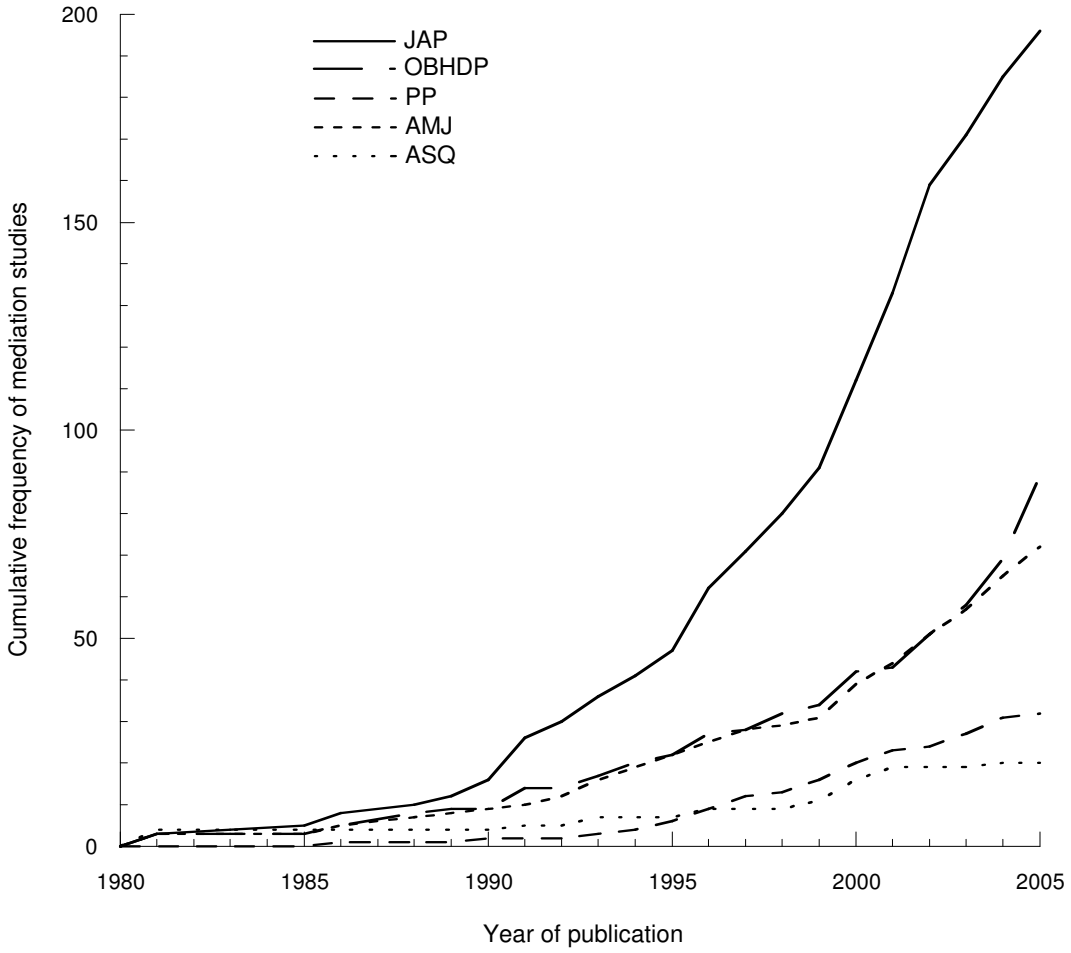

Figure 1a. Cumulative frequencies of mediation studies published in five journals by year of publication for the period 1981 to 2005.

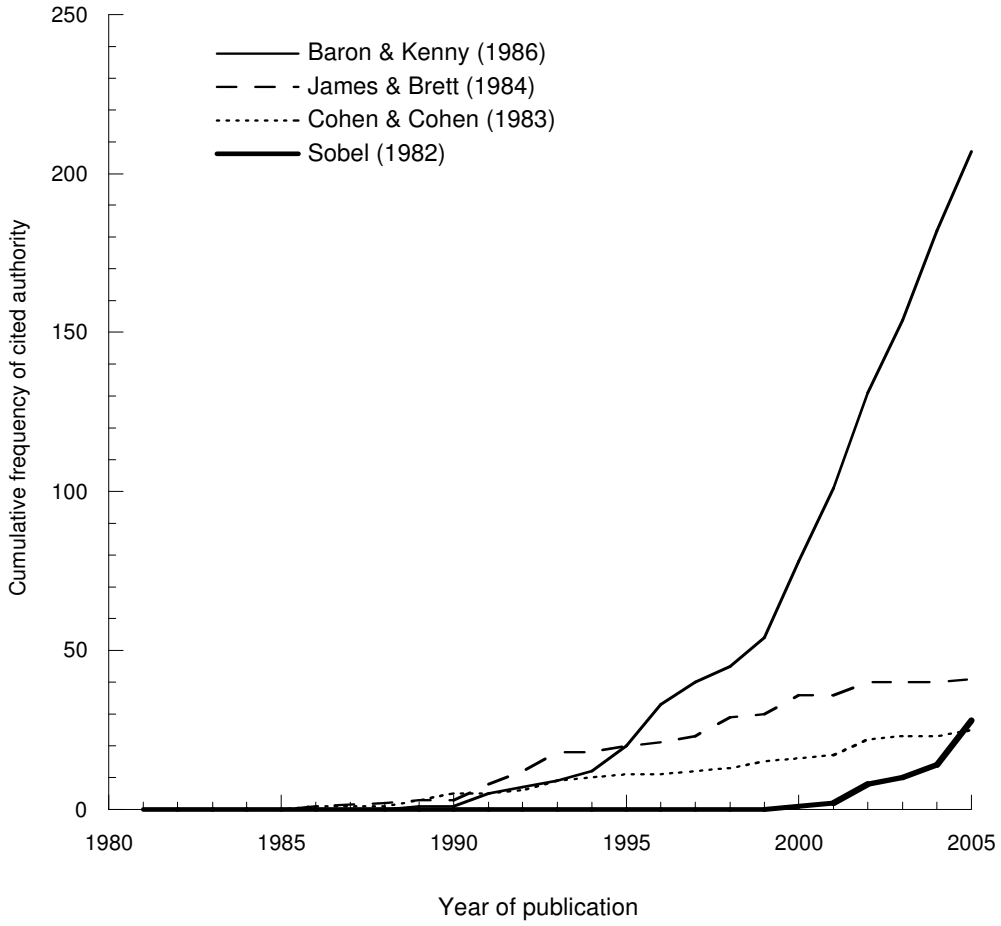

Figure 1b. Cumulative frequencies of citations for the four most commonly cited sources of methods for testing for mediation by year of publication for the period 1981 to 2005.