11 Teaching Students of Educational Psychology New Sophisticated Statistical Techniques

Bengt Muthen
Graduate School of Education
University of California, Los Angeles

BACKGROUND

The methodological training of graduate students in the area of Educational Psychology poses an exceptional challenge. While these students frequently have little background in mathematics and statistics, many of the statistical methods most appropriate for the field are technically complex. This is true both for students whose main interests and specialization lie in a substantive area, and also, perhaps even more so, for those students who specialize in research methods. While the substantive group struggles with basic statistics, the often better quantitative background of the methods group is frequently still a poor match for the sharply increased level of difficulty of the methods they are expected to master. The data analysis related research in the Educational Psychology area poses statistical problems as challenging as they come. Although many methodologically advanced solutions have been proposed in recent years, even the methods students are hard pressed to understand them well enough to make good use of them.

The statistical complexities in Educational Psychology research stem from the fact that data are frequently collected in a nonexperimental setting, often in a hierarchical, multilevel fashion (such as students observed within schools), and almost always involving constructs that are difficult to measure in a valid and reliable way. Prime examples of advanced, statistical answers to these data complexities include maximum-likelihood structural equation modeling (see, Joreskog, 1977) which attempts to simultaneously handle uncontrolled background differences and measurement unreliability, regression analysis with a variance component structure (see, Aitkin & Longford, 1986; Burstein, 1985) to take multilevel measurements into account, and item response theory analysis with a full information, three-parameter logistic approach (see Bock & Aitkin, 1981; Hambleton & Swaminathan, 1984) to properly describe the responses to dichotomously scored achievement items.

When thinking about emerging new research methods that future students of educational psychology will need to learn, the above set of advanced techniques naturally come to mind. However, there are other methods that seem important and go beyond the usually covered areas such as experimental design, multiple regression, measurement theory, and standard multivariate techniques. For instance, there have been several useful developments in regression analysis regarding diagnostic measures such as leverage and influence (see e.g., Atkinson [1985] and Cook & Weisberg [1982]). Many recent developments have been made that extend standard analysis techniques to the frequently encountered situations of categorical and other nonnormal data, such as loglinear modeling (see Bishop, Fienberg, Holland, 1975), extensions of loglinear modeling to ordinal variables (see Agresti, 1984 and references therein to work by Goodman), extensions to factor analysis (see Mislevy, 1986), and more general nonnormal data analyses (see McCullagh & Nelder, 1983). Multidimensional scaling and cluster analysis would seem to be useful emerging data analysis techniques and are described in Schiffman, Reynolds, and Young (1981). Traditional as well as more recent missing data techniques are treated in Little and Rubin (1987). Meta analysis techniques for combining information from several studies are treated in Hedges and Olkin (1985). While many important methods naturally are omitted from the above list, it should give a feel for the great amount of statistical material that is potentially relevant for an educational psychology student's data analysis.

This chapter uses structural equation modeling to discuss the general issues involved in preparing Educational Psychology researchers to properly use advanced statistical methods.

THE METHODS TOPIC

Structural equation modeling is a general term for a set of techniques that cover path analysis, confirmatory factor analysis, and general latent variable models that combine features of path and factor analysis. An attempt is made to describe the methodological issues in as nontechnical a language as possible.

In path analysis a set of linear regression relations are assumed to describe a set of correlated variables. For instance, educational aspiration may be regressed on a set of background variables, while at the same time it may be used as a predictor in another regression, e.g., using educational attainment as the dependent variable. By formulating this system of two regression relations, the researcher may attempt to separate the direct influence of background on attainment from the indirect influence on attainment via aspiration. Statistical tests may be performed to check whether all influence is indirect.

In confirmatory factor analysis, an attempt is made to go beyond the scope of ordinary exploratory factor analysis to not only cluster variables that measure the same constructs, but also test specific hypothesis on the measurement relations between the observed variables and the latent factors. For instance, using a measurement instrument in different populations, it may be of great interest to study whether the measurement relations are the same, and if they are, how the distributions of the common factors differ.

In general structural equation modeling, the above two features are special cases which can be combined in one powerful model that addresses measurement concerns and construct relationships simultaneously. Simple path analysis where all variables are observed can be done by a sequence of ordinary linear multiple regressions and has been in practice for about 20 years. Regular exploratory factor analysis can be performed by standard computer packages and has been in practice even longer. However, the more complex modeling, particularly involving hypothesis testing and latent variables, requires specialized software and did not come in general practice until the popularization of the Joreskog's LISREL program (Joreskog & Sorbom, 1984) about 10 years ago.

THE PROBLEM

Today, general structural equation analyses using LISREL are common in Ed Psych journals such as *Journal of Educational Psychology*, *Journal of Personality and Social Psychology*, *American Educational Research Journal*, *Journal of Educational Measurement*, *Sociology of Education*, *Applied Psychological Measurement*, and *Psychological Bulletin*. Grant applications have been turned down if structural equation modeling has not been mentioned among solutions to measurement concerns. Structural equation modeling has become a hot topic among social and behavioral science data analysts, and the fact that it is statistically complex only serves to make it the more alluring. At this point, however, I think we are facing a serious problem which is not confined to the use of structural equation modeling, but is a general one for all new sophisticated methods. There is a multitude of very poor applications of the methodology, including the simpler path analyses. Many research claims are based on flawed or completely erroneous analyses. If this volume of bad applications is not greatly reduced, there is, in my opinion, a strong risk that large portions of research results will not be believable. I think there is a growing credibility problem, which must be stemmed since many important research studies will continue to call for statistical tools of this kind. A few prototypical examples of kinds of poor applications may be of interest. In general, I find that the major problem lies in the transition from substantive, conceptual ideas to statistical analysis—there is a difficulty in moving from conceptual to statistical modeling. The common problem is that measurement issues and statistical assumptions that are incidental to the researchers' conceptual ideas become stumbling blocks that invalidate the statistical analysis.

One frequent misuse involves the omission of important explanatory variables ("x variables") in a certain regression relation, be it in path analysis, structural modeling, or factor analysis (where the factor assumes the role of x). It is common to see studies report regressions on a certain set of x variables of particular interest in the study without imbedding these among other x variables belonging in the equation but not necessarily of prime interest. Perhaps this stems from being trained with experimental research situations, where ANOVA on a randomized sample allows one to exclusively concentrate on the *manipulated* factors. However, with regression in a nonexperimental setting, this use of a minimal set of x's is likely to cause severe bias in the estimated regression slopes, which has been called omitted variable bias.

A frequently occurring problem is associated with the increasing use of "confirmatory" analysis, i.e. using a chi-square test of model fit to assess the appropriateness of a certain theory as specified in terms of a covariance structure. Here, the common mistake is to consider estimates and significance of parameters from models that do not fit the data, i.e., use "predictions" from a model that has little to do with the data at hand. A not uncommon version is to compare a sequence of models built from various competing theories, to discuss differences in estimates, and *choose* the model that has the best chi-square—even when each of the chi-square values indicate a strong rejection of the model at hand. Again, perhaps the misuse stems from being used to ANOVA, where one usually does not worry about the "fit of a model" (indeed ANOVA is perhaps not even thought of as a model) but merely wants to look at the "effects."

More complex misuses arise with latent variable modeling, where the researcher may not fully realize that the questionnaire format used or the particular phenomenon intended to be measured causes complications for the latent variable modeling to be carried out in a standard structural modeling framework. The indicator variables may be nonlinearly related to the latent variable; they may, by the question format, have certain direct dependencies; and measurement errors may be likely to be correlated with the latent variable and have strongly heteroscedastic variances. These examples show that structural equation modeling can be a very complex topic, which to be well mastered takes years of hard study.

Part of the methodological/statistical community is becoming skeptical about these methods. For instance, the Summer 1987 issue of *Journal of Educational Statistics* is devoted to the discussion of the strengths and weaknesses of path analysis, where the seminal article is by the statistician David Freedman who is strongly critical of path analysis usage due to violated statistical assumptions. As a discussant in a recent conference on test validity (see Wainer & Braun, 1987), the statistician Don Rubin voiced his concern over bad structural equation modeling removing the analyst to far from his or her data. Many of us interested in psychometrics are concerned. Cliff (1983) stated:

> Initially, these methods seemed a great boon to social science research, but there is some danger that they may instead become a disaster, a disaster because they seem to encourage one to suspend his normal critical faculties. Somehow the use of one of these computer procedures lends an air of unchallengeable sanctity to conclusions that would otherwise be subjected to the most intense scrutiny. These methods have greatly increased the rigor with which one can analyze his correlational data, and they solve many major statistical problems that have plagued this kind of data. However, they solve a much smaller proportion of the interpretational—inferential in the broader sense—problems that go with such data. These interpretational problems are particularly severe in those increasingly common cases where the investigator wishes to make causal interpretations of his analyses.

Five or 10 years ago, poor published applications of structural equation modeling could be excused as demonstration pieces, promoting the method itself. But the average quality has not risen as much as it should since then. One explanation may be related to the publication pressure, where researchers are not allowed to present analysis failures, but where published models must fit the data well at any cost, and where complex analyses are favored over more mundane ones. Such an atmosphere does not stimulate good applications. Some psychometricians emphasize the improvement of the statistical methods to better fit real data, and development of new computer software which is technically less demanding, giving more time to consider the basic analysis problem.

In my own view, however, the largest part of the problem and the largest part of its solution does not lie in the domain of publishing or in methodology development, but in education. Presently, there is not enough done in the education of the ultimate users of these new statistical techniques for them to learn the methodological part of their research trade well. I do not think there is a real problem with the methodology and I am by no means ready to throw out the baby with the bath water. The methods can surely be improved, and that is important. We methodologists should probably also be much more careful of not overselling our new developments. However, the most important way to change the quality of applications is to put more emphasis on training students of the topic to learn it well.

METHODOLOGICAL TRAINING

Like other areas of advanced statistical methods for the social sciences, structural equation modeling has a group of contributors of theory and methods. Hence, recently many new powerful techniques have been proposed in psychometric journals such as *Psychometrika* and *British Journal of Mathematical Psychology*.

As is also the case with item response theory, for example, the group of theoretical structural modelers is rather small, concentrated in a subgroup of *Psychometric Society*, which itself meets with only about 200 members. The ratio of providers to consumer is presumably very small, and this poses a difficult educational problem. There is a distinct lack of people who can bridge the gap between the theory provided and the intelligent use of these methods in practice.

I believe that this problem has caused a pressure for people who are not well trained themselves to assume the role of *bridgers*, teaching and advising those even less knowledgeable in less than-optimal ways. And since these people may not always have conveyed a message of methodological rigor, many purely substantive researchers may have felt that it was all right for them too to dabble on their own with the advanced methods. At least that is the impression I am often left with from reading applied journal articles and listening to talks at professional conferences. In my opinion, these methods—or computer programs, as they are often viewed—should not be taken as "everyman's causal modeling" tools.

It is time to take this problem seriously and to consciously educate students to assume various methodological roles. Not everyone should be using these methods. Most people should not use them without intimate guidance and involvement of a truly knowledgeable person (a *bridger* or a theoretical expert). Not everyone should attempt to become an expert.

For the purposes of discussing methodological training, it may be useful to distinguish between three types of Ed Psych students: those who emphasize substantive interest, those who emphasize methodological interest but do not aspire to contribute to methodology, and those who place a strong emphasis on methodology and have aspirations to in some way enhance the methodology.

The first group of substantive students will ultimately constitute a major portion of "the users" of a given methodology, here structural equation modeling. In my view, these students need only an overview of the potential of the techniques, explained in a largely conceptual way. The major message should be to seek intimate cooperation with an expert if a need for structural modeling arises. These students should not be encouraged to *wing it* on their own and should be discouraged from seeking automated sofware solutions. I believe that a large portion of today's poor state of applications is due to the enticement of such students to be self sufficient. While this is laudable and advisable regarding more straightforward, standard statistical techniques, it is not a healthy attitude regarding a new, sophisticated technique such as structural modeling, a topic which it takes years to master well.

Both the second and third group of students, choosing a methodological emphasis, are very much needed in the Ed Psych area. They can fill the void of people at various levels of bridgers and in various degrees make connections between the theorists and the users. They have the potential know-how to move beyond poor and mediocre applications. With a good and intensive methods training they have the distinct advantage over pure statisticians of having a good understanding of their substantive area. They should be trained in making the transition from conceptual, substantive *modeling* to statistical analysis. On the other hand, they should not be expected to become statisticians.

The methodological curriculum should make a distinction between the second and third group. The goal of the second group may suitably be to have a strong grasp of the methods, being able to understand advanced applications, e.g., in publications, and be able to use the methods in their own substantive research with only minor assistance from more expert colleagues. We might call this group low-level bridgers. With this goal, the students certainly need to be quantitatively adept, but need not take special courses in mathematics and statistics. They need courses in regression analysis, ANOVA, and multivariate statistics. The structural equation modeling topic may be studied in two ten week courses, covering factor analysis, path analysis with observed variables, and general structural equation modeling. Some technical detail is needed. However, the emphasis should lie on sound use of the methods, and that can be taught well without taking a large portion of time for statistical theory. This group may also cover other areas of advanced methods, such as item response theory, and so become general methodologists.

The third group, the most methodologically oriented, constitutes a relatively small group that can be trained as applied statisticians or psychometricians with an Ed Psych specialization. These can serve as high-level bridgers. They need considerable technical training, and to assimilate that, need to have previous or parallel mathematical and statistical training in areas such as calculus, matrix algebra, and mathematical statistical theory on an undergraduate introductory level. This knowledge is best achieved by a Master's degree in applied statistics. A degree of this type is strongly recommended for high level bridgers.

Following applied Ed Psych training in areas of regression analysis, experimental design, and multivariate statistics, the study of structural equation modeling for high-level bridgers necessitates a sequence of three ten week courses covering general factor analysis, path analyis, and advanced structural equation modeling. In addition, independent studies with the students working on their own data, followed up by specialized seminars would be needed. In my view, nothing less prepares the student for penetrating the topic to a degree that makes him or her serve in the role as advanced bridger. This person must be able to read and communicate to users the latest advanced developments as presented in the original sources, and must come beyond the point of initial fascination with the methods and their technical aspects to know both strengths and weaknesses well. This will admittedly leave relatively little room for penetrating studies of other topics and hence the group will be a highly specialized one. However, I believe that this is necessary to do the training well—or else, the aim should be a different one.

The last group will be a small, but very important one. Few Ed Psych students are suited for this training. The present rarity of candidates for mastering advanced methodology is certainly a problem worthy of attention. Suitable candidates, for example, are students with a Master's degree in Statistics or applied Mathematics. There is a challenging recruitment effort involved in conveying that they are needed in the Ed Psych methods arena.

CONCLUSION

Using the case of structural equation modeling, the general problem of teaching advanced statistical methods to students in Educational Psychology has been outlined. Due to a lack of attention to the difficulty level of structural modeling, the training of past students has been inadequate. This is having serious ramifications, in that a large number of poor applications of the method has been appearing in scientific journals. Research conclusions have been based on flawed analyses. The importance of breaking a trend towards such analyses not being taken seriously was pointed out. The future need of researchers who can bridge the gap between the advanced statistical methods and sound applications was stressed. The production of such bridgers can only be accomplished by stronger emphasis on thorough methods training. For high level bridgers this requires a strong degree of methods specialization.

ACKNOWLEDGMENT

This work has benefitted from discussions with Leigh Burstein and Linda K. Muthen.

REFERENCES

Aitkin, M. A., & Longford, N. (1986). Statistical modeling issues in school effectiveness studies. *Journal of the Royal Statistical Society, Series A*, 149, 1–35.

Agresti, A. (1984). *Analysis of ordinal categorical data*. New York: Wiley.

Atkinson, A. C. (1985). *Plots, transformations and regression*. London: Oxford University Press.

Bishop, Y. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: Theory and practice*. Cambridge, MA.: MIT Press.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*, 443–459.

Burstein, L. (1985). Units of analysis. *International Encyclopedia of Education*. London, England: Pergamon Press, 5368–5375. Also reprinted in International Encyclopedia of Teaching and Teacher Education (forthcoming).

Cliff, N. (1983). Some cautions concerning the application of causal modeling methods. *Multivariate Behavioral Research, 18*, 115–126.

Cook, R. D., & Weisberg, S. (1982). Residuals and influence in regression. New York: Chapman and Hall.

Hambleton, R. K. & Swaminathan, H. (1984). Item response theory. *Principles and applications.* Boston: Kluwer-Nijhoff.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* New York: Academic Press.

Joreskog, K. G. (1977). Structural equation models in the social sciences: Specification, estimation and testing. In P. R. Krishnaiah (Ed.), *Applications of statistics.* Amsterdam: North-Holland.

Joreskog, K. G., & Sorbom, D. (1984). *LISREL VI Analysis of linear structural relationships by the method of maximum likelihood.* User's guide. Mooresville, IN: Scientific Software Inc.

Little, R. & Rubin, D. (1987). *Statistical analysis with missing data.* New York: Wiley.

McCullagh, P., & Nelder, J. A. (1983). *Generalized linear models.* London: Chapman & Hall.

Mislevy, R. J. (1986). *Recent developments* in the factor analysis of categorical variables. *Journal of Educational Statistics, 11,* 3–31.

Schiffman, S. S., Reynolds, M. L., & Young, F. W. (1981). *Introduction to multidimensional scaling. Theory, methods and Applications.* New York: Academic Press.

Wainer, H. & Braun, H. (1987). *Test validity.* Hillsdale, NJ: Lawrence Erlbaum Associates.