Psychological Bulletin
1991, Vol. 110, No. 2, 315–317

0033-2909/91/$3.00

Cautions in Assessing Spurious "Moderator Effects"

James A. Shepperd
Holy Cross College

Lubinski and Humphreys (1990) illustrated how moderator effects uncovered using hierarchical multiple regression procedures may be spurious and that the predictor–criterion functional relationship may be better characterized by a nonlinear function. They recommended concurrently inspecting nonlinear functions when examining moderator effects. This article reviews the evidence presented by Lubinski and Humphreys to substantiate their argument and the generality of their findings beyond their data. Although an examination of nonlinear functions may be justified in their specific illustration, the blanket recommendation of inspecting polynomial terms when examining moderator effects is troublesome and may itself lead to spurious conclusions.

In a recent article in *Psychological Bulletin,* Lubinski and Humphreys (1990) presented a thoughtful discussion of how moderator effects uncovered using hierarchical multiple regression analyses (HMRA) may be spurious. Using data from the Project Talent Data Bank (Flanagan et al., 1962), Lubinski and Humphreys illustrated empirically that a nonlinear term (e.g., X^2 or $X^{1/2}$) may characterize the functional relationship between predictors (in their case, quantitative and spatial abilities) and the criterion (exceptional levels of mathematical sophistication) better than the interaction term. As a consequence, Lubinski and Humphreys concluded that researchers examining moderator effects should concurrently inspect nonlinear functions. In this article I review the evidence presented by Lubinski and Humphreys to substantiate their argument and the generality of their findings beyond their data set. Although an examination of nonlinear functions may be justified in their specific illustration, the blanket recommendation of inspecting polynomial terms when using HMRA to examine moderator effects is troublesome and may itself lead to spurious conclusions.

Problems With Moderate Effects

Lubinski and Humphreys (1990) noted two problems with assessing moderator effects. The first problem is low statistical power and the subsequent increased probability of a Type II error arising from using unreliable first-order predictors. Because the reliability of the interaction term can be represented as the product of the reliabilities of the first-order predictors, to the extent that the first-order predictors are unreliable (contain measurement error) the interaction term will also be unreliable. As a solution, Lubinski and Humphreys recommended using more reliable predictors.

Lubinski and Humphreys (1990) devoted most of their article to describing and illustrating empirically a second problem associated with assessing moderator effects, the Type I error. This problem arises from specifying a linear trend to characterize the predictor–criterion functional relationship when a nonlinear trend is more appropriate. To exemplify this potential problem, Lubinski and Humphreys conducted a series of analyses on data from the Project Talent Data Bank (Flanagan et al., 1962), using hierarchical multiple regression. In the first analysis, three terms were entered sequentially into a regression model: two first-order terms (denoted X and Z), followed by the Linear × Linear interaction term (XZ). The interaction term enhanced the prediction of the criterion beyond the two first-order terms, suggesting a classic moderator effect.

A second analysis was then conducted to examine whether a quadratic trend might better characterize the predictor–criterion functional relationship. Specifically, following the entrance of the first-order terms, the interaction term (XZ) and quadratic terms (denoted X^2 and Z^2) were entered simultaneously into the model in a stepwise fashion. In this procedure, the three higher order terms compete with one another to determine which function form best characterizes the predictor–criterion relationship. In seven of eight separate samples examined, Lubinski and Humphreys (1990) found that, of the three higher order predictors, the quadratic term X^2 accounted for the largest proportion of variance (i.e., R^2) and absorbed virtually all of the variance explained by the Linear × Linear interaction function examined in the previous model. Hence, they concluded that the moderator effect was spurious and that a quadratic trend best explains the predictor–criterion relationship.

An important caveat regarding these findings should be noted. The additional variance explained by the quadratic term across the eight subsamples was miniscule, ranging from 0% of the total variance for 9th-grade girls to roughly 3.4% for 12th-grade girls. Thus, although the quadratic term enhanced the prediction of the criterion relative to the interaction term, the

I thank Linda Carli, Charles Locurto, and Bill Graziano for helpful comments on a draft of this article.

Correspondence concerning this article should be addressed to James A. Shepperd, Department of Psychology, Holy Cross College, Worcester, Massachusetts 01610.

additional variance explained seems negligible. Nevertheless, one can at least imagine samples with less error in which the quadratic term explains more unique variance. Moreover, the Linear × Linear interaction term itself explained at most only 3.8% of the variance beyond the two first-order terms. Thus, it seems true, at least in this sample, that a quadratic term more accurately represents the predictor–criterion functional relationship than does the interaction term and that the hypothesized moderator effect is spurious.

The previous argument notwithstanding, there are several reasons why inspecting nonlinear functions when assessing the veracity of a moderator effect using HMRA may be unwise and, at the very least, should be viewed cautiously. First, the possibility that the predictor–criterion functional relationship is better characterized by a nonlinear trend is typically a concern only when the first-order predictors are highly correlated (i.e., when there is high multicollinearity). As Lubinski and Humphreys (1990) noted, when the predictors are highly correlated, then the interaction and quadratic terms will share substantial variance (i.e., as $r_{XZ} \rightarrow 1.0$, $r_{(XZ)(X^2)} \rightarrow 1.00$). If, however, the predictors are not highly correlated, then the higher order functions themselves will share only a moderate amount of variance. Thus, the type of spurious moderator effect that Lubinski and Humphreys cautioned against is of little concern to researchers who use orthogonal or only marginally correlated first-order predictors.

A second reason for exercising caution when inspecting nonlinear functions is that the test of such functions, similar to the test of interaction functions, may suffer unduly from low statistical power if the corresponding first-order term is not highly reliable. Lubinski and Humphreys (1990) noted that when first-order terms are multiplied to generate Linear × Linear interaction terms, the errors in measurement are multiplied as well. This results in low statistical power and an increased probability of committing the Type II error to the extent that one or both of the first-order terms is unreliable. What Lubinski and Humphreys failed to point out is that the quadratic term is also a multiplicative term (i.e., $X^2 = X \times X$) and thus shares with the interaction term the problem of low reliability if the corresponding first-order term (X) is not highly reliable. Thus, even with nonlinear terms one risks commission of a Type II error (e.g., statistically rejecting a quadratic functional relationship that actually is present) to the extent that the first-order term possesses measurement error.

A third reason for exercising caution in using the procedures recommended by Lubinski and Humphreys (1990) is that they may inadvertently lead to spurious findings. Lubinski and Humphreys recommended the inspection of quadratic and other nonlinear terms (e.g., $X^{1/2}$, $Z^{1/2}$, X^2Z, XZ^2) as a safeguard against committing a Type I error (i.e., erroneously specifying a relationship between the criterion and a higher order function). However, inspecting (or perhaps fishing for) relationships between higher order functions and a criterion will itself increase the probability of committing a Type I error. After all, each additional term inspected (nonlinear or otherwise) represents an additional significance test, and the probability of finding a spurious predictor increases with every additional significance test conducted. As a consequence, inspection of quadratic and other nonlinear terms, unless specifically predicted, should be viewed as exploratory and treated prudently.

A final reason for exercising caution when inspecting nonlinear functions is that, unless the first-order term(s) predict perfectly, with very large samples, virtually any predictor (nonlinear functions included) is likely to be statistically significant and to account (however small) for some proportion of the variance. As a result, a decision may be made to include in the final model a nonlinear function (either instead of or in addition to an interaction function) that has little theoretical meaning and adds little in the way of predictive power. Importantly, Lubinski and Humphreys (1990) noted that statistical significance becomes practically meaningless with very large samples, making it inappropriate to use conventional significance levels as a criterion for including a predictor in the model. However, given their recommendation to inspect nonlinear functions when using HMRA, the use of discretion in interpreting significant predictors (particularly predictors that were not anticipated a priori to be significant) is advised.

Summary and Conclusion

The warning regarding spurious moderator effects raised by Lubinski and Humphreys (1990) is important and should be heeded when assessing the veracity of interactions involving highly correlated first-order predictors. Moreover, they provided a compelling illustration of a spurious moderator effect in the psychological literature: On the basis of their analyses, the development of exceptional levels of mathematical sophistication indeed appears to be characterized better by a quadratic function involving math ability than by a Linear × Linear function involving math and spatial ability. As noted earlier, however, when the first-order terms are not highly correlated, the likelihood that a moderator effect is spurious, reflecting a nonlinear function in disguise, is greatly diminished. Thus, one remedy to this type of error in model specification, although certainly not practical for all research, is to avoid correlated first-order predictors.

When the use of correlated first-order predictors is necessary or unavoidable, there is a statistical method for determining the extent to which multicollinearity is a serious problem in the fitted model: One can compute the variance inflation factor for each predictor in the model. The variance inflation factors assess the extent to which the variances of the estimated regression coefficients are inflated when a linear rather than a nonlinear relationship among the predictors is assumed. Computation of the variance inflation factors (or their reciprocal, tolerance) is an option in a number of computer regression programs, and a more detailed discussion of this procedure is provided elsewhere (e.g., Neter, Wasserman, & Kutner, 1983).

Should an investigator feel the need to inspect nonlinear functions when fitting a model using multiple regression, such an inspection should be undertaken with caution unless the nonlinear function is predicted a priori. Similar to "spurious" moderator effects, nonlinear effects can be spurious as well. Moreover, one should avoid the error of including a nonlinear function in the final model that is statistically significant but has no practical meaning or predictive power. If there is theoretical justification for predicting a moderator effect and competing theoretical justification for predicting a nonlinear effect, then the two effects should be examined on equal footing. However, in the absence of theoretical justification for anticipating one effect or the other, it is wise to adopt a stricter probability level for the nonpredicted effect.

Undoubtedly, there are many mathematical functions that can describe a given empirical relationship. However, the inductive approach of letting "the data decide on the precise functional relation responsible for observed incremental validity" (Lubinski & Humphreys, 1990, p. 390) lends itself to serious problems of inflated Type I error rates. Therefore, rather than "fishing" for the best mathematical description for a given criterion–predictors relationship, the inclusion of nonlinear or multiplicative variables in multiple regression should be based on theoretical considerations. After all, what is crucial is the *explanation* for the relation, not the statistical description. To the extent that more complex equations lend themselves to better explanations, they should be used. Otherwise, the familiar rule "simpler is better" (Cohen, 1990) is still good advice.

References

Cohen, J. (1990). Things I have learned (so far). *American Psychologist, 45,* 1304–1312.

Flanagan, J. C., Dailey, J. T., Shaycoft, M. F., Gorham, W. A., Orr, D. B., & Goldberg, I. (1962). *Design for a study of American youth.* Boston: Houghton Mifflin.

Lubinski, D., & Humphreys, L. G. (1990). Assessing spurious "moderator effects": Illustrated substantively with the hypothesized ("synergistic") relation between spatial and mathematical ability. *Psychological Bulletin, 107,* 385–393.

Neter, J., Wasserman, W., & Kutner, M. H. (1983). *Applied linear regression models.* Homewood, IL: Irwin.

Received August 14, 1990
Revision received December 14, 1990
Accepted December 18, 1990 ■