Research Practices that Produce Type VI Errors

Isadore Newman
The University of Akron

John W. Fraas
Ashland University

Carole Newman
The University of Akron

Russell Brown
Summa Health Care System

A Type VI error occurs when an inconsistency exists between the study's research question and the analytic technique and/or research design used in the study. This article presents various research practices that lead to Type VI errors. Specifically discussed are the practices that (a) fail to distinguish between statistical analysis and research design issues, (b) do not match the model used in structural equation modeling with the research question, (c) analyze a research question that involves practical significance with an analytical technique that fails to do so, (d) use methods designed to control for inflated Type I error rates that do not match the nature of the research question, and (e) employ multivariate data

analysis techniques for research questions that require the application of univariate techniques. An examination of these types of mismatches between the analytical technique and/or the research design and the research question should assist researchers in avoiding Type VI errors.

Current Research Practices that Produce Type VI Errors

The research and analytical techniques used by the researcher should not lead to Type VI errors, that is, a condition in which an inconsistency exists between the study's research question and the analytic technique and/or research design (Newman, Deitchman, Burkholder, Sanders, & Ervin, 1976). Type VI errors are committed when researchers (a) fail to distinguish between statistical analysis and research design issues, (b) do not match the model used in structural equation modeling with the research question, (c) analyze a research question that involves practical significance with an analytical technique that fails to do so, (d) use methods designed to control for inflated Type I error rates that do not match the nature of the research question, and (e) employ multivariate data analysis techniques for research questions that require the application of univariate techniques.

The current state of computer hardware and software allows researchers to apply sophisticated statistical techniques with relative ease. It must be noted, however, that the point-and-click environment of today's computer statistical software allows researchers to use these sophisticated techniques in a rather mechanical fashion. Thus, while such ease of use leads to efficiency, it may also foster the inappropriate match between the analytical technique and the research question, which will lead to a Type VI error. An examination of these types of mismatches between the study's research question and the analytic technique and/or research design should assist researchers in avoiding Type VI errors.

Differences Between Issues Addressed by Statistical Analysis and Research Design

A Type VI error is likely to occur when a researcher does not distinguish between the issues addressed by the analytical techniques employed in a study and those issues dealt with by the study's research design. Although this distinction may seem obvious, it is vastly overlooked by too many researchers. To illustrate, in 1980 the United States Supreme Court based decisions in racial and sex discrimination suits on statistical analyses that reported statistically significant relationships (i.e., correlation values). However, these studies were void of the types of research design controls discussed by Fisher (1971), Campbell and Stanley (1963), and Newman and Newman (1994). These decisions were based on the misunderstanding that the use of sophisticated statistical analyses allows one to assume causation. If the research question was designed to determine a cause for the discrepancies between opportunities for advancement or pay, the use of correlations would be one form of a Type VI error. If, however, the research question attempted to establish the existence of relationships, the applications of correlation analytical techniques would not produce a Type VI error unless the researcher attempted to make casual inferences from those correlation results.

To further understand the appropriate role of all statistical tests, it is important to note that all tests of significance are tests of relationships. Therefore, each statistical test employed by researchers can be mathematically translated into the degree of relationship between the variables. Conversely, if one knows the sample size, all mathematical relationships can be converted into tests of significance. For example, as noted by McNeil, Newman, and Kelley (1996), a *t* test is equivalent to a point-biserial correlation; an *F* test is equivalent to an *eta* value or an *R* value; and a 2X2 chi-square value is equivalent to a *phi* coefficient. The point biserial correlation, *phi*, *eta*, *t*, *R*, *F* and chi-square values are all related. Thus, any test of significance merely gives a probability estimate that a relationship is not due to random variation. It does not yield any information about the causal nature of that relationship (i.e., the degree of internal validity). Only to the extent that a research design has a high degree of internal validity can one assume causality.

Type VI Error Concerns with Structural Equation Modeling

Structural Equation Modeling (SEM) has been conceptualized as the "marriage of path analysis to factor analysis." Thus, SEM is purported to have many advantages including the ability to (a) represent latent variables (or factors) that are theoretically presumed to be the cause of observed variables, (b) model theoretical relationships, (c) estimate measurement error, and (d) estimate the overall "fit" of the model (Quintana & Maxwell, 1999). Of these advantages, perhaps, the most important is that the models used in conjunction with SEM tend to be theoretically based. It is important to note, however, that researchers who mechanically make certain assumptions and interpretations with SEM increase the chances of committing Type VI errors. Five of these assumptions and interpretations are (a) the assumption of causation, (b) the practice of not including interaction terms, (c) the interpretation of the goodness-of-fit indices, (d) assumptions regarding factor fit, and (e) the assumption regarding measurement error due to factor instability.

The assumption of causation. Some authors who use SEM have the tendency to interpret their results in a manner that implies, if not directly states, that a causal relationship exists. For example, in a study by Aquino, Russell, Cutrona, and Altmaier (1996), which examined the relationship between social support, employment and life satisfaction in the elderly, the authors provided a causal interpretation of an ex post facto design. With respect to the appropriateness of this causal interpretation, it is important to note that SEM is an "extension of the collection of statistics belonging to the general linear model (GLM)" (Quintana & Maxwell, 1999, p. 486), and is, therefore, an analytic technique and not a research design. Yet, SEM has been explicitly described as causal modeling (Fassinger, 1987). The components of SEM (factor analysis and path analysis) are both based on the analysis of the covariance/correlation matrix. As such, SEM is still based on the degree of relationship (correlation) among variables. Thus SEM does not inherently control for threats to internal validity. Correlation and causation are not the same. Since causal inferences depend on appropriate research design and a statistical method does not allow for causal inference, researchers must be keenly aware of what researchers can conclude based on the statistical analysis and what limitations the study's research design imposes on those conclusions, even with SEM.

The practice of not including interaction terms. SEM is infrequently used to test interaction effects (Newman & Marchant, 1993; Ridenour, Marchant, & Newman, 1994). Therefore, if the theory assumes interaction, which should be reflected in the research question, an inconsistency between the analysis and the research question would exist. It should be noted, however, that researchers who use SEM have dealt with interaction effects by analyzing separate models for variables that interact. This is a simple and elegant solution when (a) there are very few variables that interact, (b) it is a first order interaction, and (c) the variables are dichotomous (e.g., gender) in nature rather than ordinal or interval.

The interaction problem is relatively easy to resolve for a dichotomous variable as was modeled by Robitschek and Kashubeck (1999). They developed separate models for males and females in order to model the relationship of family functioning and alcoholism to a person's well-being and level of distress. The inclusion of interaction terms in a SEM analysis becomes considerably more complicated when there are many interactions, the variables are ordinal or interval in nature (e.g., intelligence), and higher order interaction exists (e.g., intelligence by age, and intelligence by age by gender). It is not unlikely for researchers to encounter these types of interactions and variables when testing social science theories. Since there is no mathematical model, of which we are aware, that can adequately assess such research questions when using SEM (Newman & Marth, 1995), the use of SEM to test research questions that include these types of interactions would produce Type VI errors. Equally important is the possibility that researchers who are committed to using SEM may be less inclined to propose research questions involving these types of interaction effects. Thus, an inconsistency between the analytic technique used and the "appropriate" research question would exist.

Interpretation of the goodness-of-fit indices. An inconsistency may exist between the method used with SEM to assess the model's goodness of fit and the intent of the research question. Hair, Anderson, Tatham, and Black (1998) present a plethora of goodness-of-fit measures that are used in conjunction with SEM. They divided these goodness-of-fit measures into three types — measures of absolute fit, measures of incremental fit, and measures of parsimonious fit. The measures of absolute fit include: (a) the likelihood-ratio chi-square statistic, (b) the noncentrality parameter, (c) the goodness-of-fit-index, (d) the root mean square residual, (e) the root mean square error of approximation, (f) the expected cross-validation index, and (g) the cross-validation index. Some of the measures of incremental fit are: (a) the adjusted goodness-of-fit index, (b) the Tucker-Lewis index, (c) the normed fit index, (d) the relative fit index, (e) the incremental fit index, and (f) the comparative fit index. Parsimonious fit measures include: (a) the parsimonious normed fit index, (b) the parsimonious goodness-of-fit index, (c) the normed chi-square, and (d) the Akaike information criterion.

These goodness-of-fit measures are based on a similar concept (i.e., the model's ability to reproduce the correlation matrix). This method of assessing model fit can lead to an incorrect assessment under two possible scenarios. First, it is possible for a model that contains numerous paths to produce a goodness-of-fit value that indicates a good level of fit in spite of the fact that none of the path values are statistically significant. Second, it is possible for the model to effectively reproduce the correlation matrix even when some of the relationships are in the opposite direction from the direction predicted by the theory. While each of these methods assesses the ability of the model to reproduce the correlation matrix, it does not provide a direct means to assess the number of relationships that are statistically significant and in the direction predicted by the theory. In fact, a number of the assessment measurements could be high and all of the relationships could be in the opposite direction of the theory! Interpretation of fit indices, alone, as support or lack of support for a theoretical model could produce a Type VI error.

Consideration of the number of path values that are statistically significant in the predicted direction would decrease the likelihood of committing a Type VI error when using SEM. This method of assessing the model's goodness of fit is not based on the model's ability to reproduce the correlation matrix, but rather on the ability of the model to reflect the path values as suggested by theory. The use of this approach, which has been referred to as the Binomial Index of Model Fit (Fraas & Newman, 1994; Newman, Fraas & Norfolk, 1995), would reduce the chance of an inconsistency between the assessment of the model's goodness of fit and the research question.

Assumptions regarding factor fit. A Type VI error can be committed when SEM is used to determine if the factors are good estimates of the theoretical factor structure (i.e., SEM is used for confirmatory factor analysis). In a measurement model with three factors, it is very possible that two factors fit the structure well and one does not. For such a case, the traditional goodness-of-fit techniques may simply indicate a poor fit without revealing which factors do not fit the structure. If the research question requires the analysis to reveal which factors do and do not fit the structure and the assessment technique is not capable of providing such evidence, a Type VI error will be committed. To avoid this error, researchers could employ the Kaiser Factor Matching technique (Newman, Dimitrov, & Waechter, 2000). This technique would indicate which of the two factors fit well. Newman, et al. suggested that Kaiser Factor Matching be used along with the more traditional estimates to get a better picture of the relationships of the factors to the underlying theoretical constructs and thus avoid a Type VI error.

The assumption regarding measurement error due to factor instability. The use of SEM could lead to another form of Type VI error based on how it purports to control for measurement error. That is, the underlying factors generated by SEM may not adequately represent the variables contained in the research question. A somewhat over simplification of this procedure may assist in the explanation of this mismatch. Instead of using the indicator variables, SEM uses the underlying factors as predictor variables. The assumption is that the factors are stable and therefore control for measurement error. It is likely that this procedure does reduce measurement error but it does not eliminate it. Thus, the use of these factors may cause other types of errors due to the fact that the factors may be quite specific to the sample analyzed. This problem is likely compounded by the tendency of researchers to use only a small number of indicators, such as three or four per latent variable, a practice that has been advocated in the literature (Quintana Maxwell &, 1999). In some cases, single indicators are used to represent latent variables. For example, Tuason and Friedlander (2000), in a study of Bowen's theory of differentiation, used single indicators to represent three of six of the latent variables in their study. This was accomplished by supplying estimates of the reliability of the measurement of each of the indicator variables. Traditionally, it has been reported that five or more items are necessary to create a stable factor (Nunnally, 1972). Thus, a good but infrequently used practice would be to cross-validate the model to estimate the stability of the latent variables. The use of cross-validation would assist researchers in avoiding Type VI errors.

Importance of Statistical Significance and Practical Significance

Numerous researchers and journal editors are calling for the inclusion of practical significance (e.g., effect sizes) in research results and, thus, presumably in research questions (Cohen, 1977, Fraas & Newman, 2000; Huberty, 1993; Levin & Robinson, 2000; Newman & Fraas, 2001; Robinson & Levin, 1997; Shaver, 1993; Thompson, 1996, 1998, 1999a, 1999b, 1999c). If research questions contain a practical significance component and the analytic techniques used to test that research question do not incorporate a practical significance component, a Type VI error will be committed. Thus, in order to avoid a Type VI error for such research questions, it has been argued that the analytical techniques used must address both the issues of statistical significance and practical significance (Fraas & Newman, 2000; Newman & Fraas, 2001; Robinson & Levin, 2000).

Fraas and Newman (2000) and Newman and Fraas (2001) discussed the testing of non-nil null hypotheses (i.e., hypotheses that incorporate non-zero values) as a means of incorporating both the test of statistical and practical significance. To illustrate the use of a non-nil null hypothesis, consider a study designed to test the effectiveness of two treatments intended to produce weight losses for cardiac patients. An important issue for researchers to initially address is the establishment of how large the difference in the weight-loss means of the two groups must be in order for it to be considered important, that is, practically significant. Fraas and Newman expressed the view that this process is probably best undertaken by involving practitioners (i.e., people directly involved in the process being studied) and researchers in the field.

In this illustration, we are assuming that practitioners and researchers have identified that any difference between the group means greater than 10 pounds would be considered practically significant. For this practical significance level, the non-nil null hypothesis that the mean for Group 1 minus the mean for Group 2 equals 10 pounds is more meaningful clinically than the nil null hypothesis that the mean for Group 1 minus the mean for Group 2 equals zero pounds. That is, instead of saying we are 95% confident that the difference is greater than zero, we can say we are 95% confident that the difference is greater than 10. If experts determined that a difference of 10 pounds was minimally required for clinical benefit, the latter statement is far more meaningful and useful for the practitioners who will utilize the study's results.

It is important to note that the use of non-nil null hypotheses require the researchers and practitioners to identify what is clinically or pragmatically meaningful before initiating the research. The decisions determining the appropriate effect size need to be related to the purpose for doing the research. Newman, Ridenour, Newman, and DiMarco (2002) indicate it is important that researchers frame their research question so it is consistent with their purpose. This clear understanding of purpose is necessary to help the researcher identify the appropriate effect size. Without looking at the research question in the context of its purpose, one cannot identify an appropriate and useful clinical effect size. The incorporation of the effect size into the research question dictates that statistical significance is not considered independent of the practical significance level.

An additional point should be noted with respect to establishing the practical significance level. When dealing with the issue of how large should an effect size be in order for it to be considered practically significant, Newman and Newman (2000) presented the case that even small effect sizes, as measured by small R-square values, may be important, that is, practically significant. They concluded that small R-square values could be very valuable and useful, depending on the research question being asked and the size of the population. This position, which is supported by Rosenthal and Rosnow (1991) and Deming (1982), suggests that practical significance should be measured in a relative sense rather than an absolute manner. This position again suggests that the involvement of practitioners in establishing the practical significance level is an essential component in establishing a meaningful research question.

It is essential that the established practical significance level be incorporated into the analytic technique used to test the research question in order to avoid committing a Type VI error. Fraas and Newman (2000) suggested that one technique that could be used to test a non-nil null hypothesis is a randomization test. The use of a randomization test, which generates its own distributions, could be used to determine if the observed difference adjusted for the practical significance level is unlikely to be due to chance variation. It should be noted, however, that Newman and Fraas (2001) conducted a Monte Carlo study regarding the impact of including the practical significance level into the research question on the Type I error rates of independent *t* tests of two group means. The results revealed negligible impacts of such tests on the Type I error rates. Thus, non-nil null hypotheses may, under certain conditions, be tested with parametric tests.

Controlling Type I Error Rates

Researchers are familiar with the practice of controlling for inflated Type I error rates (i.e., the chance of rejecting a true null hypothesis) in studies that involve multiple statistical tests. If the adjustment technique the researchers use, however, does not match the research question with respect to the assumed error rate unit and directional and non-directional relationships, a Type VI error will be committed. Newman, Fraas and Laux (2000) suggested a non-mechanical approach that requires the researcher reflect on three elements of the adjustment process.

First, the researcher identifies the error rate unit or units (i.e., the various groupings of tests for which adjustments are to be made). Second, adjustments are not made for directional tests (i.e., the hypotheses reflect a direction based on theory and/or previous empirical results). Newman, Fraas, and Laux (2000) suggested that if a researcher adjusts the alpha levels for tests that are predicted from previous experience or theory, the researcher would probably overcorrect for possible inflated Type I errors. Such an outcome would increase the likelihood of Type II errors being committed. Third, the alpha levels for the non-directional tests in a given error rate unit are adjusted for the number of such tests. This approach requires researchers to reflect on the adjustment process at two stages of the research process (a) the point at which the research question is being formulated and (b) the point at which the analytic technique is being chosen and implemented. Reflecting on the adjustment process at both of these stages should decrease the likelihood that the adjustment technique used does not match the research question, that is, decrease the chance of a Type VI error.

Multivariate Tests versus Univariate Tests

Various statistics books recommend that if one has multiple dependent variables, multivariate analyses should be used. One reason for this recommendation relates to controlling for inflated Type I error rates when multiple statistical tests are conducted on a set of dependent variables (Stevens 1996). Unfortunately this recommendation is frequently misunderstood since the overall multivariate test of significance is only one aspect of the multivariate testing procedure. When a significant multivariate test is followed by univariate tests to determine where the significant differences lie, a control for multiple tests is still necessary to keep the Type I error rate from being inflated (Croom, 1987; Newman, Croom, Mugrage & Hoedt, 1983).

The most important consideration in determining whether one should use multivariate or univariate measures should be how well the statistical procedures reflect the question of interest (Newman & Benz, 1987). If the question of interest is not related to the underlying construct of a set of independent variables, but is instead related to each dependent variable, univariate analyses should be used. Multivariate analyses should only be used when one is interested in the underlying hypothetical construct of the set of dependent variables, which is frequently not the case (J. Finn, personal communication, October, 2001; Thompson, 1994). The use of multivariate analyses with a research question that reflects an interest in each dependent variable will produce a mismatch between the analytic technique and the research question, which is a Type VI error.

An additional point is important to consider regarding the use of multivariate analyses. All multivariate analysis of variance techniques are subsets of canonical analysis, just as traditional analysis of variance techniques are subsets of multiple regression analyses. A canonical correlation measures the degree of association between the underlying construct of a set of predictor variables with the underlying construct of a set of criterion variables. It does not answer the question: What is the degree of association among the individual predictor variables and the individual criterion variables simultaneously? If a researcher is interested in each dependent variable separate from the others and the researcher conducts a multivariate analysis, a Type VI error would be committed. Similarly, if a researcher is interested in a linear combination of the dependent variables and the researcher conducts a univariate analysis, a Type VI error would also be committed. The research must match the multivariate and univariate analytical techniques with the appropriate research questions.

Implications for Researchers

The intent of this article is to alert researchers to possible research practices that may lead to inconsistencies among the research questions, the analytical techniques employed and/or the research designs (i.e., Type VI errors). To avoid Type VI errors, researchers must be careful not to approach the research process in a mechanical fashion, but rather they must evaluate whether their analytic techniques and research design are appropriate for their research questions.

Specifically, a researcher can decrease their chances of committing Type VI errors by addressing a number of key questions:

1. When stating the conclusions resulting from my study, did I clearly delineate and separate the issues addressed by my research design from those issues related to the analytic techniques employed?

2. If I used SEM, did my model and analysis reflect my research question, including the need to: (a) incorporate interaction variables, (b) use appropriate goodness-of-fit measures, (c) reveal which factors do and do not fit the structure, and (d) have the underlying factors generated by SEM adequately represent the variables.

3. If my research question contains a practical significance component, do my analytic techniques allow me to adequately test that research question? If my research question does not contain a practical significance component, should it?

4. If my study involves multiple statistical tests, does the method I used to control for inflated Type I error rates match the nature of my research question? That is, have I accurately identified my error rate units and appropriately considered the directional versus the nondirectional nature of my hypotheses?

5. Did the analytic technique that I employed accurately reflect whether the nature of my research question was univariate or multivariate?

We believe that Type VI errors would be less common in today's research if researchers would reflect on the issues addressed by these questions.

References

Aquino, J., Russell, D., Cutrona, C., & Altmaier, E. (1996). Employment status, social support, and life satisfaction among the elderly. *Journal of Counseling Psychology, 43*(4), 480-489.

Campbell, D.T., & Stanley, J.C. (1963). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences.* New York: Academic Press, Inc.

Croom, W. (1987). *Selected psychometric properties and measures of interpretability of two methods of estimating population parameters: Canonical correlation and multiple linear regression.* Unpublished doctoral dissertation, The University of Akron, Akron, OH.

Deming, E. (1982). *Out of crisis: Center for advanced engineering study.* Cambridge: Massachusetts Institute of Technology.

Fassinger (1987). Use of structural equation modeling in counseling psychology research. *Journal of Counseling Psychology, 34*(4), 425-436.

Fisher, R.A. (1971). *Collected papers of R.A. Fisher.* Edited by J.H. Bennett. Adelaide: The University of Adelaide.

Fraas, J.W., & Newman, I. (1994). A binomial test of model fit. *Structural Equation Modeling: A Multidisciplinary Journal, 1*(3), 268-273.

Fraas, J.W., & Newman, I. (2000, October). *Testing for statistical and practical significance: A suggested technique using a randomization test.* Paper presented at the annual meeting of the Mid-Western Educational Research Association, Chicago, IL.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Huberty, C. J. (1993). Historical origins of statistical testing practices: The treatment of Fisher versus Neyman-Pearson views in textbooks. *The Journal of Experimental Education, 61*(4), 317-333.

Levin, J.R., & Robinson, D.H. (2000) Rejoinder: Statistical hypothesis testing, effect size estimate, and the conclusion of coherence of primary research studies. *Educational Researcher, 29*(1), 34-36.

McNeil, K., Newman, I., & Kelly, F. (1996). *Testing research hypotheses with the general linear model.* Carbondale, IL: Southern Illinois University Press.

Newman I. & Benz, C. (1987). *Comparing canonical correlation strategies to model comparison multiple regression on one data set* (technical report). Akron, OH: The University of Akron, Office of Educational Research and Evaluation.

Newman, I., Croom, W., Mudgrage, B., & Hoedt, K. (1983). Note: A suggested method for correcting alpha error build-up on orthogonal dependent variables. *Multiple Linear Regression Viewpoints, 12*(1), 56-60.

Newman, I., Deitchman, R., Burkholder, J., Sanders, R., & Ervin, L. (1976). Type VI Error: Inconsistency between the statistical procedure and the research question. *Multiple Linear Regression Viewpoints, 6*(4), 1-19.

Newman, I., Dimitrov, D., & Waechter, D. (2000). Factor structure of perceived individualization of instruction: Argument for multiple perspective. *Educational Research Quarterly, 24*(1), 20-29.

Newman, I., & Fraas, J.W. (2001, October). *Testing non-nil null hypotheses with randomization and t tests: A Monte Carlo study.* Paper presented at the annual meeting of the Mid-Western Educational Research Association, Chicago, IL.

Newman, I, Fraas, J.W., & Laux, J. (2000). A three-step adjustment procedure for Type I error rates. *Journal of Research in Education, 10*(1), 84-90.

Newman, I., Fraas, J.W., & Norfolk, T. (1995). Binomial index of model fit: An elaboration. *Structural Equation Modeling: A Multidisciplinary Journal, 2*(2), 155-162.

Newman, I. & Marchant, G. J. (1993, April) *Type VI error in path analysis: Testing for interactions.* Paper presented at the annual meeting of the American Educational Research Association, Atlanta, Ga.

Newman, I., & Marth, J. (1995). To path analyze or to not path analyze: Is there an alternative approach. *Multiple Linear Regression Viewpoints, 22*(1), 7-12.

Newman, I., & Newman, C. (1994). *Conceptual statistics for beginners.* (2nd ed.). New York: University Press of America.

Newman, I., & Newman, C. (2000). A discussion of low R^2: Concerns and use. *Educational Research Quarterly, 24*(2), 3-9.

Newman, I., Ridenour, C., Newman, C., & DiMarco, G. (2002). A typology of research purposes and its relationship to mixed methods. In A. Tashakkori & C. Teddie (Eds.), *Handbook of mixed methods in social and behavioral research.* Thousand Oaks, CA: Sage Publications.

Nunnally, J. (1972). Educational measurement and evaluation. New York: McGraw-Hill.

Quintana, S., & Maxwell, S. (1999). Implications of recent developments in structural equation modeling for counseling psychology. *The Counseling Psychologist, 27*(4), 485-527.

Ridenour, T. A., Marchant, G. J., & Newman, I. (1994, August). *Interaction effects in path analysis in psychology.* Paper presented at the annual meeting of the American Psychological Association, Los Angeles, CA.

Robinson, D.H., & Levin, J.R. (1997). Reflection on statistical and substantive significance, with a slice of replication. *Educational Researcher, 26*(5), 21-26.

Robitchek, C., & Kashubeck, S. (1999). A structural model of parental alcoholism, family functioning, and psychological health: The mediating effects of hardiness and personal growth orientation. *Journal of Counseling Psychology, 46*(2), 159-172.

Rosenthal, R., & Rosnow, R. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York: McGraw-Hill.

Shaver, J. P. (1993). What statistical significance testing is, and what it is not. *Journal of Experimental Education, 61*(4), 293-316.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences.* Mahwah, NJ: Lawrence Erlbaum Associates.

Thompson, B. (1994, February). *Why multivariate methods are usually vital in research: Some basic concepts.* Paper presented as a Featured Speaker at the biennial meeting of the Southwestern Society for Research in Human Development (SWSRHD), Austin, TX.

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. *Educational Researcher, 26*(5), 29-32.

Thompson, B. (1998). Review of what if there were no significance tests? *Educational and Psychological Measurement, 58*(2), 334-346.

Thompson, B. (1999a). If statistical significance tests are broken/misused, what practices should supplement or replace them? *Theory and Psychology, 9*(2), 165-181.

Thompson, B. (1999b). Statistical significance tests, effect-size reporting and the vain pursuit of pseudo-objectivity. *Theory and Psychology, 9*(2), 191-196.

Thompson, B. (1999c). Journal editorial policies regarding statistical significance tests: Heat is to fire as p is to importance. *Educational and Psychological Review, 11*(2), 157-169.

Tuason, M., Friedlander, M. (2000). Do parents' differentiation levels predict those of their adult children? And other tests of Bowen theory in a Philippine sample. *Journal of Counseling Psychology, 47*(1), 27-35.

Dr. Isadore Newman is a Distinguished Professor at the University of Akron, and a Professor in the Department of Educational Foundations and Leadership where he teaches research methods, advanced statistics and Ph. D. research seminars. He has also held the Harrington Endowed chair. Dr. Newman has writen several books on qualitative and quantitative research methods and has over 100 published articles.

Dr. Carole Newman is an Associate Professor in the Department of Curricular and Instructional Studies at the University of Akron where she teaches graduate and undergraduate courses in teaching methods. She is an educational consultant for area school districts and educational service centers.

John W. Fraas is a professor of business administration and a Trustees' Professor at Ashland Universitiy. He was

the initial appointee to the position of Trustees' Professor, a position awarded by the University's Board of Trustees to faculty members who have excelled in the areas of teaching, scholarship, and University service.

Russell Brown is completing his doctorate in Counseling Psychology at The University of Akron. He is currently an adjunct faculty member at Cleveland State University and the University of Akron where he teaches Statistics, Research Methodology and Developmental Psychology.