Research design options for testing mediation models and their implications for facets of validity

Article *in* Journal of Managerial Psychology · September 2010

DOI: 10.1108/02683941011075256

2 authors, including:

Patrick Rosopa
Clemson University

Research design options for testing mediation models and their implications for facets of validity

Eugene F. Stone-Romero
University of Texas at San Antonio, San Antonio, Texas, USA, and
Patrick J. Rosopa
Clemson University, Clemson, South Carolina, USA

Received October 2009
Revised January 2010
Accepted February 2010

Abstract

Purpose – Tests of assumed mediation models are common in research in many disciplines, including managerial psychology, industrial and organizational psychology, organizational behavior, and organizational theory. Thus, the purpose of this paper is to detail experimental design options for conducting such tests in a manner that has the potential to yield results that have high levels of internal and construct validity.

Design/methodology/approach – The paper presents a logical analysis of strategies for testing mediation models so as to insure valid inferences about causal relations between variables.

Findings – The most appropriate strategy for testing assumed mediation models is research that uses randomized experimental designs.

Practical implications – Managers should base their actions on valid evidence about phenomena. More specifically, managerial actions should be predicated on research results that have high levels of internal, construct, and statistical conclusion validity. Thus, this paper encourages managers to base decisions about organizational policies and practices on well-designed experimental research.

Originality/value – This paper addresses a number of points about issues involving internal and construct validity in tests of assumed causal models that have not been covered in previous work.

Keywords Experimental design, Business policy

Paper type Technical paper

Validity issues in tests of mediation models

Tests of mediation models are very common in managerial psychology (Wood *et al.*, 2008) and a host of other fields, including clinical psychology, marketing, personality psychology, social psychology, and sociology. Mediation models are always concerned with causal relations between variables. In complete mediation, the independent variable (X) causes the mediator (M) and M, in turn, causes the dependent variable (Y). In partial mediation, there are both direct and indirect effects of X on Y. More specifically, there is:

- a direct effect of X on Y, i.e. $X \rightarrow Y$; and
- an indirect effect of X on Y via the mediator, i.e. $X \rightarrow M \rightarrow Y$.

Simple complete and partial mediation models are shown in Figure 1(a) and (b), respectively.

Journal of Managerial Psychology
Vol. 25 No. 7, 2010
pp. 697-712
© Emerald Group Publishing Limited
0268-3946
DOI 10.1108/02683941011075256

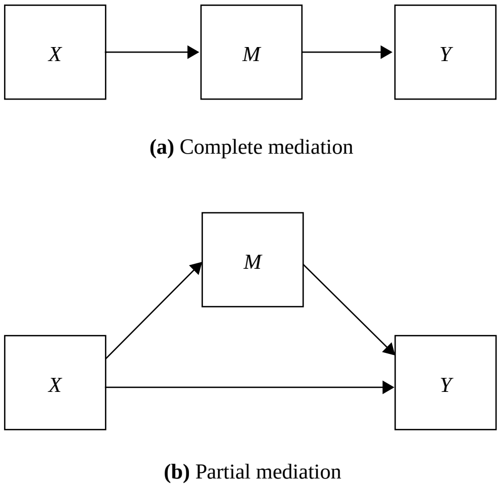

Figure 1.
Complete and partial mediation models

Purposes of paper

In view of the importance of mediation models, this paper has several major purposes, including:

- explaining several validity facets in the context of tests of mediation models;
- describing experimental design options (i.e. nonexperimental, quasi-experimental, and randomized experimental) for testing such models;
- providing explanations of several randomized experimental strategies for testing mediation models;
- considering construct validity issues in tests of mediation models;
- describing how commonly used statistical tests can be conducted using data from experimental tests of mediation models;
- offering some conclusions and recommendations associated with tests of assumed mediation models; and
- specifying implications for theory, research, and practice.

Our paper adds to the literature on mediation models by dealing with several issues not considered in the extant literature. One is that we deal explicitly with a concern about construct validity in experimental tests of mediation models that was expressed by Kenny (2008). Another is that we explain how data from experimental tests of mediation models can be used to test for complete versus partial mediation. It is worth noting that the primary audience for this paper is researchers because they often serve in consultative roles, translating research to better advise practicing managers.

Validity facets in tests of mediation models

Several types of validity are important in tests of mediation models, i.e. internal, construct, statistical conclusion, and external (Mathieu *et al.*, 2008; Rosopa and Stone-Romero, 2008; Shadish *et al.*, 2002; Stone-Romero and Rosopa, 2004). We describe them below.

Internal validity

An empirical study of mediation has internal validity to the degree that it yields correct inferences about causal relations between the independent, mediator, and dependent variables (Mathieu *et al.*, 2008; Rosopa and Stone-Romero, 2008; Shadish *et al.*, 2002; Stone-Romero and Rosopa, 2004, 2008). Thus, in mediation research, it is vital to show, for example, that $X \rightarrow M \rightarrow Y$. Doing so, hinges on the ability to satisfy the conditions vital to inferring causality. They are temporal precedence, statistical covariation, and ruling out of alternative causes (Bollen, 1989; Mathieu *et al.*, 2008; Rosopa and Stone-Romero, 2008; Shadish *et al.*, 2002; Stone-Romero and Rosopa, 2004, 2008). Thus, internal validity inferences are most justified when they are based randomized experimental studies and least warranted when predicated on nonexperimental research (Bollen, 1989; Mathieu *et al.*, 2008; Rosopa and Stone-Romero, 2008; Shadish *et al.*, 2002; Stone-Romero and Rosopa, 2004, 2008).

Construct validity

The second type of inference that is critical in tests of mediation models is construct validity, i.e. the correctness of the inference that an empirical realization (operational definition) of a construct mirrors the relevant construct (Shadish *et al.*, 2002). Construct validity is vital in any test of an assumed mediation model, whatever its experimental design (Campbell, 1976; Guion, 1976, 2002; Nunnally and Bernstein, 1998; Shadish *et al.*, 2002; Stone-Romero, 1994, 2009, 2010).

Statistical conclusion validity

The third type of inference that is important in tests of mediation models concerns statistical conclusion validity, i.e. the correctness of inferences stemming from statistical tests (Shadish *et al.*, 2002). Several data-analytic strategies for testing mediating effects (e.g. the hierarchical multiple regression (HMR) approach advocated by Baron and Kenny (1986)) have suspect statistical conclusion validity (Rosopa and Stone-Romero, 2008; Stone-Romero and Rosopa, 2004, 2008).

External validity

External validity exists to the degree that a causal relation (e.g. $X \rightarrow M$) found in one study generalizes to studies that use different samples of units, treatments, settings, and outcome measures (Mook, 1983; Shadish *et al.*, 2002). It only is an issue in randomized experimental studies if one is concerned with generalizing a mediation effect across combinations of units (e.g. type of subjects), treatments (i.e. manipulations), observations (i.e. measures), and settings (Cook and Campbell, 1979; Mook, 1983; Shadish *et al.*, 2002).

Research settings

Research aimed at testing mediation models can be conducted in special purpose and/or non special purpose settings (Stone-Romero, 2002, 2009, 2010). A special purpose setting is one that was developed for the purpose of conducting a study, and ceases to exist when it has been completed. In contrast, a non special purpose setting is one that was created for a purpose other than research (e.g. manufacturing automobiles and providing health care).

It is important to recognize that research settings are independent of the experimental designs that are used in testing mediation models (Stone-Romero, 2002, 2009, 2010). Thus, tests of mediation models can be conducted in any setting experimental design. For example, Eden and his colleagues (Dvir *et al.*, 2002; Eden, 1985; Eden and Aviram, 1993; Eden and Kinnar, 1991; Eden and Moriah, 1996) have performed numerous randomized experiments in non special purpose settings.

Experimental design alternatives for tests of mediation models

Numerous design alternatives exist for conducting tests of mediation models, i.e. nonexperimental studies, quasi-experimental studies, randomized experimental studies.

Each alternative has advantages and disadvantages that are considered below. Ideally, the results of studies using all such alternatives should converge. This provides for the triangulation of findings concerning mediation. In addition, convergent evidence strengthens external validity inferences. It is most likely to be produced by tests of mediation models that are based on sound theories.

Nonexperimental designs

In nonexperimental tests of mediation models, the assumed independent, mediator, and dependent variables are all measured. Such tests are very common in research. Note, however, that the results of nonexperimental tests do not provide a sound basis for inferring that the causal model posited by the researchers is superior to alternative models (Bollen, 1989; Gelman and Hill, 2007; Mathieu *et al.*, 2008; Rogosa, 1987; Rosopa and Stone-Romero, 2008; Shadish *et al.*, 2002; Stone-Romero and Rosopa, 2004, 2008).

Relative to other designs, nonexperimental tests of mediation have a number of advantages. One is that they are fairly easy to conduct in non special purpose settings. However, for the reasons considered above, the results of such tests afford a very weak basis for inferring that relations between measured variables are actually causal (Bollen, 1989; Gelman and Hill, 2007; Mathieu *et al.*, 2008; Rogosa, 1987; Rosopa and Stone-Romero, 2008; Shadish *et al.*, 2002; Stone-Romero and Rosopa, 2004, 2008).

Quasi-experimental designs

Tests of assumed mediation models also may be tested using quasi-experimental designs (Stone-Romero and Rosopa, 2008). In such tests, the independent variable is manipulated and its effects on several measured outcomes are assessed (Hackman *et al.*, 1978). However, because subjects are not randomly assigned to treatment conditions, causal inferences are less justified than they are in research that uses randomized experimental designs. However, causal inferences are more justified in quasi-experimental than in nonexperimental research (Cook and Campbell, 1979; Shadish *et al.*, 2002; Stone-Romero, 2009, 2010; Stone-Romero and Rosopa, 2008).

Quasi-experimental tests of mediation models can be of considerable value in basic and applied research. They are typically much easier to conduct in non special purpose settings (e.g. work organizations) because of the possibility of taking advantage of naturally occurring changes in organizations (Cook and Campbell, 1979; Grant and Wall, 2009; Hackman *et al.*, 1978; Shadish *et al.*, 2002).

Randomized experimental designs

Mediation models can be tested using randomized experimental designs (Word *et al.*, 1974). Stone-Romero and Rosopa (2008) explained how two randomized experimental studies can be used for this purpose. Table I illustrates this option. In the first experiment, the inference that the independent variable (X) produces changes in the measured mediator (M_{obs}) is based on a randomized experiment in which X is manipulated to assess its effects on M_{obs} and Y. It is important to add that because both M_{obs} and Y are measured in Experiment 1, it is possible to infer that X affected both of these outcomes. In addition, as is detailed below, it also is possible to use the data from these experiments to derive estimates of covariances that can be used to test models of assumed complete and partial mediation.

In the second experiment, the inference that the manipulated mediator (M_{man}) causes Y stems from a randomized experiment in which the mediator is manipulated and its effect on the dependent variable is measured. Note that:

- R is used in Table I to denote the random assignment of subjects to conditions;
- different subscripts are used for M in the first and second experiments (i.e., respectively, M_{obs}, measured mediator variable, and M_{man}, manipulated mediator variable); and
- the subscripts of E and C are used to denote experimental (treatment) and control conditions, respectively.

It is important to recognize that experiments need not have a no-treatment control group (as shown in Table I). Instead, they may have two or more levels of a manipulated variable.

Assuming that both of the above-described experiments produce hypothesis consistent findings, it is possible to infer the existence of mediation (Stone-Romero and Rosopa, 2008). In addition, because both the mediator and the dependent variable were measured in Experiment 1, it is possible to rule out the argument that the two randomized experiment strategy "broke" the causal chain between the independent variable and the dependent variable. Moreover, the fact that the mediator was measured before the dependent variable in Experiment 1, serves as a powerful basis for arguing that the strength of the relation between X and M_{obs} was not affected by the measurement of Y.

Experiment 1				Experiment 2		
(a) Two randomized experiments (option 1)						
R	X_C	M_{obs}	Y	R	M_{manC}	Y
R	X_E	M_{obs}	Y	R	M_{manE}	Y
(b) Two randomized experiments (option 2)						
R	X_C	M_{obs}	Y	R	M_{manC}	Y
R	X_E	M_{obs}	Y	R	M_{manE}	Y
R	X_E		Y			

Notes: R, random assignment; X_C, control condition; X_E, treatment condition for independent variable; M_{obs} , measured mediator variable; Y, measured dependent variable; M_{manC}, mediator variable control condition; M_{manE}, treatment condition for mediator variable

Table I. Design options for experimental tests of assumed mediation models

Internal validity inferences are strong with the randomized experimental design alternative for testing mediation models. Thus, it is the preferred strategy for testing such models when causal inferences are important. However, it is typically the most difficult strategy for testing mediation models, especially when they are tested using data collected in non special purpose settings (Cook and Campbell, 1979; Shadish *et al.*, 2002; Stone-Romero, 2009, 2010).

There are nontrivial differences in the extent to which causal inferences are warranted in research using each of the three general designs described above. Although this is well-known in the literature (Shadish *et al.*, 2002), it is important to highlight that extant research is replete with examples where inappropriate causal inferences were advanced based on data from nonexperimental studies (Stone-Romero and Gallaher, 2006; Wood *et al.*, 2008).

Randomized experimental design strategies in tests of mediation

Internally, valid inferences about mediation can be based on research that uses a number of experimental designs. We consider options involving two randomized experiments.

Two randomized experiments (option 1)

Stone-Romero and Rosopa (2008) suggested the two randomized experiment strategy. One option is depicted in Table I. In the first experiment, X is manipulated and M_{obs} and Y are measured, and in the second experiment the effects of M_{man} on Y are assessed.

A study by Davidson and Eden (2000) provides an example of the first experiment. The researchers randomly assigned leaders from the Israel Defense Forces to an experimental or control group. Leaders' expectations about their subordinates (X) were manipulated. The researchers measured such outcomes as subordinate self-efficacy (M_{obs}) and performance (Y). Among the hypotheses tested, the researchers found empirical support for their assertion that leader expectations (X) affected subordinate self-efficacy (M_{obs}) and performance (Y).

To illustrate the second experiment, we describe a study by Eden and Zuk (1995). The researchers randomly assigned cadets in the Israel Defense Forces to an experimental or control group. Using verbal persuasion and modeling (Bandura, 1986), the researchers manipulated self-efficacy (M_{man}) and measured seasickness and performance (Y). The cadets in the treatment condition reported less seasickness and received higher performance ratings by naïve instructors than those in the control condition (namely self-efficacy $\rightarrow$ performance).

The use of the two randomized experiment strategy has important advantages over other research designs when testing mediation. Randomized experiments provide a firm basis for advancing causal inferences because randomization of units to conditions equates groups probabilistically on all measured and unmeasured variables prior to treatment. Thus, alternative explanations of an effect can be ruled out. In addition, the temporal precedence requirement (Shadish *et al.*, 2002) for making inferences about cause is satisfied. That is, in Experiment 1, X is manipulated to assess its impact on M_{obs} and Y. Then, Experiment 2 tests for the effect of M_{man} on Y. Overall, assuming that treatments are implemented validly, randomized experiments allow for unambiguous inferences about cause-effect relations (e.g. $X \rightarrow M$ and $M \rightarrow Y$).

Another advantage is that by measuring both M_{obs} and Y in the first experiment, one has all of the information needed to test models that posit both direct and indirect effects of X on Y. That is, the results of the first experiment provide a basis for estimating the strength of the relations between:

- X and M_{obs};
- X and Y; and
- M_{obs} and Y.

As detailed below, these estimates allow for testing complete and partial mediation models using such techniques as multiple regression, path analysis (PA), and structural equation modeling (SEM) (Davidson and Eden, 2000).

The use of the two randomized experiment strategy, however, is not without its disadvantages. Randomized experiments can be very difficult to conduct, especially in non special purpose settings where control over confounds may be difficult. In addition, there may be instances where the manipulation of variables or random assignment of units to treatments may not be allowed for ethical, logistical, or administrative reasons.

Overall, executing randomized experiments properly can be very demanding relative to other experimental designs. Noteworthy, however, research by Eden and his colleagues (Dvir *et al.*, 2002; Eden, 1985; Eden and Aviram, 1993; Eden and Kinnar, 1991; Eden and Moriah, 1996) illustrates multiple instances in which tests of assumed causal models were tested with randomized experiments. Although executing randomized experiments can be very challenging, Shadish *et al.* (2002, pp. 269-75) delineate various conditions conducive to conducting them, especially in non special purpose settings.

Two randomized experiments (option 2)

A modification of the two randomized experiment option is shown in Table I. It adds a third group that enables the researcher to infer that $X \rightarrow Y$, independent of the measurement of M_{obs} or the manipulation of M_{man}; that is, it provides a simple test of the validity of the $X \rightarrow Y$ claim. This can prove useful in ruling out rival explanations of the relation between X and Y. One is that it is an artifactual byproduct of response consistency, i.e. subjects' responses to the measure of Y being influenced by their responses to the measure of M_{obs}.

Construct validity concerns in tests of mediation models

A potential concern in tests of mediation models is the construct validity of M. Kenny (2008) raised this issue in a recent article. He criticized the two randomized experiment strategy advocated by Stone-Romero and Rosopa (2008) on construct validity grounds, arguing that the measures M of the first experiment (M_{obs}) may not be equivalent to the manipulated M of the second (M_{man}). Thus, we consider this issue here. Prior to addressing it, however, we stress that construct validity is always an issue in empirical research. For example, in a nonexperimental study that tests the assumed mediation model of $X \rightarrow M \rightarrow Y$, all variables are measured. Thus, responses to a measure of Y could very well be affected by responses to a measure of M. For instance, if subjects sequentially completed measures of X, M, and Y, they might respond to any given measure in such a way as to appear consistent with responses to previously completed measures. Consistency would be especially likely if they were able to guess a study's hypotheses and responded in such a way as to provide support for them (Shadish *et al.*, 2002; Webb *et al.*, 1981; Weber and Cook, 1972). This would pose severe threats to the construct validity of the measures of M and Y. Note, moreover, that having subjects complete only subsets of measures in a nonexperimental study (e.g. the assumed mediator and dependent variables) does not lessen the possibility of their responses being a function of such artifacts as hypothesis guessing, consistency-driven responding, evaluation apprehension, and mono-method bias (Rosenthal and Rosnow, 1969; Shadish *et al.*, 2002; Webb *et al.*, 1981). The latter problem may be especially important in instances where questionnaire measures of assumed X, M, and Y variables are all completed in a single session.

There are numerous instances in which measuring one variable influences responses to others. With respect to attitude measurement, when a respondent completes measures, prior items/measures can:

- influence the respondent's interpretation of an item;
- prime certain beliefs making these more accessible; and
- create pressures to appear consistent or moderate (Tourangeau and Rasinski, 1988).

Consistent with this, with respect to personality measurement, Knowles (1988) found that an item's correlation with the rest of a measure increased linearly with the item's serial position in the measure. Similarly, Feldman and Lynch (1988) argued that attitudes, beliefs, and intentions can be created by measurement if these do not exist in long-term memory and that responses to earlier items are used by respondents to determine how to respond to subsequent items. They referred to this as self-generated validity. Clearly, construct validity may be quite problematic in tests of mediation models that are based on nonexperimental studies because these designs are based on the measurement of all variables in them. Nevertheless, it is important to specify how construct validity issues related to M can be addressed in experimental research. Thus, the next section considers this issue.

Construct validity in experimental tests of mediation models

A number of options exist in addressing the construct validity of M in experimental tests of mediation models. Our paper presents one of them. The general strategy involves three randomized experiments.

In Experiment 1, the researcher manipulates X and measures both M_{obs} and Y. In the simplest case, a researcher might have two levels of X. Then, to address the response consistency issue noted above, measurement order of M_{obs} and Y can be varied. That is, in one condition, M_{obs} is measured before Y (Sequence 1). In another condition, Y is measured before M_{obs} (Sequence 2). Thus, this experiment is a 2 (levels of X) $\times$ 2 (measurement order) factorial design.

A researcher can statistically compare the means of and the covariances (correlations) between variables in these two sequence-based conditions. Specifically, statistical tests for the equality of independent correlation coefficients can be conducted which compare the correlation coefficient for conditions where M_{obs} is measured before Y (Sequence 1) to the correlation coefficient for conditions where Y is measured before M_{obs} (Sequence 2). Note that differential correlations between the sequences would suggest that the order of measurement of M and Y made a difference. For example, the finding that the correlation between these measures was stronger in Sequence 1 than in Sequence 2 would suggest that the former sequence led to greater artifactual (spurious) correlation than the latter sequence.

In Experiment 2, the researcher manipulates M (via M_{man}) and assesses its effect on a measure of M (i.e. M_{obs}). For example, in the simplest case involving two levels of M_{man}, and assuming that M_{obs} is a continuous variable, a two-sample t-test would provide an assessment of the construct validity of both M_{obs} and M_{man}. In this study, M_{obs} serves as a check on the construct validity of M_{man}.

In Experiment 3, the researcher assesses the effects of M_{man} on Y. Statistical evidence of a relation between M_{man} and Y establishes that M causes Y.

It deserves noting that Experiment 2 can be embedded in Experiment 3, thus, reducing the experimental test of mediation models to two randomized experiments (Stone-Romero and Rosopa, 2008). Specifically, the researcher manipulates M (via M_{man}) and measures M_{obs} and Y. This establishes that M_{man} causes Y and concurrently assesses the construct validity of M_{obs} and M_{man}.

Illustration

The study described above by Eden and Zuk (1995) provides an illustration of construct validity of a mediator variable in experimental research. Using persuasion and vicarious experience as the treatment, the researchers manipulated self-efficacy (M_{man}) and measured performance (Y). As hypothesized, when self-efficacy was augmented, performance increased. Importantly, the researchers used measures of self-efficacy to confirm that those in the experimental condition did in fact report higher average self-efficacy scores than those in the control group. Note that these measures of self-efficacy were used in prior studies as an outcome measure (Eden and Kinnar, 1991) where an independent variable was manipulated and self-efficacy measured (M_{obs}).

Statistical issues in testing mediation models

A number of data-analytic strategies can be used in tests of assumed mediation models. One of the most common is the HMR approach advocated by Baron and Kenny (1986). Another is PA (Blalock, 1964, 1971). Yet another is SEM (Bollen, 1989). Assuming that a researcher has unstandardized or standardized covariances (i.e. correlation coefficients) among X, M, and Y, these techniques can be used to provide evidence about the consistency between a proposed mediation model and the data derived from a study. The same analyses can be used to support tentative inferences about complete and partial mediation. Note, however, that the legitimacy of inferences about actual causal connections between variables hinges on experimental design, not statistical evidence alone (Bollen, 1989; Rogosa, 1987; Rosopa and Stone-Romero, 2008; Shadish *et al.*, 2002; Stone-Romero, 2009, 2010; Stone-Romero and Rosopa, 2004, 2008); that is, statistical evidence allows for inferences about the degree to which the results of a study are consistent with an assumed mediation model. However, the same results also may be consistent with other mediation models. Thus, statistical evidence alone provides an insufficient basis for inferring that the relations revealed by a study allow for valid inferences about causal connections among variables (Bollen, 1989; Rogosa, 1987; Rosopa and Stone-Romero, 2008; Shadish *et al.*, 2002; Stone-Romero, 2009, 2010; Stone-Romero and Rosopa, 2004, 2008). As Tabachnick and Fidell (2007, p. 682) noted, "There is nothing causal, in the sense of inferring causality, about the use of SEM (and other statistical procedures). Attributing causality is a design issue, not a statistical issue".

Tests of mediation models with any of the above-noted data-analytic strategies require estimates of three covariances or correlation coefficients (i.e. ρ_{XM}, ρ_{XY}, and ρ_{MY}). For example, in terms of the very commonly used three-step HMR strategy (Baron and Kenny, 1986), regression coefficients are estimated based on three regression analyses which are used to infer either complete or partial mediation. Note, however, that even though the relevant regression coefficients may be consistent with an assumed mediation model, this evidence does not provide a valid basis for inferring that the same model is correct (MacKinnon, 2008; Stone-Romero and Rosopa, 2008). As MacKinnon (2008, p. 67) notes:

> [...] there are situations in which these coefficients may not reflect a true causal relation such as if X does not represent random assignment to conditions or the random assignment has been compromised.

He continues by saying that "Even if X represents random assignment, the [other] [...] coefficients are still potentially problematic because M is not randomly assigned but is determined or self-selected by study participants" (p. 67).

It deserves adding that the HMR strategy has been shown to have very low statistical power (MacKinnon *et al.*, 2002; Rosopa and Stone-Romero, 2008; Stone-Romero and Rosopa, 2004) and frequently results in incorrect inferences regarding mediation (Rosopa and Stone-Romero, 2008; Stone-Romero and Rosopa, 2004). Moreover, researchers (James *et al.*, 2006) have noted that the HMR approach inappropriately uses partial mediation as its baseline model when complete mediation should serve this purpose, because the baseline model is more parsimonious than the model tested by the Baron and Kenny (1986) procedure. In addition, assuming that mediation is tested with SEM, the baseline model allows for the test of goodness of fit (James *et al.*, 2006).

From the foregoing, it should be clear that all that is needed to compute the above-referenced regression coefficients (assuming an HMR analysis) or structural coefficients (assuming an SEM analysis) is the set of covariances among X, M, and Y. Note, in addition, that the formulas needed to estimate the regression or structural coefficients are "blind" to the nature of the study (nonexperimental, quasi-experimental, randomized experimental) that was used to generate the covariances. HMR or SEM can be used to derive estimates from any study that yields information on the above-noted covariances, independent of its experimental design.

In nonexperimental studies, the needed estimates are available, assuming X, M, and Y are measured. All that the researcher must do is estimate the degree of correlation among these variables. In addition, in quasi-experimental studies, these estimates also are available, assuming that X is manipulated and M and Y are measured. (The relevant statistical procedures are noted below.) Finally, in randomized experimental studies, the estimates are easy to obtain. We offer several examples.

The two randomized experiment (option 1). Consider the two randomized experiment, Option 1 shown in Table I. In Experiment 1, X is manipulated to assess its effects on M_{obs} and Y. The data from such a study can be used to produce estimates of ρ_{XM}, ρ_{XY}, and ρ_{MY}. If X is dichotomous, the estimates of ρ_{XM} and ρ_{XY} can be obtained using formulas for the point-biserial correlation coefficient (Hays, 1994). And, if X takes on three or more values, estimates of ρ_{XM} and ρ_{XY} can be derived from analysis of variance (ANOVA). For example, consider the correlation between X and M_{obs}. The effect size estimate for a fixed factor (e.g. X) is SSX/SSM$_{obs}$, where SSM$_{obs}$ (sum of squares for M_{obs}) is equivalent to SST. In addition, the ratio of SSX/SSM$_{obs}$ provides an estimate of the proportion of variance in M_{obs} that is explained by X. Thus, the square root of this ratio affords an estimate of the degree of correlation between X and M_{obs}.

Note that a similar analysis can be conducted to assess the effects of X on Y. Assuming that the effects of X on M_{obs} and Y are statistically significant, the relevant inference is that the mean values of M_{obs} and Y differ across levels of X.

In Experiment 2, the researcher can obtain an estimate of the effect size for the manipulation of M_{man} by SSM$_{man}$/SSY, where SSY is equivalent to SST. The inference here is about the means of Y across the levels of M_{man}.

The correlation coefficients derived from the just-described analyses can be used, for example, in conjunction with the above-noted formulas for estimating the standard partial regression coefficients to obtain estimates of direct and indirect mediation. Although the distribution of X is a fixed, there is no problem obtaining effect size estimates. See, for example, Neter *et al.* (1985, Section 26.3) or Hays (1994, Chapter 10).

It is worth noting that the inferences from a fixed effects model are not directly comparable to the inferences that stem from an analysis based on a random effects model (Hays, 1994, p. 529). However, assuming that the effect sizes (e.g. R^2-values) are comparable for fixed and random effects model-based analyses, the researcher can be confident that the R^2-values stemming from the fixed effects based analysis are not an artifact of the levels of X or M in the randomized experiments. In addition, if different researchers perform randomized experiments with different fixed levels of either X or M, the inferences stemming from the set of studies are equivalent to those from a study that is based on a random effects analysis (Hays, 1994, p. 529).

Two randomized experiments (Option 2). Next, we consider the two randomized experiment, Option 2 in Table I. Consistent with the arguments noted above, the data from either of these experiments afford a means of estimating all of the correlation coefficients needed to estimate relations between X, M, and Y. Thus, they allow for testing direct and indirect mediation using either the HMR or SEM.

Implications for theory, research, and practice

It is important to consider the implications of our arguments for constructing theory, conducting empirical research, and informing practice. We deal with each of these issues below.

Theory construction

Theories are critical in terms of specifying relations between constructs. For example, the theory of reasoned action argues that:

- attitudes toward a behavior, moral beliefs, subjective norms, and perceived behavioral control are the immediate antecedents of behavioral intentions (a mediator variable); and
- behavioral intentions are the proximal causes of behavior.

A great deal of experimental and nonexperimental research has provided support for this theory. Note, however, that had all of the research been of the nonexperimental variety, it would be difficult to specify the directions of causal connections between constructs.

One theorist might argue that moral beliefs cause behavioral intentions. However, another might view moral beliefs as a consequence of behavior. Thus, experimental research would be vital in terms of the objective of justifying the ordering of variables in a theory or model. Because of this, we believe that theorists should always attempt to show that their predictions are supported by evidence derived from experimental research.

Conducting empirical research

In early stages of research on a phenomenon (e.g. the relation between job satisfaction and employee turnover), nonexperimental research is common and may be of considerable value in terms of developing hypotheses about the antecedents and consequences of focal variables. However, as the findings of nonexperimental research accumulate, research needs to shift from nonexperimental to experimental designs. Little or nothing is gained from having convergent evidence from hundreds of studies on a specific relation. For example, the results of a meta-analysis by Judge *et al.* (2001) involving 312 samples and an overall n of 54,417 resulted in an average r of 0.30. Unfortunately, this finding does nothing whatsoever to justify conclusions about the causal connection between job satisfaction and job performance. Moreover, after almost a century of nonexperimental research, virtually nothing is known about the causal relation between these variables.

Guiding practice

The formulation of or changes in managerial processes and practices always should be guided by sound evidence about causal connections between variables. Thus, for example, it would be very unwise for a manager to use the accumulated evidence of a 0.30 average correlation between job satisfaction and job performance (Judge *et al.*, 2001) as a basis for organizational interventions aimed at increasing job satisfaction (e.g. through such actions as increasing pay, providing greater benefits, reducing hours of work, and decreasing performance standards). Although they all may serve to increase job satisfaction, there is no sound basis for arguing that they would have positive effects on performance. Thus, managers should base their actions on sound evidence on phenomena, i.e. research results that have high levels of internal, construct, and statistical conclusion validity (Shadish *et al.*, 2002). Unfortunately, nonexperimental studies do not allow for sound inferences about causal connections between variables. Thus, we encourage managers to base decisions about organizational policies and practices on well-designed experimental research. Noteworthy, researchers can play an important role in shaping management practices by advising practicing managers to utilize knowledge gleaned from experimental research.

Conclusions

On the basis of the foregoing, we offer several important conclusions. First, it should be clear that internal, construct, statistical conclusion, and external validity issues are of importance in most empirical studies. Thus, researchers should design studies in a manner that insures sufficiently high levels of all important validity facets.

Second, construct validity concerns are an issue in all empirical studies, whether they are of the nonexperimental, quasi-experimental, or randomized experimental variety.

Thus, we believe that Kenny's (2008) criticism of construct validity in the two randomized experimental strategy for testing mediating effects is unwarranted.

Third, the construct validity evidence derived from studies that use two randomized experiments appears more convincing than that derived from tests of mediation models based on nonexperimental research. Note, in addition, that irrespective of the experimental design used in a study construct validity must be demonstrated, as opposed to being assumed.

Fourth, nonexperimental, quasi-experimental, and randomized experimental options are available for testing assumed mediation models. They each have the relative advantages and disadvantages that must be considered in testing such models.

Fifth, we believe that it is far more important to show that there are causal relations between variables than to demonstrate that operational definitions of variables have construct validity (Campbell and Stanley, 1963; Cook and Campbell, 1979; Shadish *et al.*, 2002; Stone-Romero, 2009, 2010). As noted by Campbell and Stanley (1963), internal validity is the *sine qua non* of research.

Sixth, we showed that estimates of covariances (e.g. correlation coefficients) can be derived from studies that use all types of experimental designs. Thus, data from studies using all design types can be used to test partial and/or complete mediation models with such analytic strategies as HMR and SEM. As noted by James *et al.* (2006), the latter strategy is superior. However, tests based on data from nonexperimental research afford no sound basis for causal inferences.

Seventh, and finally, we considered the implications of our analysis for developing theory, conducting research, and guiding practice. In terms of these considerations, the results of research using randomized experiments are more useful than research that uses alternative designs.

Overall, therefore, researchers should not shy away from testing mediation models through research that uses randomized experimental designs. Indeed, research that uses them provides the firmest basis for:

- developing evidence about the validity of theoretical models; and
- the design of organizational interventions aimed at changing the levels of important outcome (dependent) variables (e.g. individual, group, and organizational performance).

References

Bandura, A. (1986), *Social Foundations of Thought and Action: A Social-cognitive View*, Prentice-Hall, Englewood Cliffs, NJ.

Baron, R.M. and Kenny, D.A. (1986), "The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations", *Journal of Personality and Social Psychology*, Vol. 51, pp. 1173-82.

Blalock, H.M. (1964), *Causal Inferences in Nonexperimental Research*, W.W. Norton, New York, NY.

Blalock, H.M. (1971), *Causal Models in the Social Sciences*, Aldine, Chicago, IL.

Bollen, K.A. (1989), *Structural Equations with Latent Variables*, Wiley, New York, NY.

Campbell, D.T. and Stanley, J.C. (1963), *Experimental and Quasi-experimental Designs for Research*, Rand McNally, Chicago, IL.

Campbell, J.P. (1976), "Psychometric theory", in Dunnette, M.D. (Ed.), *Handbook of Industrial and Organizational Psychology*, Rand McNally, Chicago, IL, pp. 185-222.

Cook, T.D. and Campbell, D.T. (1979), *Quasi-experimentation: Design and Analysis Issues for Field Settings*, Houghton Mifflin, Boston, MA.

Davidson, O.B. and Eden, D. (2000), "Remedial self-fulfilling prophecy: two field experiments to prevent Golem effects among disadvantaged women", *Journal of Applied Psychology*, Vol. 85, pp. 386-98.

Dvir, T., Eden, D., Avolio, B.J. and Shamir, B. (2002), "Impact of leadership development on follower development and performance: a field experiment", *Academy of Management Journal*, Vol. 45, pp. 735-44.

Eden, D. (1985), "Team development: a true field experiment at three levels of rigor", *Journal of Applied Psychology*, Vol. 70, pp. 94-100.

Eden, D. and Aviram, A. (1993), "Self-efficacy training to speed reemployment: helping people to help themselves", *Journal of Applied Psychology*, Vol. 78, pp. 352-60.

Eden, D. and Kinnar, J. (1991), "Modeling Galatea: boosting self-efficacy to increase volunteering", *Journal of Applied Psychology*, Vol. 76, pp. 770-80.

Eden, D. and Moriah, L. (1996), "Impact of internal auditing on branch bank performance: a field experiment", *Organizational Behavior and Human Decision Processes*, Vol. 68, pp. 262-71.

Eden, D. and Zuk, Y. (1995), "Seasickness as a self-fulfilling prophecy: raising self-efficacy to boost performance at sea", *Journal of Applied Psychology*, Vol. 80, pp. 628-35.

Feldman, J.M. and Lynch, J.G. (1988), "Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior", *Journal of Applied Psychology*, Vol. 73, pp. 421-35.

Gelman, A. and Hill, J. (2007), *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press, New York, NY.

Grant, A.M. and Wall, T.D. (2009), "The neglected science and art of quasi-experimentation: why-to, when-to, and how-to advice for organizational researchers", *Organizational Research Methods*, Vol. 12, pp. 653-86.

Guion, R.M. (1976), "Recruiting, selection, and job placement", in Dunnette, M.D. (Ed.), *Handbook of Industrial and Organizational Psychology*, Rand McNally, Chicago, IL, pp. 777-828.

Guion, R.M. (2002), "Validity and reliability", in Rogelberg, S.G. (Ed.), *Handbook of Research Methods in Industrial and Organizational Psychology*, Blackwell, Malden, MA, pp. 57-76.

Hackman, J.R., Pearce, J.L. and Wolfe, J.C. (1978), "Effects of changes in job characteristics on work attitudes and behaviors: a naturally occurring quasi-experiment", *Organizational Behavior and Human Performance*, Vol. 21, pp. 289-304.

Hays, W.L. (1994), *Statistics*, 5th ed., Harcourt Brace, Fort Worth, TX.

James, L.R., Mulaik, S.A. and Brett, J.M. (2006), "A tale of two methods", *Organizational Research Methods*, Vol. 9, pp. 233-44.

Judge, T.A., Thorensen, C.J., Bono, J.E. and Patton, G.K. (2001), "The job satisfaction-job performance relationship: a qualitative and quantitative review", *Psychological Bulletin*, Vol. 127, pp. 376-407.

Kenny, D.A. (2008), "Reflections on mediation", *Organizational Research Methods*, Vol. 11, pp. 353-8.

Knowles, E.S. (1988), "Item context effects on personality scales: measuring changes the measure", *Journal of Personality and Social Psychology*, Vol. 55, pp. 312-20.

MacKinnon, D.P. (2008), *Introduction to Statistical Mediation Analysis*, Erlbaum, New York, NY.

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

Mathieu, J.E., DeShon, R.P. and Bergh, D.D. (2008), “Mediational inferences in organizational research: then, now, and beyond”, *Organizational Research Methods*, Vol. 11, pp. 203-23.

Mook, D.G. (1983), “In defense of external invalidity”, *American Psychologist*, Vol. 38, pp. 379-87.

Neter, J., Wasserman, W. and Kutner, M.H. (1985), *Applied Linear Statistical Models*, 2nd ed., Irwin, Homewood, IL.

Nunnally, J.C. and Bernstein, I.H. (1998), *Psychometric Theory*, 3rd ed., McGraw-Hill, New York, NY.

Rogosa, D. (1987), “Causal models do not support scientific conclusions: a comment in support of Freedman”, *Journal of Educational Statistics*, Vol. 12, pp. 185-95.

Rosenthal, R. and Rosnow, R.L. (1969), *Artifacts in Behavioral Research*, Academic Press, New York, NY.

Rosopa, P.J. and Stone-Romero, E.F. (2008), “Problems with detecting assumed mediation using the hierarchical multiple regression strategy”, *Human Resource Management Review*, Vol. 18, pp. 294-310.

Shadish, W.R., Cook, T.D. and Campbell, D.T. (2002), *Experimental and Quasi-experimental Designs for Generalized Causal Inference*, Houghton Mifflin, Boston, MA.

Stone-Romero, E.F. (1994), “Construct validity issues in organizational behavior research”, in Greenberg, J. (Ed.), *Organizational Behavior: The State of the Science*, Erlbaum, Hillsdale, NJ.

Stone-Romero, E.F. (2002), “The relative validity and usefulness of various empirical research designs”, in Rogelberg, S.G. (Ed.), *Handbook of Research Methods in Industrial and Organizational Psychology*, Blackwell, Malden, MA, pp. 77-98.

Stone-Romero, E.F. (2009), “Implications of research design options for the validity of inferences derived from organizational research”, in Buchanan, D. and Bryman, A. (Eds), *Handbook of Organizational Research Methods*, Sage, London, pp. 302-27.

Stone-Romero, E.F. (2010), “Research strategies in industrial and organizational psychology: Nonexperimental, quasi-experimental, and randomized experimental research in special purpose and non-special purpose settings”, in Zedeck, S. (Ed.), *Handbook of Industrial and Organizational Psychology*, American Psychological Association Press, Washington, DC, pp. 35-70.

Stone-Romero, E.F. and Gallaher, L. (2006), “Inappropriate use of causal language in reports of nonexperimental research”, paper presented at the meeting of the Society for Industrial and Organizational Psychology, Dallas, TX, May.

Stone-Romero, E.F. and Rosopa, P.J. (2004), “Inference problems with hierarchical multiple regression-based tests of mediating effects”, *Research in Personnel and Human Resources Management*, Vol. 23, pp. 249-90.

Stone-Romero, E.F. and Rosopa, P.J. (2008), “The relative validity of inferences about mediation as a function of research design characteristics”, *Organizational Research Methods*, Vol. 11, pp. 326-52.

Tabachnick, B.G. and Fidell, L.S. (2007), *Using Multivariate Statistics*, 5th ed., Pearson, Boston, MA.

Tourangeau, R. and Rasinski, K.A. (1988), “Cognitive processes underlying context effects in attitude measurement”, *Psychological Bulletin*, Vol. 103, pp. 299-314.

Webb, E.J., Campbell, D.T., Schwartz, R.D., Sechrest, L. and Grove, J.B. (1981), *Nonreactive Measures in the Social Sciences*, 2nd ed., Houghton Mifflin, Boston, MA.

Weber, S.J. and Cook, T.D. (1972), "Subject effects in laboratory research: an examination of subject roles, demand characteristics, and valid inference", *Psychological Bulletin*, Vol. 77, pp. 273-95.

Wood, R.E., Goodman, J.S., Beckmann, N. and Cook, A. (2008), "Mediation testing in management research: a review and proposals", *Organizational Research Methods*, Vol. 11, pp. 270-95.

Word, C.O., Zanna, M.P. and Cooper, J. (1974), "The nonverbal mediation of self-fulfilling prophecies in interracial interaction", *Journal of Experimental Social Psychology*, Vol. 10, pp. 109-20.

Further reading

Cook, T.D. and Campbell, D.T. (1976), "The design and conduct of quasi-experiments and true experiments in field settings", in Dunnette, M.D. (Ed.), *Handbook of Industrial and Organizational Psychology*, Rand McNally, Chicago, IL, pp. 223-326.

Mathieu, J.E. and Taylor, S.R. (2006), "Clarifying conditions and choice points for mediational type inferences in organizational behavior", *Journal of Organizational Behavior*, Vol. 27, pp. 1031-56.

Meyer, J.P. and Gellatly, I.R. (1988), "Perceived performance norm as a mediator in the effect of assigned goal on personal goal and task performance", *Journal of Applied Psychology*, Vol. 73, pp. 410-20.

About the authors

Eugene F. Stone-Romero (PhD, University of California-Irvine) is a Professor of Management at the University of Texas at San Antonio. He is a Fellow of the Society for Industrial and Organizational Psychology, the Association for Psychological Science, and the American Psychological Association. The results of his research have appeared in such outlets as the *Journal of Applied Psychology*, *Organizational Behavior and Human Performance*, *Personnel Psychology, Journal of Vocational Behavior*, *Academy of Management Journal*, *Journal of Management*, *International Review of Industrial and Organizational Psychology*, *Research in Personnel and Human Resources Management*, *Applied Psychology: An International Review*, and the *Journal of Applied Social Psychology*. He is also the author of numerous book chapters. Finally, he is the author of a book entitled *Research Methods in Organizational Behavior*, and the co-author of books entitled *Job Satisfaction: How People Feel about Their Jobs and How It Affects Their Performance,* and *The Influence of Culture on Human Resource Management Processes and Practices.* Eugene F. Stone-Romero is the corresponding author and can be contacted at: wolfcub1@satx.rr.com

Patrick J. Rosopa is an Assistant Professor of Industrial/Organizational Psychology in the Department of Psychology at Clemson University. He earned his PhD and MS in Industrial/Organizational Psychology from the University of Central Florida and his BS in Psychology from Tulane University. He has co-authored a statistics textbook entitled *Statistical Reasoning for the Behavioral Sciences* (6th edition, 2010, Wiley, with B.M. King and E.W. Minium). His research interests include: applied statistics and research methods, personality and cognitive ability, stereotypes and fairness in the workplace, and cross-cultural issues in organizational psychology. His research has been published in such outlets as the *Encyclopedia of Measurement and Statistics*, *Human Resource Management Review*, *Journal of Vocational Behavior*, *Organizational Research Methods*, *Personality and Individual Differences*, and *Research in Personnel and Human Resource Management*.