

Commentary on Quantitative Methods in I/O Research

Robert MacCallum

Department of Psychology

Ohio State University

To appear in *The Industrial-Organizational Psychologist* (1998, Vol. 35, No. 4).
Bowling Green, OH: Society of Industrial and Organizational Psychology.

This document can be downloaded from worldwide web site

<http://quantrm2.psy.ohio-state.edu/maccallum/>

Address correspondence to Robert MacCallum, Department of Psychology,
142 Townshend Hall, 1885 Neil Avenue, Columbus, OH 43210-1222.

Email: maccallum.1@osu.edu.

Commentary on Quantitative Methods in I/O Research

Robert MacCallum

Ohio State University

I am pleased and flattered to have been invited to contribute this commentary on the use of quantitative methods in research in I/O psychology. I offer my views and perspectives as an interested outsider. I have been teaching quantitative psychology at Ohio State University since 1974 and have had the pleasure of teaching many fine students in the OSU I/O program as well as interacting and collaborating with faculty colleagues in that program. A number of the views expressed in this commentary have benefited from recent discussions with those colleagues, including Bob Billings, Mary Roznowski, Jim Austin, and Phyllis Panzano. I have an abiding interest in how quantitative methods are used in empirical research in various fields of psychology, including I/O. I hope my comments will serve to reinforce much that is good with regard to this dimension of research in your field, and also that I can suggest some perspectives and approaches that might further enhance the use of quantitative methods in your research.

The Role of Quantitative Methods

Researchers in I/O psychology have long appreciated and embraced the critical role of quantitative methodology in the research enterprise. Strong I/O graduate programs typically put great emphasis on methodological training. The research literature in your major journals is characterized for the most part by sound design and use of appropriate, often sophisticated, methods of data analysis. Clearly your journal editors and reviewers place great value on this aspect of research.

With such strong training and emphasis, however, comes a potential trap wherein the research process can become driven by methods. Katzell (1994) expresses this concern in his discussion of meta-trends in I/O psychology. If a researcher focuses too much on methods, he or she may formulate research questions based on methodology, or may choose to study trivial questions that can be studied using a fashionable technique. Clearly, formulation of research questions should be theory-driven or problem-driven rather than technique-driven. Given a theory-based or problem-based question, one then determines a research design and method of analysis that will provide the best answer. Such an approach might or might not involve use of a sophisticated quantitative technique. We sometimes forget that important research is not necessarily characterized by the application of sophisticated or complex statistical techniques. In fact, important research might well involve very simple approaches, including case studies or qualitative analyses. Thus, I would urge researchers to avoid method-driven research, as well as to avoid biases against simple designs and analyses and in favor of complex analyses and fashionable methods. We shouldn't be impressed by flashy analyses that address trivial questions. Rather, we should place highest value on theory-driven and problem-driven studies that address important questions and that use sophisticated quantitative knowledge and insight to choose the best methodological approach.

There is another worrisome aspect of this concern about method-driven research. Just as researchers can fall into the trap of letting their research be driven by sophisticated quantitative techniques, they can also become locked into thinking about research problems from only one, possibly very simple, method-based perspective. The most obvious example of this phenomenon is the affliction I'll call "ANOVA Mindset

Syndrome.” In the most serious cases, victims of AMS are unable to conceive of a research problem in any terms other than those defined by an ANOVA design, and are unable to analyze data by any other method. For example, suppose an AMS victim is interested in a problem involving a relationship between two observed variables measured on numerical scales such as job satisfaction and job performance. He or she might resort to such approaches as gathering data from extreme groups on the independent variable of job satisfaction, or else gathering data on the full range of the variable but then converting it into a high-low categorical variable by a median split. Then analyses can be conducted by ANOVA to test whether high and low satisfaction “groups” differ in performance. If you know colleagues who do these things, you might suggest that they consult a regression specialist soon while there is still time!

Facetiousness aside, conducting research from such a perspective is limiting in many ways, and is also completely unnecessary. Such a narrow perspective limits the sort of problems that can be studied effectively. The failure to appreciate individual differences and study them using correlational methods incurs a great cost in loss of information and understanding of the phenomena under study. Dichotomization of variables, in particular, carries severe costs in terms of loss of power, effect size, and reliability (Cohen, 1983). Researchers can get stuck in other “mindsets” as well. The cure is to choose the design and analysis method that provide the best answers to the research questions, whether that be an ANOVA approach or a correlational approach or something else.

Of course, given my own background, I am most interested in those situations where an important research question is best addressed using a sophisticated quantitative

method. I am especially interested in models and methods for studying correlational data, and the I/O field has a long tradition of extensive and rigorous use of such methods. Much research in your field is non-experimental (about 50% according to Stone-Romero, Weaver, & Glenar, 1995), and there is an appreciation of individual differences along with a desire to understand and explain such differences. This perspective has led to wide usage of methods such as regression, factor analysis, and structural equation modeling (Stone-Romero et al., 1995). Given this pattern of usage, and given that my own methodological interests focus on such techniques, I wish to offer some specific comments about the application of such methods in empirical research.

Structural Equation Modeling

I begin with structural equation modeling (SEM), whose use in your field has rapidly increased in recent years (Stone-Romero et al., 1995). Many SEM applications published in I/O journals are well done and yield insights about relationships among central constructs in the field. SEM is a very alluring technique because it seems to provide a tool for determining the validity of hypotheses about patterns of relationships among hypothetical constructs. But we must take care to use SEM from an appropriate perspective. A fundamental principle that users of this method should always keep in mind is that *all structural equation models are wrong*. Some are more wrong than others. Of course, this principle applies to modeling in virtually any discipline. In the present context and in any specific study it is unproductive and invalid to think that there exists some parsimonious true model that holds exactly in the population and that our task is to find it. The world is far too complex and structural equation models are far too simple for that view to work. In fact, the best we can hope for in an SEM study is to find

a model that is parsimonious and substantively meaningful with regard to its structure and the resulting parameter estimates, and which fits our empirical data adequately well. If we achieve that goal, it would be wonderful if we could then believe that we have identified the true pattern of relationships among our variables and that we could make interpretations, draw conclusions, and take action on that basis. But that, of course, is never the case. In fact, such an outcome yields a model that can be viewed only as providing one plausible and approximate explanation of the real-world phenomena we are trying to explain. There will almost certainly exist other models that fit our data as well or better, and some of these alternatives may be sufficiently parsimonious and as meaningful as, or even more meaningful than, the original model (MacCallum, Wegener, Uchino, & Fabrigar, 1993). Of course, none of these models represents the “true model,” which doesn’t exist. Thus, we must always temper and moderate our conclusions by taking these principles into account. If such a perspective makes a researcher less enthusiastic about using SEM, then so be it. We must recognize the limitations of our techniques and not reach beyond those limitations to grasp at unjustified interpretations. (Keep in mind that these comments are coming from someone who would be considered by most methodologists to be a “believer” in SEM.)

For those of you still interested in using SEM, I offer some specific comments about technique. The first involves overall strategy. A common approach in empirical applications of SEM is to specify and evaluate a single model, or perhaps a couple of alternatives. If a model is found to fit data poorly, then an investigator might modify that model to improve its fit to the data and report such modifications along with the final model, making interpretations about the final model. I, along with many of my

methodological colleagues, have come to believe that a much better strategy is to investigate a set of alternative competing models ranging from fairly simple to relatively complex. It is quite difficult to evaluate a single model in isolation without a reference point. Specification of multiple models forces the researcher to consider a variety of alternatives, possibly representing competing theories or simply logical alternatives. Some of these models might not be expected to work well at all, but can still serve as meaningful reference points. Estimation of the alternative models yields abundant comparative information, including parameter estimates and measures of fit.

Let's consider the issue of assessment of model fit, a topic, which has received much attention in the methodological literature and is central to empirical applications of SEM. The most commonly used indexes of model fit are incremental measures such as NFI, NNFI, RNI, and CFI. These measures are based on a comparison of a given model to a null model, which specifies all measured variables as being uncorrelated in the population. I have used such measures often in the past. But I and many colleagues have developed concerns about such indexes for two reasons. First, they often seem to convey an overly positive picture of model fit, especially when the model contains latent variables each represented by several very good indicators (i.e., indicators with very high loadings on the desired latent variable and corresponding low unique variances). In such a case the quality of the measurement model can result in high values of these fit indexes even if the structural model is only mediocre. Overall, the model appears to fit well, relative to the null model, and the user is happy. The user concludes the whole model is working well, even though the structural portion of the model might be relatively poor. The second concern about incremental fit measures is that distributional properties are

unknown for nearly all such indexes, thus making it impossible to obtain confidence intervals so as to have information about precision of these estimates of model fit.

Because of these concerns, I would urge SEM users to make more use of fit indexes that are not based on a null model and for which distributional properties are known. The best such index currently available is probably RMSEA, which was first proposed nearly 20 years ago (Steiger & Lind, 1980) and has come into common usage in recent years. The availability of confidence intervals for this index aids greatly in interpretation. Some colleagues and I have developed a method for power analysis in SEM based on RMSEA (MacCallum, Browne, & Sugawara, 1996). One important result from that project was the finding that when a model has very low degrees of freedom, tests of hypotheses about model fit have very low power, and confidence intervals for RMSEA are correspondingly wide. The implication of this is that it is difficult to make precise inferences about model quality when degrees of freedom are low. In effect, it is difficult to “reject” a model with low degrees of freedom, so we should be skeptical about studies that yield evidence of support for such models.

When comparing overall fit of alternative models, researchers should be aware of the role of sample size. The degree of complexity of a model that can be supported is a function of sample size (Cudeck & Henly, 1991). With large samples we can support the estimation of parameters of more complex models, but when sample size is small we are in effect restricted to simpler models. One way to take this issue into account in assessment of model fit is through the use of the ECVI (expected cross-validation index). The ECVI estimates the degree to which a solution obtained from the sample at hand would generalize to the population. Use of the ECVI under varying levels of sample size

will show that simpler models are preferred when sample size is small, whereas more complex models can be supported when sample size is large. This index is not useful for evaluation of single models because it has no inherent reference point, but is very useful for comparison of alternative models. Confidence intervals are available for ECVI. Use of indexes such as RMSEA and ECVI can help the investigator identify the best model from among a set of alternatives, rather than attempt to determine the quality of a single model. Browne and Cudeck (1992) provide a detailed discussion of these indexes along with illustrations.

A final comment about assessment of model fit is warranted. Although the χ^2 test of overall fit continues to be given considerable weight in many empirical studies, this test is viewed by methodologists as being of little value. It tests a null hypothesis of no empirical interest (perfect fit in the population), is highly influenced by sample size, and has very low power when degrees of freedom are low. Virtually no weight should be given to this test in model evaluation.

In wrapping up my comments about SEM, I note that investigators could often take more advantage of the flexibility of this technique. For example, there is a capability for fitting models simultaneously to data from samples from distinct populations, which allows one to investigate explanations for similarities and differences among such groups. Also, whereas conventional SEM applications involve modeling the structure of covariances or correlations, it is also possible to model the structure of means; e.g., to represent means of measured variables as functions of means of latent variables. Models with structured means are especially useful in multi-group analyses to test group differences on means of latent variables, as well as in longitudinal studies

where one wishes to evaluate change in level of a selected variable over time. Millsap and Everson (1991) describe and illustrate a variety of models of this kind. SEM is particularly useful for studying change. Katzell (1994) refers to the study of change over time as a meta-trend in the I/O field, and SEM is a valuable tool for specifying and testing models of change along with investigating predictors, correlates, and consequences of change (Willett & Sayer, 1994). More generally, the development of techniques for analyzing change has been a major focus of methodological work in recent years (Collins & Horn, 1991).

Factor Analysis

Let me now turn to factor analysis, which remains a commonly used technique in I/O research for scale development, evaluation of measurement models, and studying the nature of latent variables underlying measured variables. Stone-Romero et al. (1995) report a relatively stable level of about 10% of the papers in the *JAP* using exploratory factor analysis. The application of factor analysis requires a series of choices involving methods of factor extraction, determination of the number of factors, and rotation, among other things. Unfortunately, there still exists considerable misunderstanding among users about the importance of these choices and the differences among some of the options. It is still fairly common for users of factor analysis to conduct a principal components analysis, retain components with eigenvalues greater than 1.0, and then carry out varimax rotation and interpret the resulting dimensions as latent variables. Such a procedure is easy, requires little thought, but, unfortunately, doesn't work well at all. This approach has been soundly discredited by methodologists, but its use persists in practice. Allow me to explain the basic problem. The objective of factor analysis is to identify latent

variables (common factors) that account for correlations among measured variables, with the unique portion of each measured variable being represented and estimated separately. Principal components analysis does not identify such latent variables, but rather identifies composites of measured variables that represent a mixture of common and unique effects. Components are thus different animals from common factors. They represent a mixture of that which variables have in common with each other, along with unique influences, which include random error. *Components will not account for correlations among measured variables as well as will common factors.* Further, retaining components with eigenvalues greater than 1.0 is a rule of thumb that works poorly as a method for determining the number of major common factors (Hakstian, Rogers, & Cattell, 1982; Tucker, Koopman, & Linn, 1969), often over-estimating or under-estimating the appropriate number of factors. Finally, varimax rotation restricts dimensions to being uncorrelated. In practice, there is rarely a good reason to assume that the underlying latent variables are in fact uncorrelated.

So what should one do instead? That's easy. First, extract factors by fitting the common factor model to the data. This simply involves use of a method that estimates communalities along with factor loadings, such as iterative principal factors (least squares) or maximum likelihood. Then one decides on an appropriate number of factors, or alternative numbers of factors, by considering several criteria. If using maximum likelihood factoring, one can make use of SEM-type fit measures such as RMSEA and ECVI. (See Browne and Cudeck (1992) for a nice example of using such indexes to estimate the number of factors.) Finally, it makes most sense to use oblique rotation. I would recommend direct quartimin, but almost any standard oblique rotation method

would be preferable to varimax. Oblique rotation makes the realistic allowance that factors are correlated. If an orthogonal solution is available that exhibits good simple structure, it can still be recovered using oblique rotation.

Let me now address the obvious question: Does it really matter? Won't we get essentially the same results regardless of the choices we make with respect to these methods? In 1967 Armstrong published a paper in *The American Statistician* with the subtitle, "Tom Swift and his Electric Factor Analysis Machine." Armstrong described generating a set of artificial data with known factor structure and then analyzing those data using principal components, retaining components with eigenvalues greater than 1.0, and rotating the components using varimax. The resulting solution bore virtually no resemblance to the known structure, and Armstrong argued that his results showed that factor analysis was not able to recover such structure and therefore was not useful for studying latent structure. However, if one generates data in the same fashion as did Armstrong and then fits the common factor model to those data, makes a careful decision about the number of factors, and conducts oblique rotation, one recovers the structure built into the data very clearly. (These results were first shown to me by my mentor, Ledyard Tucker, in 1970 and are included in a paper that is currently under review.) Armstrong's paper made an important point, but not the one he intended! He inadvertently provided a clear illustration of some of the effects of poor choice of technique in factor analysis. So, yes, it does matter. Fortunately, it's almost as easy to do it the right way as to do it the wrong way. Unfortunately, we have to be skeptical about the substantial array of published findings that are based on the faulty techniques just described.

Before I leave the topic of factor analysis, I should also comment on the issue of factor scores. It is not unusual in applications of factor analysis for investigators to compute factor scores and do further analyses on those scores. For instance, one might use such scores as dependent variables in ANOVA in order to test group differences on factors, or one might correlate such scores with other variables in order to evaluate relationships between the factors and other measures. When investigating questions that seem to call for such analyses, researchers should consider two issues. First, factor scores are scores on underlying latent variables and are thus indeterminate and unobservable. The common practice of computing composite scores, which are weighted or unweighted sums of those variables that load highly on a given factor, does not produce actual factor scores, nor even direct estimates of them. So such scores should be referred to as “composite scores” or “scale scores” rather than factor scores, and results of analyses of such scores apply *only* to those composites and not to the latent variables themselves.

A second point is more important. In many cases, research questions about relationships of factors to other variables, or about group differences on factors, can be addressed without computing such scores at all. Rather than view such questions in terms of a two-stage analysis involving first a factor analysis and computation of scores followed by analysis of those scores using regression or ANOVA methods, users can address such problems in a unified structural equation model. Questions involving relationships between factors and other variables can be studied using conventional SEM by specifying and fitting a model wherein latent variables are related to the other variables of interest. Questions about group differences on factors can be addressed

using multi-group SEM with structured means, as mentioned earlier. Such approaches eliminate the need to compute composite scores or other factor score estimates, and, more importantly, provide results that apply to the latent variables rather than to the estimated scores.

Measurement

Let me turn next to the issue of measurement. I think it is fair to say that the I/O research literature devotes relatively little attention to measurement problems. There is too heavy a reliance on coefficient alpha as an indicator of measurement quality and relatively little study of measurement properties of items and scales other than by factor analysis. This is unfortunate because measurement lies at the heart of much of our applied research. If we don't assure that we have good measures, then all that we do with those measures is called into question. Such a situation is unfortunate because it isn't that difficult to do better. How? The field would clearly benefit from routine use of simple traditional methods for assessing the quality of measures. These include classical test theory methods for estimation of reliability and validity, as well as simple item-level statistics such as means, item intercorrelations, and measures of difficulty and discrimination. Further benefit could be gained from the use of item response theory (IRT) (Drasgow & Hulin, 1990). Although some applied researchers may be scared off by IRT, the principles are really fairly simple. The basic concept is that an individual's response on an item is a function of that individual's true level on some underlying trait. There are alternative models specifying the nature of that function, which is represented by an "item characteristic curve" for each item showing the relationship between that item and the underlying trait. Application of the method results in estimation of

properties of items (such as difficulty and discrimination), as well as estimation of the trait score for each individual. Furthermore, although IRT was developed and is generally applied in the domain of ability testing, it is applicable in any area where the principle of item-trait relationship is applicable. For example, Roznowski (1989) provides an illustration of the use of IRT in studying a measure of job satisfaction.

Levels of Analysis

As a final major topic of comment, I wish to turn to the problem of level of analysis. This is of course a long-standing and difficult issue in the I/O field because many research questions involve individuals functioning in groups. Difficulties arise with regard to defining variables and levels at which they should be measured (e.g., climate, leadership), as well as in defining the level(s) at which the research question or theory is to be investigated. A number of important papers in recent years (Klein, Dansereau, & Hall, 1994; House, Rousseau, & Thomas-Hunt, 1995; Rousseau, 1985) have emphasized the point that most problems studied in organizational research are inherently multilevel in nature and that researchers should take this into account at all stages of research, from theory development to data collection to data analysis. Katzell (1994) identifies the study of multilevel phenomena as a meta-trend in the I/O field.

When a problem is multilevel in nature, micro or macro views will lead to misspecification of theory by ignoring the multilevel nature of the phenomena and the relevance of variables at one level to variables at another level. For example, in studying individual job performance, there are undoubtedly both individual level variables (such as motivation and ability) and group level variables (such as climate and norms) that are relevant predictors. When a theory appropriately takes into account the multilevel nature

of the problem, the gathering of appropriate data is facilitated. Units can be sampled at whatever levels are relevant (e.g., sampling organizations and sampling individuals within organizations), and variables can be measured at appropriate levels. Data can then be analyzed so as to take into account the multilevel structure of both the research questions and the data. There is a considerable methodological literature about problems caused by aggregation and disaggregation of measures, thereby ignoring the multilevel structure of the data. Such procedures can yield severely biased results as well as invalid conclusions, such as the well-known ecological fallacy wherein one uses group-level data to draw conclusions about individuals.

Important methodological developments in recent years offer a new tool for addressing some types of multilevel research questions. Methods called hierarchical linear modeling or multilevel modeling (Bryk & Raudenbush, 1992) provide a framework for specifying and fitting models to multilevel data, where units at one level (e.g., individuals) are nested within units at another level (e.g., organizations) and variables may be measured at both levels. In this framework the primary outcome variable, or dependent variable, is defined at the lowest level, usually individuals. Let's again use job performance as an example. Predictor variables may be measured at both the individual level (e.g., job satisfaction) and the group or organizational level (e.g., employee ownership of the organization). The modeling framework allows one to evaluate a variety of models involving within and between-level effects on the outcome variable. For instance, we could investigate the relationship between satisfaction and performance, and whether that relationship varies across organizations. We could determine whether variation in that relationship is predictable from the measure of

employee ownership. We could study the cross-level effect of ownership on performance. Thus, the multilevel modeling framework provides a system for specifying and testing some kinds of theories about within-level and between-level effects and thereby offers another tool for addressing some aspects of the age-old levels-of-analysis problem.

Things I Wish I Had More Space to Discuss

There are a number of additional methodological issues on which I would comment in some detail if space permitted. So I will simply offer some bullet-style notes on a few other points.

- Event history analysis: I mentioned earlier that the study of change over time has been a major focus of methodological work in psychology in recent years and that Katzell (1994) identifies such study as a meta-trend in I/O psychology. In addition to structural equation modeling and multilevel modeling, event history analysis (or survival analysis) is another technique especially useful for this purpose. Event history analysis is a technique for modeling length of time spent in various states or situations (e.g., employment). In this approach, the probability of an individual remaining in a specified state (e.g., being employed) is modeled as a function of time (the “survival function”). It is then feasible to investigate effects of individual-level variables (e.g., gender, qualifications) on the nature of this function. A discussion of this method is provided by Singer and Willett (1991), and an illustration of event history analysis of employee turnover is presented by Dickter, Roznowski, and Harrison (1996).

- Archival data: I urge I/O researchers to consider more frequent and extensive use of archival data sets. There are many such data sets available and many are of high-quality. They provide a way to test new ideas on existing data, often with large samples, while saving great amounts of time and other resources. For example, the National Longitudinal Survey of Youth (U.S. Department of Labor, 1997) is highly useful for researchers interested in work behavior. Information about some other longitudinal and archival data sets that may be useful to I/O researchers can be found in Howard and Bray (1988).
- Moderated regression: This technique has been used for many years in I/O research, but users may not be aware of a significant concern by methodologists. Moderator effects can be artifacts created by nonlinear effects of separate predictor variables (MacCallum & Mar, 1995), so users should routinely check for nonlinear effects as well. It's as easy to check for a quadratic effect as for an interaction effect in a regression model.
- Modeling multitrait-multimethod correlation matrices: The use of restricted (confirmatory) factor analysis models, specifying trait and method factors, simply doesn't work very well. In this context, such models are over-parameterized and will fit well in virtually every case, but often are subject to serious estimation problems and may provide nonsensical parameter estimates (Brannick & Spector, 1990; Coover, Craiger, & Teachout, 1997). It might be better to consider multiplicative models (Cudeck, 1988).

- Significance testing: Many of you are familiar with aspects of the debate about significance testing. Regardless of your point of view about the value of significance tests, I think it is self-evident that our empirical research literature can be enhanced by reporting confidence intervals and effect sizes, whether they are reported in place of or in addition to results of significance tests. I urge researchers to start doing so routinely, and for editors and reviewers to start insisting on such reporting. It is important for researchers to take a broader and more integrated view regarding statistical conclusion validity and to avoid a narrow focus on significance tests (Austin, Boyle, & Lualhati, in press).

Closing Remarks

Although many of my comments have involved somewhat sophisticated quantitative methods, I wish to reiterate the first major point I raised in this commentary. Researchers must let the research questions drive the selection of methods. Use methods that answer the questions, whether those methods are simple or complex. The most important aspect of our research is the significance of the questions and answers, not the complexity of the methods we use.

I close by again commending I/O researchers for their appreciation of the value of quantitative methods. From a personal perspective, such an appreciation helps to make worthwhile my own efforts at teaching methodology and attempting to bridge the gap between methodologists and substantive researchers. It also gives me confidence that my comments here will be taken seriously and may contribute in some way to further enhancement of the research enterprise in your field.

References

Armstrong, J. S. (1967). Derivation of theory by means of factor analysis or Tom Swift and his electric factor analysis machine. *The American Statistician*, *21*, 17-21.

Austin, J. T., Boyle, K., & Lualhati, J. (In press). Statistical conclusion validity in organizational research: A review. *Organizational Research Methods*.

Brannick, M. T., & Spector, P. R. (1990). Estimation problems in the block-diagonal model of the multitrait-multimethod matrix. *Applied Psychological Measurement*, *14*, 325-339.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research*, *21*, 230-258.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, *7*, 249-253.

Collins, L. M., & Horn, J. L. (Eds.). (1991). *Best methods for the analysis of change*. Washington, DC: APA.

Coovert, M. D., Craiger, J. P., & Teachout, M. S. (1997). Effectiveness of the direct product model versus confirmatory factor model for reflecting the structure of multitrait-multirater job performance data. *Journal of Applied Psychology*, *82*, 271-280.

Cudeck, R. (1988). Multiplicative models and MTMM matrices. *Journal of Educational Statistics*, *13*, 131-147.

Cudeck, R., & Henly, S. J. (1991). Model selection in covariance structures analysis and the “problem” of sample size. *Psychological Bulletin*, *109*, 512-519.

Dickter, D. N., Roznowski, M., & Harrison, D. A. (1996). Temporal tempering: An event history analysis of the process of voluntary turnover. *Journal of Applied Psychology, 81*, 705-716.

Drasgow, F., & Hulin, C. L. (1990). Item response theory. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 1, pp. 577-636). Palo Alto, CA: Consulting Psychologists Press.

Hakstian, A. R., Rogers, W. T., & Cattell, R. B. (1982). The behavior of number-of-factor rules with simulated data. *Multivariate Behavioral Research, 17*, 193-219.

House, R., Rousseau, D. M., & Thomas-Hunt, M. (1995). The meso paradigm: A framework for the integration of micro and macro organizational behavior. *Research in Organizational Behavior, 17*, 71-114.

Howard, A., & Bray, D. W. (1988). *Managerial lives in transition: Advancing age and changing times*. New York: Guilford Press.

Katzell, R. A. (1994). Contemporary meta-trends in industrial and organizational psychology. In H. Triandis, M. D. Dunnette, & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., vol. 4, pp. 1-89). Palo Alto, CA: Consulting Psychologists Press.

Klein, K. J., Dansereau, F., & Hall, R. J. (1994). Levels issues in theory development, data collection, and analysis. *Academy of Management Review, 19*, 195-229.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*, 130-149.

MacCallum, R. C., & Mar, C. M. (1995). Distinguishing between moderator and quadratic effects in multiple regression. *Psychological Bulletin, 118*, 405-421.

MacCallum, R. C., Wegener, D. R., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in covariance structure analysis. *Psychological Bulletin, 114*, 185-199.

Millsap, R. E., & Everson, H. (1991). Confirmatory measurement model comparisons using latent means. *Multivariate Behavioral Research, 26*, 479-497.

Rousseau, D. M. (1985). Issues of level in organizational research: Multi-level and cross-level perspectives. *Research in Organizational Behavior, 7*, 1-37.

Roznowski, M. (1989). Examination of the measurement properties of the Job Descriptive Index with experimental items. *Journal of Applied Psychology, 74*, 805-814.

Singer, J. D., & Willett, J. B. (1991). Modeling the days of our lives: Using survival analysis when designing and analyzing longitudinal studies of duration and the timing of events. *Psychological Bulletin, 110*, 268-290.

Steiger, J. H., & Lind, J. M. (1980). *Statistically based tests for the number of common factors*. Paper presented at the Annual Meeting of the Psychometric Society, Iowa City, Iowa.

Stone-Romero, E. F., Weaver, A. E., & Glenar, J. L. (1995). Trends in research design and data analytic strategies in organizational research. *Journal of Management, 21*, 141-157.

Tucker, L. R., Koopman, R. F., & Linn, R. L. (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika*, 34, 421-459.

U. S. Department of Labor. (1997). *NLS Handbook*. Washington, DC: Bureau of Labor Statistics.

Willett, J. B., & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin*, 116, 363-381.