Journal of Personality and Social Psychology
1986, Vol. 51, No. 6, 1173–1182

0022-3514/86/$00.75

The Moderator–Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations

Reuben M. Baron and David A. Kenny
University of Connecticut

In this article, we attempt to distinguish between the properties of moderator and mediator variables at a number of levels. First, we seek to make theorists and researchers aware of the importance of not using the terms *moderator* and *mediator* interchangeably by carefully elaborating, both conceptually and strategically, the many ways in which moderators and mediators differ. We then go beyond this largely pedagogical function and delineate the conceptual and strategic implications of making use of such distinctions with regard to a wide range of phenomena, including control and stress, attitudes, and personality traits. We also provide a specific compendium of analytic procedures appropriate for making the most effective use of the moderator and mediator distinction, both separately and in terms of a broader causal system that includes both moderators and mediators.

The purpose of this analysis is to distinguish between the properties of moderator and mediator variables in such a way as to clarify the different ways in which conceptual variables may account for differences in peoples' behavior. Specifically, we differentiate between two often-confused functions of third variables: (a) the moderator function of third variables, which partitions a focal independent variable into subgroups that establish its domains of maximal effectiveness in regard to a given dependent variable, and (b) the mediator function of a third variable, which represents the generative mechanism through which the focal independent variable is able to influence the dependent variable of interest.

Although these two functions of third variables have a relatively long tradition in the social sciences, it is not at all uncommon for social psychological researchers to use the terms *moderator* and *mediator* interchangeably. For example, Harkins, Latané, and Williams (1980) first summarized the impact of identifiability on social loafing by observing that it "moderates social loafing" (p. 303) and then within the same paragraph proposed "that identifiability is an important mediator of social loafing." Similarly, Findley and Cooper (1983), intending a moderator interpretation, labeled gender, age, race, and socioeconomic level as mediators of the relation between locus of control and academic achievement. Thus, one largely pedagogical function of this article is to clarify for experimental researchers the importance of respecting these distinctions.

This is not, however, the central thrust of our analysis. Rather, our major emphasis is on contrasting the moderator–mediator functions in ways that delineate the implications of this distinction for theory and research. We focus particularly on the differential implications for choice of experimental design, research operations, and plan of statistical analysis.

We also claim that there are conceptual implications of the failure to appreciate the moderator–mediator distinction. Among the issues we will discuss in this regard are missed opportunities to probe more deeply into the nature of causal mechanisms and integrate seemingly irreconcilable theoretical positions. For example, it is possible that in some problem areas disagreements about mediators can be resolved by treating certain variables as moderators.

The moderator and mediator functions will be discussed at three levels: conceptual, strategic, and statistical. To avoid any misunderstanding of the moderator–mediator distinction by erroneously equating it with the difference between experimental manipulations and measured variables, between situational and person variables, or between manipulations and verbal self-reports, we will describe both actual and hypothetical examples involving a wide range of variables and operations. That is, moderators may involve either manipulations or assessments and either situational or person variables. Moreover, mediators are in no way restricted to verbal reports or, for that matter, to individual-level variables.

Finally, for expository reasons, our analysis will initially stress the need to make clear whether one is testing a moderator or a mediator type of model. In the second half of the article, we provide a design that allows one to test within the structure of the same study whether a mediator or moderator interpretation is more appropriate.

Although these issues are obviously important for a large number of areas within psychology, we have targeted this article for a social psychological audience because the relevance of this distinction is highest in social psychology, which uses experi-

This research was supported in part by National Science Foundation Grant BNS-8210137 and National Institute of Mental Health Grant R01MH-40295-01 to the second author. Support was also given to him during his sabbatical year (1982–83) by the MacArthur Foundation at the Center for Advanced Studies in the Behavioral Sciences, Stanford, California.

Thanks are due to Judith Harackiewicz, Charles Judd, Stephen West, and Harris Cooper, who provided comments on an earlier version of this article. Stephen P. Needel was instrumental in the beginning stages of this work.

Correspondence concerning this article should be addressed to Reuben M. Baron, Department of Psychology U-20, University of Connecticut, Storrs, Connecticut 06268.

mental operations and at the same time retains an interest in organismic variables ranging from individual difference measures to cognitive constructs such as perceived control.

The Nature of Moderators

In general terms, a moderator is a qualitative (e.g., sex, race, class) or quantitative (e.g., level of reward) variable that affects the direction and/or strength of the relation between an independent or predictor variable and a dependent or criterion variable.

Specifically within a correlational analysis framework, a moderator is a third variable that affects the zero-order correlation between two other variables. For example, Stern, McCants, and Pettine (1982) found that the positivity of the relation between changing life events and severity of illness was considerably stronger for uncontrollable events (e.g., death of a spouse) than for controllable events (e.g., divorce). A moderator effect within a correlational framework may also be said to occur where the direction of the correlation changes. Such an effect would have occurred in the Stern et al. study if controllable life changes had reduced the likelihood of illness, thereby changing the direction of the relation between life-event change and illness from positive to negative.

In the more familiar analysis of variance (ANOVA) terms, a basic moderator effect can be represented as an interaction between a focal independent variable and a factor that specifies the appropriate conditions for its operation. In the dissonance–forced compliance area, for example, it became apparent that the ability of investigators to establish the effects of insufficient justification required the specification of such moderators as commitment, personal responsibility, and free choice (cf. Brehm & Cohen, 1962).

An example of a moderator-type effect in this context is the demonstration of a crossover interaction of the form that the insufficient justification effect holds under public commitment (e.g., attitude change is inversely related to incentive), whereas attitude change is directly related to level of incentive when the counterattitudinal action occurs in private (cf. Collins & Hoyt, 1972). A moderator-interaction effect also would be said to occur if a relation is substantially reduced instead of being reversed, for example, if we find no difference under the private condition.[1]

Toward Establishing an Analytic Framework for Testing Moderator Effects

A common framework for capturing both the correlational and the experimental views of a moderator variable is possible by using a path diagram as both a descriptive and an analytic procedure. Glass and Singer's (1972) finding of an interaction of the factors stressor intensity (noise level) and controllability (periodic–aperiodic noise), of the form that an adverse impact on task performance occurred only when the onset of the noise was aperiodic or unsignaled, will serve as our substantive example. Using such an approach, the essential properties of a moderator variable are summarized in Figure 1.

The model diagrammed in Figure 1 has three causal paths that feed into the outcome variable of task performance: the impact of the noise intensity as a predictor (Path *a*), the impact of controllability as a moderator (Path *b*), and the interaction or product of these two (Path *c*). The moderator hypothesis is supported if the interaction (Path *c*) is significant. There may also be significant main effects for the predictor and the moderator (Paths *a* and *b*), but these are not directly relevant conceptually to testing the moderator hypothesis.

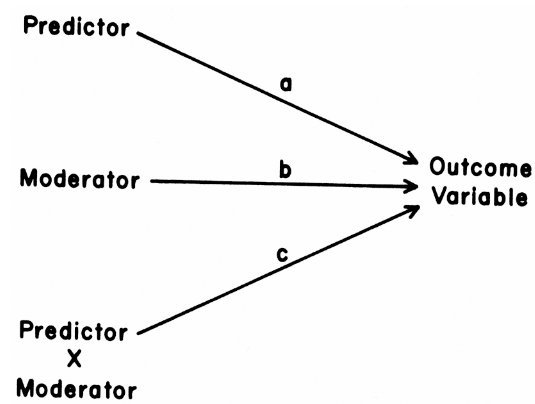


Figure 1. Moderator model.

In addition to these basic considerations, it is desirable that the moderator variable be uncorrelated with both the predictor and the criterion (the dependent variable) to provide a clearly interpretable interaction term. Another property of the moderator variable apparent from Figure 1 is that, unlike the mediator–predictor relation (where the predictor is causally antecedent to the mediator), moderators and predictors are at the same level in regard to their role as causal variables antecedent or exogenous to certain criterion effects. That is, moderator variables always function as independent variables, whereas mediating events shift roles from effects to causes, depending on the focus of the analysis.

Choosing an Appropriate Analytic Procedure: Testing Moderation

In this section we consider in detail the specific analysis procedures for appropriately measuring and testing moderational hypotheses. Within this framework, moderation implies that the causal relation between two variables changes as a function of the moderator variable. The statistical analysis must measure and test the differential effect of the independent variable on the dependent variable as a function of the moderator. The way to measure and test the differential effects depends in part on the level of measurement of the independent variable and the moderator variable. We will consider four cases: In Case 1, both moderator and independent variables are categorical variables; in Case 2, the moderator is a categorical variable and the independent variable a continuous variable; in Case 3, the modera-

[1] At a conceptual level, a moderator may be more impressive if we go from a strong to a weak relation or to no relation at all as opposed to finding a crossover interaction. That is, although crossover interactions are stronger statistically, as they are not accompanied by residual main effects, conceptually no effect shifts may be more impressive.

tor is a continuous variable and the independent variable is a categorical variable; and in Case 4, both variables are continuous variables. To ease our discussion, we will assume that all the categorical variables are dichotomies.

Case 1

This is the simplest case. For this case, a dichotomous independent variable's effect on the dependent variable varies as a function of another dichotomy. The analysis is a 2 × 2 ANOVA, and moderation is indicated by an interaction. We may wish to measure the simple effects of the independent variable across the levels of the moderator (Winer, 1971, pp. 435–436), but these should be measured only if the moderator and the independent variable interact to cause the dependent variable.

Case 2

Here the moderator is a dichotomy and the independent variable is a continuous variable. For instance, gender might moderate the effect of intentions on behavior. The typical way to measure this type of moderator effect is to correlate intentions with behavior separately for each gender and then test the difference. For instance, virtually all studies of moderators of the attitude–behavior relation use a correlational test.

The correlational method has two serious deficiencies. First, it presumes that the independent variable has equal variance at each level of the moderator. For instance, the variance of intention must be the same for the genders. If variances differ across levels of the moderator, then for levels of the moderator with less variance, the correlation of the independent variable with the dependent variable tends to be less than for levels of the moderator with more variance. The source of this difference is referred to as a restriction in range (McNemar, 1969). Second, if the amount of measurement error in the dependent variable varies as a function of the moderator, then the correlations between the independent and dependent variables will differ spuriously.

These problems illustrate that correlations are influenced by changes in variances. However, regression coefficients are not affected by differences in the variances of the independent variable or differences in measurement error in the dependent variable. It is almost always preferable to measure the effect of the independent variable on the dependent variable not by correlation coefficients but by unstandardized (not betas) regression coefficients (Duncan, 1975). Tests of the difference between regression coefficients are given in Cohen and Cohen (1983, p. 56). This test should be performed first, before the two slopes are individually tested.

If there is differential measurement error in the independent variable across levels of the moderator, bias results. Reliabilities would then need to be estimated for the different levels of the moderator, and slopes would have to be disattenuated. This can be accomplished within the computer program LISREL-VI (Jöreskog & Sörbom, 1984) by use of the multiple-group option. The levels of the moderator are treated as different groups.

Case 3

In this case, the moderator is a continuous variable and the independent variable is a dichotomy. For instance, the independent variable might be a rational versus fear-arousing attitude-change message and the moderator might be intelligence as measured by an IQ test. The fear-arousing message may be more effective for low-IQ subjects, whereas the rational message may be more effective for high-IQ subjects. To measure moderator effects in this case, we must know a priori how the effect of the independent variable varies as a function of the moderator. It is impossible to evaluate the general hypothesis that the effect of the independent variable changes as a function of the moderator because the moderator has many levels.

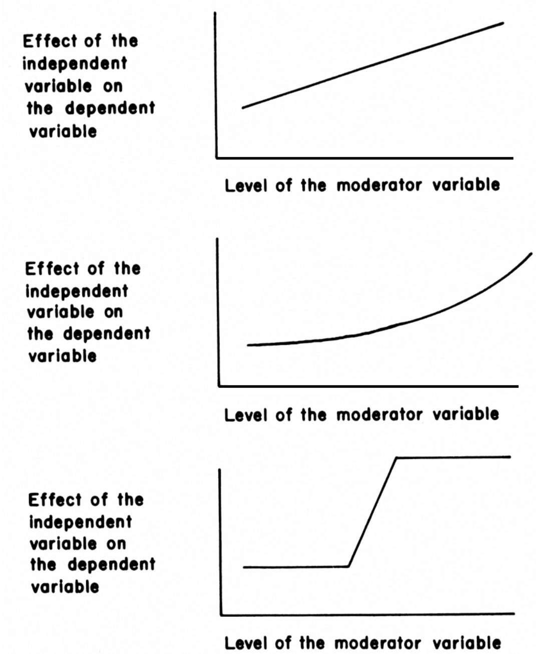


Figure 2. Three different ways in which the moderator changes the effect of the independent variable on the dependent variable: linear (top), quadratic (middle), and step (bottom).

Figure 2 presents three idealized ways in which the moderator alters the effect of the independent variable on the dependent variable. First, the effect of the independent variable on the dependent variable changes linearly with respect to the moderator. The linear hypothesis represents a gradual, steady change in the effect of the independent variable on the dependent variable as the moderator changes. It is this form of moderation that is generally assumed. The second function in the figure is a quadratic function. For instance, the fear-arousing message may be more generally effective than the rational message for all low-IQ subjects, but as IQ increases, the fear-arousing message loses its advantage and the rational message is more effective.

The third function in Figure 2 is a step function. At some critical IQ level, the rational message becomes more effective than the fear-arousing message. This pattern is tested by dichotomizing the moderator at the point where the step is supposed to occur and proceeding as in Case 1. Unfortunately, theories in social psychology are usually not precise enough to specify the exact point at which the step in the function occurs.

The linear hypothesis is tested by adding the product of the moderator and the dichotomous independent variable to the regression equasion, as described by Cohen and Cohen (1983) and Cleary and Kessler (1982). So if the independent variable is denoted as X, the moderator as Z, and the dependent variable as Y, Y is regressed on X, Z, and XZ. Moderator effects are indicated by the significant effect of XZ while X and Z are controlled. The simple effects of the independent variable for different levels of the moderator can be measured and tested by procedures described by Aiken and West (1986). (Measurement error in the moderator requires the same remedies as measurement error in the independent variable under Case 2.)

The quadratic moderation effect can be tested by dichotomizing the moderator at the point at which the function is presumed to accelerate. If the function is quadratic, as in Figure 2, the effect of the independent variable should be greatest for those who are high on the moderator. Alternatively, quadratic moderation can be tested by hierarchical regression procedures described by Cohen and Cohen (1983). Using the same notation as in the previous paragraph, Y is regressed on X, Z, XZ, Z^2, and XZ^2. The test of quadratic moderation is given by the test of XZ^2. The interpretation of this complicated regression equation can be aided by graphing or tabling the predicted values for various values of X and Z.

Case 4

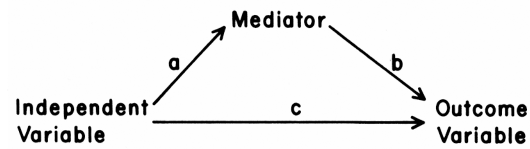
In this case both the moderator variable and the independent variable are continuous. If one believes that the moderator alters the independent–dependent variable relation in a step function (the bottom diagram in Figure 2), one can dichotomize the moderator at the point where the step takes place. After dichotomizing the moderator, the pattern becomes Case 2. The measure of the effect of the independent variable is a regression coefficient.

If one presumes that the effect of the independent variable (X) on the dependent variable (Y) varies linearly or quadratically with respect to the moderator (Z), the product variable approach described in Case 3 should be used. For quadratic moderation, the moderator squared must be introduced. One should consult Cohen and Cohen (1983) and Cleary and Kessler (1982) for assistance in setting up and interpreting these regressions.

The presence of measurement error in either the moderator or the independent variable under Case 4 greatly complicates the analysis. Busemeyer and Jones (1983) assumed that the moderation is linear and so can be captured by an XZ product term. They showed that measuring multiplicative interactions when one of the variables has measurement error results in low power in the test of interactive effects. Methods presented by Kenny and Judd (1984) can be used to make adjustments for measurement error in the variables, resulting in proper estimates of interactive effects. However, these methods require that the variables from which the product variable is formed have normal distributions.

The Nature of Mediator Variables

Although the systematic search for moderator variables is relatively recent, psychologists have long recognized the importance of mediating variables. Woodworth's (1928) S-O-R



model, which recognizes that an active organism intervenes between stimulus and response, is perhaps the most generic formulation of a mediation hypothesis. The central idea in this model is that the effects of stimuli on behavior are mediated by various transformation processes internal to the organism. Theorists as diverse as Hull, Tolman, and Lewin shared a belief in the importance of postulating entities or processes that intervene between input and output. (Skinner's blackbox approach represents the notable exception.)

General Analytic Considerations

In general, a given variable may be said to function as a mediator to the extent that it accounts for the relation between the predictor and the criterion. Mediators explain how external physical events take on internal psychological significance. Whereas moderator variables specify when certain effects will hold, mediators speak to how or why such effects occur. For example, choice may moderate the impact of incentive on attitude change induced by discrepant action, and this effect is in turn mediated by a dissonance arousal–reduction sequence (cf. Brehm & Cohen, 1962).

To clarify the meaning of mediation, we now introduce a path diagram as a model for depicting a causal chain. The basic causal chain involved in mediation is diagrammed in Figure 3. This model assumes a three-variable system such that there are two causal paths feeding into the outcome variable: the direct impact of the independent variable (Path *c*) and the impact of the mediator (Path *b*). There is also a path from the independent variable to the mediator (Path *a*).

A variable functions as a mediator when it meets the following conditions: (a) variations in levels of the independent variable significantly account for variations in the presumed mediator (i.e., Path *a*), (b) variations in the mediator significantly account for variations in the dependent variable (i.e., Path *b*), and (c) 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. In regard to the last condition we may envisage a continuum. When Path *c* is reduced to zero, we have strong evidence for a single, dominant mediator. If the residual Path *c* is not zero, this indicates the operation of multiple mediating factors. Because most areas of psychology, including social, treat phenomena that have multiple causes, a more realistic goal may be to seek mediators that significantly decrease Path *c* rather than eliminating the relation between the independent and dependent variables altogether. From a theoretical perspective, a significant reduction demonstrates that a given mediator is indeed potent, albeit not both a necessary and a sufficient condition for an effect to occur.

Testing Mediation

An ANOVA provides a limited test of a mediational hypothesis as extensively discussed in Fiske, Kenny, and Taylor (1982). Rather, as recommended by Judd and Kenny (1981b), a series of regression models should be estimated. To test for mediation, one should estimate the three following regression equations: first, regressing the mediator on the independent variable; second, regressing the dependent variable on the independent variable; and third, regressing the dependent variable on both the independent variable and on the mediator. Separate coefficients for each equation should be estimated and tested. There is no need for hierarchical or stepwise regression or the computation of any partial or semipartial correlations.

These three regression equations provide the tests of the linkages of the mediational model. To establish mediation, the following conditions must hold: First, the independent variable must affect the mediator in the first equation; second, the independent variable must be shown to affect the dependent variable in the second equation; and third, the mediator must affect the dependent variable in the third equation. If these conditions all hold in the predicted direction, then the effect of the independent variable on the dependent variable must be less in the third equation than in the second. Perfect mediation holds if the independent variable has no effect when the mediator is controlled.

Because the independent variable is assumed to cause the mediator, these two variables should be correlated. The presence of such a correlation results in multicollinearity when the effects of independent variable and mediator on the dependent variable are estimated. This results in reduced power in the test of the coefficients in the third equation. It is then critical that the investigator examine not only the significance of the coefficients but also their absolute size. For instance, it is possible for the independent variable to have a smaller coefficient when it alone predicts the dependent variable than when it and the mediator are in the equation but the larger coefficient is not significant and the smaller one is.

Sobel (1982) provided an approximate significance test for the indirect effect of the independent variable on the dependent variable via the mediator. As in Figure 3, the path from the independent variable to the mediator is denoted as a and its standard error is s_a; the path from the mediator to the dependent variable is denoted as b and its standard error is s_b. The exact formula, given multivariate normality for the standard error of the indirect effect or ab, is this:

$$\sqrt{b^2 s_a^2 + a^2 s_b^2 + s_a^2 s_b^2}$$

Sobel's method omits the term $s_a^2 s_b^2$, but that term ordinarily is small. His approximate method can be used for more complicated models.

The use of multiple regression to estimate a mediational model requires the two following assumptions: that there be no measurement error in the mediator and that the dependent variable not cause the mediator.

The mediator, because it is often an internal, psychological variable, is likely to be measured with error. The presence of measurement error in the mediator tends to produce an underestimate of the effect of the mediator and an overestimate of the effect of the independent variable on the dependent variable when all coefficients are positive (Judd & Kenny, 1981a). Obviously this is not a desirable outcome, because successful mediators may be overlooked.

Generally the effect of measurement error is to attenuate the size of measures of association, the resulting estimate being closer to zero than it would be if there were no measurement error (Judd & Kenny, 1981a). Additionally, measurement error in the mediator is likely to result in an overestimate in the effect of the independent variable on the dependent variable. Because of measurement error in the mediator, effects of the mediator on the dependent variable cannot totally be controlled for when measuring the effects of the independent variable on the dependent variable.

The overestimation of the effects of the independent variable on the dependent variable is enhanced to the extent that the independent variable causes the mediator and the mediator causes the dependent variable. Because a successful mediator is caused by the independent variable and causes the dependent variable, successful mediators measured with error are most subject to this overestimation bias.

The common approach to unreliability is to have multiple operations or indicators of the construct. Such an approach requires two or more operationalizations or indicators of each construct. One can use the multiple indicator approach and estimate mediation paths by latent-variable structural modeling methods. The major advantages of structural modeling techniques are the following: First, although these techniques were developed for the analysis of nonexperimental data (e.g., field-correlational studies), the experimental context actually strengthens the use of the techniques. Second, all the relevant paths are directly tested and none are omitted as in ANOVA. Third, complications of measurement error, correlated measurement error, and even feedback are incorporated directly into the model. The most common computer program used to estimate structural equation models is LISREL-VI (Jöreskog & Sörbom, 1984). Also available is the program EQS (Bentler, 1982).

We now turn our attention to the second source of bias in the mediational chain: feedback. The use of multiple regression analysis presumes that the mediator is not caused by the dependent variable. It may be possible that we are mistaken about which variable is the mediator and which is the dependent variable.

Smith (1982) has proposed an ingenious solution to the problem of feedback in mediational chains. His method involves the manipulation of two variables, one presumed to cause only the mediator and not the dependent variable and the other presumed to cause the dependent variable and not the mediator. Models of this type are estimated by two-stage least squares or a related technique. Introductions to two-stage least squares are in James and Singh (1978), Duncan (1975), and Judd and Kenny (1981a). The earlier-mentioned structural modeling procedures can also be used to estimate feedback models.

Overview of Conceptual Distinctions Between Moderators and Mediators

As shown in the previous section, to demonstrate mediation one must establish strong relations between (a) the predictor