The Impact of Measurement Error:
A Study of Four Correction Approaches

By Kimberly S. Maier, Qiu Wang, Chueh-an Hsieh, Changhui Zhang,
Chun-lung Lee, David Reyes-Gastelum, and Jonghwan Lee

Michigan State University

Background:

Measurement error is ubiquitous in educational research. Many of the variables of interest in educational research are measured imperfectly since these constructs cannot be measured directly, but rather by the use of indicators such as the students' responses of test items or survey questions. In order to create measures of these constructs, analytical techniques or paradigms are used, such as item response theory (IRT). Using modern test theory IRT, a measure can be constructed for the trait that is a function of the respondent's answers to the items of an instrument. The error in measurement occurs when the estimate differs from the value being estimated using the instrument data (Hansen, Hurwitz, Marks, and Mauldin, 1951). This is considered a non-sampling error. As with any measurement error, the difference between the true latent trait and the measure may lead to biased estimates of interest such as the average value in the population and can also lead to an underestimation of the total variance of the statistics of interest (Särndal, Swensson and Wretman, 1993).

Most recommendations to treat measurement error involve adjustment methods. Approaches to the adjustment method to treat measurement errors include the moment method (Fuller, 1987; Cheng and Van Ness, 1992, 1994, 1997) and the simulation method (Carroll et al, 2006). The moment method uses the reliability coefficient to adjust the biased estimators back toward their true values (Cheng and Van Ness, 1994). Simulation methods use the Monte Carlo sampling simulation extrapolation method (SIMEX, e.g., Caroll, et al 2006) to reduce the bias due to measurement error.

Purpose:

The purpose of this study is to quantify the amount of measurement error introduced by IRT estimates used as the independent variable in a logistic regression model when different approaches to treating measurement error are used. The IRT estimates have a standard error which is a function of both the latent trait and the number of items in the instrument (Embretson and Reise, 2000). We considered the efficacy of four different approaches to the treatment of measurement error when the latent trait is estimated using a two-parameter logistic model. Two of the approaches might be considered more common approaches to treating measurement error in the modeling context described above. The remaining approaches are more novel approaches.

Method:

This study used data from the 500 Family Study, conducted by the Alfred P. Sloan Center on Parents, Children, and Work at the University of Chicago. The 500 Family Study contains rich, detailed information on over 500 middle-class, dual-earner families in eight communities across the U.S (Schneider and Waite, 2005). The Study contains families with either adolescents, young children, or both and collected data from participants using surveys, interviews, and time diaries. For this study we have included only those families with adolescents who had participated in all study components and had at least 15 entries in their time diaries. These selection criteria resulted in a final sample of 216 adolescents (81.4 percent of the adolescents in the original sample).

For this analysis, data were obtained from the adolescent surveys. The survey included a number of items used in other national studies of adolescents, including extensive information on adolescents' home and school experiences. The measures used in the models included aspirations and parental strategic support.

The aspiration variable used was an item asking how far the student thought they would go in school. It was originally coded as an eight category variable ranging from less than high school education to graduate school. Given the high aspirations of this sample of adolescents (over 89% indicated that they would at least graduate from a four year college), this variable was recoded as a dichotomous variable indicating whether or not the student aspired to attend graduate school.

The Parental Strategic Support (PSS) measure was constructed from adolescents' responses on the student survey. This measure is considered to be a latent variable that captures facets of familial support and challenge offered by the parent as perceived by the adolescent, and was created from adolescents' responses to a series of statements asking them about various supportive or challenging aspects of their family environment. The thirteen items included in the Parental Strategic Support measure inquired about aspects of the responsiveness of the family to the adolescent such as providing appreciation, attention, love, and acceptance and aspects of the home environment that encourage the responsible and autonomous development of the adolescent. To construct the Parent Strategic Support (PSS) measure, a two-parameter logistic item response model was used.

Data was simulated for the study based on the parameters estimated from the real data. Using the item parameter estimates and the estimated correlation between the outcome variable (graduate school attendance) and the PSS measure, dichotomous item responses were simulated. Based on these parameter estimates, a complete data set was simulated for sample sizes of 500 (reflecting a moderate sample size) and 1500 (reflecting a large sample size).

Four different approaches to measurement error were examined for each of the two sample sizes. The model parameters of each of these four approaches were estimated using the Bayesian estimation framework; analyses were implemented in BUGS (Bayesian inference Using Gibbs Sampling; Lunn, Thomas, Best, and Spiegelhalter, 2000) software. The first two approaches constitute two-stage approaches to model estimation. The first approach (A1) used the known item parameters and initially estimated the latent variable measures. This was followed by estimation of the logistic model parameters using the latent measures as fixed and known. The second approach (A2) first estimated the latent trait measures for each person along with the item parameters. Then the parameters of the logistic regression model were estimated, again treating the latent measures as fixed and known. The third and fourth approaches constitute one-stage approaches to model estimation. The third approach (A3) combined the estimation of the IRT item parameters, the latent variable measures, and the parameters of the logistic regression in one model, constituting a one-stage estimation approach. The fourth (A4) and final approach combines the third approach plus adjustment to estimators using the estimated reliability of the PSS instrument. The first two approaches were considered to be the more traditional approaches to measurement error treatment; the final two are more novel approaches.

Findings:

We found that in general, approach A3 performed better than both approaches A1 and A2. This implies that a one-stage approach can effectively reduce the impact of measurement error on estimation. However, approach A4 outperformed approach A3. Approach A4 provided the most accurate estimates, suggesting that the use of the reliability coefficient provides additional correction in the presence of measurement error. We also found that the estimates for the larger sample size were better than those for the smaller sample size, not a surprising finding. However, it was encouraging that approaches A3 and A4 were still able to produce estimates that were not seriously biased for the smaller sample size.

Conclusions:

Regardless of the analytical technique or paradigm used to construct the measure, latent constructs are often not precisely estimated, and produce relatively large measurement error when introduced into models as independent variables. Four different analytic approaches to the treatment of this measurement error were explored using simulated data. Comparisons of the results from each indicate that the use of a one-stage approach to model estimation is the most efficacious. It is important to note that this study was very limited in the scope of models and sample sizes that were considered. However, the results indicate a promising avenue of research in the treatment of measurement error when independent variables are resultant measures of a 2PL IRT model. Future research includes the consideration of more complex IRT models such as those for polytomous data, a more complicated logistic model, as well as a more fine-grained exploration of the impact of sample size.

References

Carroll, Raymond J., Ruppert, David, Stefanski, Leonard A., and Crainiceanu, Ciprian M. (2006). *Measurement error in nonlinear models: A modern perspective* (2nd ed.). New York: Chapman & Hall.

Cheng, C-L. and Van Ness, J. W. (1992). Generalized M-estimators for errors-in-variables regression. *The Annals of Statistics, 20*(1), 385-397.

Cheng, C-L. and Van Ness, J. W. (1994). On estimating the linear relationships when both variables are subject to errors. *Journal of the Royal Statistical Society Series B, 56*(1), 167-183.

Cheng, C-L. and Van Ness, J. W. (1997). Robust calibration. *Technometrics, 39*(4), 401-411.

Embretson, Susan E. and Reise, Steven P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.

Fuller, Wayne A. (1987). *Measurement Error Models.* Wiley.

Hansen, M.H., Hurwitz, W. N., Marks E. S., and Mauldin, W. P. (1951). Response errors in surveys. *Journal of the American Statistical Association, 46*, 147-90.

Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. (2000) WinBUGS -- a

Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, **10**:325--337.

Särndal, C.E., Swensson, B., and Wretman, J. (1992). *Model assisted survey sampling*. New York: Springer-Verlag.

Schneider, B. & Waite, L.J. (2005). *Being Together, Working Apart: Dual-Career Families and the Work-Life Balance.* New York: Cambridge University Press.