EDITOR'S COMMENTS

An Update and Extension to SEM Guidelines for Administrative and Social Science Research

By: **David Gefen**
Professor of Management Information Systems
Drexel University
gefend@drexel.edu

Edward E. Rigdon
Professor of Marketing
Georgia State University
erigdon@gsu.edu

Detmar Straub
Editor-in-Chief , *MIS Quarterly*
Professor of CIS
Georgia State University
dstraub@gsu.edu

Appendix A: Modeling Interaction Effects in CBSEM

Among the several approaches proposed for estimating interactions in latent variable models in CBSEM, Bollen and Paxton (1998) and Little et al. (2006) describe procedures that do not require special-purpose software, and so are accessible to all CBSEM users. Bollen and Paxton described an approach using two-stage least squares (2SLS), an analytical approach used in regression to overcome estimation problems which would confound OLS. 2SLS associates each predictor in a regression model with an instrumental variable. In the first stage, each predictor is regressed on its instrument, then in the second stage the ultimate dependent variable is regressed on the expected or predicted portion of each predictor from the first stage.

In this case, the researcher wants to estimate an interaction model involving latent variables, using only observed variables, each of which is contaminated with error. Under assumptions, two indicators x_1 and x_2, reflecting the same latent variable share variance due only to the latent variable, while their error terms are mutually uncorrelated. If x_1 is regressed on x_2, then $\hat{x}_1$, the portion of x_1 explained by x_2, can be used as an error-free substitute for the latent variable. Thus, Bollen and Paxton's 2SLS approach involves 2SLS estimation with the ultimate dependent variable regressed on indicators of the main effect latents and the interaction latent variable, using other indicators of those latents as instruments. Klein and Moosbrugger (2000) demonstrated that the 2SLS approach is not as statistically efficient as more recent methods. Still, this approach can be used to estimate latent variable interaction models using most any standard statistics package.

Little et al. (2006) proposed an "orthogonalizing" strategy to overcome the problem of collinearity between main effects and the interaction term. First, create indicators of the interaction term by multiplying indicators of the main effect constructs. Then regress each product indicator on the main effect indicators used to form each product. The residuals from these regressions are retained as indicators of an interaction latent variable, which is completely orthogonal to the main effect latent variables. Marsh et al. (2007) showed that researchers using this approach could dispense with special and difficult parameter constraints, allowing estimation of the continuous latent variable interaction model using any standard CBSEM package.

References

Bollen, K. A., and Paxton, P. 1998. "Interactions of Latent Variables in Structural Equation Models," *Structured Equation Modeling* (5:3), pp. 267-293.

Klein, A., and Moosbrugger, H. 2000. "Maximum Likelihood Estimation of Latent Interaction Effects with the LMS Method," *Psychometrika* (65:4), pp. 457-474.

Little, T. D., Bovaird, J. A., and Widaman, K. F. 2006. "On the Merits of Orthogonalizing Power and Product Terms: Implications for Modeling Interactions Among Latent Variables," *Structural Equation Modeling* (13:4), pp. 497-519.

Marsh, H. W., Wen, Z., Hau, K.-T., Little, T. D., Bovaird, J. A., and Widaman, K. F. 2007. "Unconstrained Structural Equation Models of Latent Interactions: Contrasting Residual- and Mean-Centered Approaches," *Structural Equation Modeling* (14:4), pp. 570-580.

Appendix B: Recommended Rigor When Using SEM

SEM Quality Assurance Steps that Should Be Taken

Statistical rigor is crucial in SEM, as it is in linear regression. The following recommendations apply.

Number of Observed Variables/Measures Needed

How many measurement/indicators there should be for each reflective latent variable is still an open debate in the literature. With PLS, the bias in parameter estimates is an inverse function of the number of observed variables per construct (Dijkstra 2010; McDonald 1996). With CBSEM, the impact depends on whether measurement models are formative or reflective. With formative scales (Diamantopoulos and Winklhofer 2001) the main issue is completeness. Are all components of a hypothetical formatively measured construct represented in the data? With reflective, factor-analytic measurement models, multiple observed variables per construct contribute strongly to the model's degrees of freedom, which is a primary driver of statistical power (MacCallum et al. 1996). Three observed variables loading exclusively on one common factor make the individual factor measurement models statistically identified. More than three observed variables make each model over-identified, increasing the researcher's ability to detect lack of fit. Encompassing the measurement models of multiple correlated constructs within a single confirmatory factor model or imposing suitable constraints on parameter estimates will allow a model to achieve identification with fewer observed variables per construct, but with a more limited ability to detect poorly performing measures.

A design that includes only a few measures per construct also puts a study at risk: if one or two of a small number of measures behave in unexpected ways, researchers may be forced to choose methods based on their analytical feasibility rather than based on their ability to best address the research questions that motivated the study.

Common Method Bias

When developing data sources for a study, researchers must be wary of common method bias (CMB). CMB can inflate estimates of structural parameters in the model and may result in erroneous conclusions (Podsakoff et al. 2003). Addressing CMB is not an integral part of SEM and so auxiliary analysis needs to be carried out to assess it, although the cross-loadings CMB creates will be evident in CBSEM through observed lack of fit.

Several steps can be taken when designing the data collection to reduce CMB. One is collecting data with different methods or sources or at different points in time (Podsakoff et al. 2003). These different sources may include multiple informants (Kumar et al. 1993), corporate databases, transactional data, financial statements, or other independent sources.

CMB can also be addressed to some degree in questionnaires by including reverse-scored items to reduce acquiescence, and through statistical analysis such as Harman's single-factor test (Podsakoff et al. 2003). In this method, all of the items are examined in an exploratory factor analysis and the researchers verify that in the unrotated solution there is more than one dominant factor. A better way to test for CMB is to add a latent variable to the model, representing conceptually the shared variance of all the measurement items, and have all the measurement items load also onto it in addition to the constructs onto which they theoretically load (Podsakoff et al. 2003). This method uses the ability of PLS and CBSEM, in contrast to linear regression, to model a loading of a measurement item on more than one latent construct.[1]

Unidimensionality

Another crucial aspect of SEM that is almost always ignored (Malhotra et al. 2006) is distributional assumption checking (especially unidimensionality). Unidmensionality means that there is only one theoretically and statistically underlying factor in all of the measurement items associated with the same latent construct. Inadequate unidimensionality challenges reflective construct validity at its most fundamental level and results in parameter estimate biases (Anderson et al. 1987; Gefen 2003; Gerbing and Anderson 1988; Segars 1997). Researchers who design a scale to reflect multiple latent dimensions face special challenges. The simple proportionality test described previously is not directly applied, and standard tools like reliability indices also will not directly apply. Unidimensionality is assumed in linear regression and PLS, and can be tested for in CBSEM. See an example of how to do so and the consequences of ignoring it in the tutorial on unidimensionality by Gefen (2003). Like CBSEM, PLS path modeling as a method also has the ability to incorporate multidimensional constructs, although few researchers seem to exploit this opportunity.

Alternative Models

Another issue that has been around for a long time but is still not applied as rigorously as it should be is the comparison of the theoretical model with a saturated model. In a seminal paper, Anderson and Gerbing (1988) highlighted the need to compare alternative models, including the theoretical model containing the hypotheses with a saturated model containing paths among all pairs of latent variables, even those assumed to be unrelated in the model. In essence, this allows the researchers to verify that no significant path has been left out of the model. This is crucial because omitting significant predictors can bias other path estimates. In some cases, these omissions bias estimated paths leading to significance. While in CBSEM omitted paths may reveal themselves through lack of fit, with PLS and linear regression the consequences of these omissions may only be seen in careful diagnostics involving plots of residuals. It should be imperative, therefore, that authors include such analysis in their papers.

Possible additional checks related to alternative models include testing for moderating effects. In CBSEM this can be done using χ^2 nested model tests (Jöreskog and Sörbom 1994). Other tests include testing for a significant change in χ^2 when comparing the original model with an alternative nested model in which one pair of the latent variables are joined (Gefen et al. 2003). If the resulting difference in χ^2 is insignificant, then, at least statistically, there is no reason not to join the two latent constructs. If the model χ^2 fit index is not significantly improved by joining any two latent variables, then there is reason to believe that there is significant discriminant validity.

Sample Size and Power

Sample size has long been an important issue in PLS path modeling and in CBSEM, but regression has the most straightforward sample size guidance. Ordinary least squares, the most popular estimation method in regression, does not demand especially large sample sizes, so the choice of sample size is primarily a matter of statistical power (Cohen 1988). Statistical power in regression is a well-understood function of effect size (f^2), sample size, number of predictors and significance level. Cohen (1988) offers detailed guidance on how to use these variables to choose a minimum necessary sample size for regression users aiming to achieve a given level of statistical power. Across the social sciences, convention specifies 80 percent as the minimum acceptable power. Websites are available to calculate the minimum sample sizes needed to achieve adequate power under different conditions in linear regression (e.g., http://www.danielsoper.com/statcalc/).

With PLS and CBSEM, sample size plays a more complex role. PLS path modeling parameter estimates are biased, as noted previously, with the bias diminishing as both the number of indicators per construct and sample size increase. Researchers can calculate the expected degree of bias and determine the likely impact of investing in a larger sample size (Dijkstra 2010; McDonald 1996). It has been argued that PLS path modeling has advantages over OLS regression in terms of power (Chin et al. 2003) although this claim has been challenged (Goodhue et al. 2007). The core of the PLS estimation method—ordinary least squares—is remarkably stable even at low sample sizes. This gave rise to a

[1]However, not all authorities endorse *ex post* techniques for dealing with common method variance (Richardson et al. 2009; Spector 2006).

rule of thumb specifying minimum sample size as 10 times the largest number of predictors for any dependent variable in the model (Barclay et al. 1995; Gefen et al. 2000). This is only a rule of thumb, however, and is not backed up with substantive research.

With CBSEM, the role of sample size is also complex, but in different ways. Statistical power itself takes on multiple meanings. On the one hand, there is power to reject the researcher's overall model. In contrast to regression, PLS path modeling, and most statistical methods in the social sciences, with CBSEM the model being proposed by the researcher is (typically, but not necessarily) equated with the null hypothesis, rather than being associated with the alternative hypothesis. In regression and PLS path modeling, however, power typically refers to the ability to reject a null hypothesis of no effect in favor of an alternative hypothesis which is identified with the researcher's proposed model. So, with CBSEM, increasing sample size means constructing a more stringent test for the model, while with regression and PLS path modeling, a large sample size may work in the researcher's favor. Power versus the overall model is related to the model degrees of freedom (MacCallum et al. 1996). A model with high degrees of freedom can obtain sufficient power at very low sample size. However, power can also be addressed in terms of individual path estimates, as in regression and PLS path modeling (Satorra and Saris 1985). In this setting, researchers may find a link between power and sample size that is more consistent with that observed in the other methods.

Additionally and entirely apart from the issue of power, CBSEM users must be concerned about stability of the estimation method. CBSEM long relied on maximum likelihood (ML) estimation, a method that achieves stable results only at larger sample sizes. Other estimation methods, designed to overcome ML shortcomings or violations of distributional assumptions, may require even larger sample sizes.

There is no guideline backed by substantial research for choosing sample size in CBSEM. At the absolute minimum, there should be more than one observation per free parameter in the model. Sample size should also be sufficient to obtain necessary statistical power. Experience argues for a practical minimum of 200 observations for a moderately complicated structural equation model with ML estimation. When other estimation methods, such as WLS, are used or when there are special features, such as mixture modeling and models with categorical latent variables, a larger sample may be needed.[2] Rules of thumb alone do not automatically tell us what the sample size for a given study should be.

Other Overall Issues of Statistical Rigor that Apply Also to SEM

Data Description

Researchers should be expected to share enough information with their readers about their empirical data that their readers can make informed judgments about the soundness of the choices made and conclusions reached by the researchers. Other researchers should find enough information in a research report to allow them to make comparisons to their own results, or to include the study as a data point in a meta-analysis. Inferior reporting limits contribution to knowledge and potential impact.

Certain reporting requirements are common across regression, PLS path modeling, and CBSEM. Sharing this basic summary information allows researchers to determine whether differences in results across studies may be due to differences at the univariate level. For example, restriction of range, which may be revealed in the standard deviations, can attenuate correlations in a particular sample (Cohen et al. 2003, p. 57). If the reader does not have the raw data available, then researchers have an even greater obligation to fully examine their data and report thoroughly. Researchers must demonstrate, for example, that data appear to be consistent with stated distributional assumptions. With regression and PLS path modeling, which both use ordinary least squares estimation, explicit distributional assumptions during estimation focus on error terms, not on the observed variables themselves, although violated assumptions such as multicolinearity will bias the estimation results. Still, skewed distributions can reduce correlations between variables, and this has direct implications for the results. Extreme violations of distributions assumptions or skewed data may suggest the need for transformation to achieve better fit, such as using a log transformation as is customary in linear regression in such cases (Neter et al. 1990), although obviously the results need to be interpreted accordingly because the slope of the relationship between Y and ln(X) is not the same as the slope of the relationship between Y and X.

---

[2]Muthén and Muthén (2002) recommend and illustrate the use of Monte Carlo simulations as a tool to help determine necessary sample size in a particular situation. Stability improves when path relationships are strong and when there are several indicators per construct. Researchers should use literature related to their particular choice of estimation method to support an argument regarding adequate sample size. For example, simulation research regarding the combination of polychoric correlation coefficients and weighted least squares estimation point to the need for sample sizes in the thousands (Yung and Bentler 1994). Even then, researchers must be alert for evidence of instability. This evidence may include unreasonable χ^2 values, parameter estimates, or estimated standard errors.

Researchers should always report means and standard deviations for all continuous observed variables, and proportions for all categorical variables, as a most basic requirement. Readers of studies that use CBSEM have additionally come to expect authors to provide the correlation or covariance matrix of the observed variables. In many cases, having these matrixes enables readers to literally reconstruct the analysis, allowing for fuller understanding; it also makes the report useful in the training of future researchers. These matrixes are also valuable when researchers use regression or PLS path modeling, although those methods rely more on the raw data. We strongly encourage researcher to add the correlation or covariance matrix as an appendix to their paper.

Missing Data

One issue too often ignored is proper handling of incomplete data. Partially missing observations are a fact of life. Unfortunately, some popular analytical programs still promote outdated procedures which at best are statistically inefficient and at worst bias results. When there are missing data and nothing can be done to rectify or simulate the missing data, we recommend the authors say so explicitly. Realizing that some methodologists do not approve of deleting data, researchers can choose and have chosen in the past listwise deletion. Discarding an entire case when only few observations are missing does not bias results under many circumstances, but it is wasteful of data as compared with currently available alternatives. The alternative pairwise deletion—estimating, say, each covariance based on all observations where both variables are present—conserves data but may, as in linear regression (Afifi et al. 2004), result in inaccurate estimations and introduce inconsistencies because of varying sample sizes within the data. Varying sample size across the model also complicates the analysis.

There may be cases, however, when there are not enough data to allow listwise deletion or there is a theoretical reason not to listwise delete these data points, such as when those data come from a unique category of respondents about whom these measurement items are irrelevant. In such cases, researchers can apply the EM algorithm with multiple imputation (Enders 2006), an approach which usually produces a complete data rectangle with minimal bias in the results. Standard statistical packages now include this approach, bringing it within reach of almost all researchers. Some statistical packages offer more sophisticated approaches such as the full information maximum likelihood approach, which is or will soon be standard in CBSEM packages. This approach models the missing data points along with the free parameters of the structural equation model. With this approach, estimation proceeds as if each missing observation is an additional free parameter to be estimated, although, unlike imputation, the method does not generate an actual set of replacement values. The modeling of missing values is specific to the larger model being estimated, so it would be inappropriate to establish specific replacement values as if they were model-independent. Techniques are available which can address missing data issues even in cases where missingness is tied to the phenomena under study under certain conditions (Enders 2011; Muthén et al. forthcoming).

Distribution Assumptions

With CBSEM, data distribution becomes an important factor in selecting the estimation method. The default maximum likelihood method assumes conditional multivariate normality (Muthén 1987), although the method is also robust against mild violations of this assumption (Bollen 1989). Therefore, researchers should report standard skewness and kurtosis measures for their observed variables. Substantial departures from conditional multivariate normality call for either transformation or the use of a different estimator.

Nonresponse Bias

Another topic related to missing data is nonresponse bias. This issue arises when more than a relatively slight proportion of the respondents did not return the survey or complete the experimental treatment. In these cases it cannot be ruled out that there is a selection bias among those who did complete the survey or treatment. There is nothing that can be done *ex post* to recreate these missing data within the data itself, but steps should be taken to verify at least at a minimal level that the nonresponders are not a unique group.[3] This can be done partially by comparing the demographics of respondents and nonrespondents and by comparing the answers of waves of responders; presumably, if these are no significant differences among these groupings, then there is less reason for concern (Armstrong and Overton 1977).

[3]Although other steps can be taken such as follow up phone interviews to see if there is a difference between the respondents and nonrespondents. Also see Sivo et al. (2006).

Population: Appropriate and Homogeneous

As with any research project involving collected data, researchers must carefully describe and justify their choices of population, sampling frame, and sampling plan. Researchers must support the appropriateness of the population and address concerns about possible heterogeneity. Heterogeneity means that different members of the population are subject to either different models or different parameter values within the same model structure. In such cases, parameters estimated across the aggregated population are at best overall averages, which may not accurately represent even one single individual respondent, much less all of them. In some cases, the sampling frame itself presents a risk of heterogeneity, as when a researcher aggregates respondents across nations or cultures in order to reach a sufficiently large total sample size.

Currently, linear regression, PLS, and CBSEM (Rigdon et al. 2010) all offer tools for comparing across populations and for seeking multiple latent populations within a single data set. Researchers should evaluate their populations for the potential for heterogeneity. Where heterogeneity represents a plausible confound, researchers should investigate and report their comparative results.

References

Afifi, A. A., Clark, V., and May, S. 2004. *Computer-Aided Multivariate Analysis, Volume 62*, Boca Raton, FL: CRC Press.

Anderson, J. C., and Gerbing, D. W. 1988. "Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach," *Psychological Bulletin* (103:3), pp. 411-423.

Anderson, J. C., Gerbing, D. W., and Hunter, J. E. 1987. "On the Assessment of Unidimensional Measurement: Internal and External Consistency, and Overall Consistency Criteria," *Journal of Marketing Research* (24), pp. 432-437.

Armstrong, J .S., and Overton, T. S. 1977. " Estimating Non-Response Bias in Mail Surveys," *Journal of Marketing Research* (14), pp. 396-402.

Barclay, D., Thompson, R., and Higgins, C. 1995. "The Partial Least Squares (PLS) Approach to Causal Modeling: Personal Computer Adoption and Use an Illustration," *Technology Studies* (2:2), pp. 285-309.

Bollen, K. A. 1989. *Structural Equations with Latent Variables*, New York: Wiley-Interscience

Chin, W. W., Marcolin, B. L., and Newsted, P. R. 2003. "A Partial Least Squares Latent Variable Modeling Approach for Measuring Interaction Effects: Results from a Monte Carlo Simulation Study and an Electronic-Mail Emotion/Adoption Study," *Information Systems Research* (14:2), pp. 189-217.

Cohen, J. B. 1988. *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.), Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J. B., Cohen, P., West, S. G., and Aiken, L. S. 2003. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3rd ed.), Mahwah, NJ: Lawrence Erlbaum Associates.

Diamantopoulos, A., and Winklhofer, H. M. 2001. "Index Construction with Formative Indicators: An Alternative to Scale Development," *Journal of Marketing Research* (38:2), pp. 269-277.

Dijkstra, T. K. 2010. "Latent Variables and Indices: Herman Wold's Basic Design and Partial Least Squares," in *Handbook of Partial Least Squares,* V. E. Vinzi, W. W. Chin, J. Henseler, and H. Wang (eds.), Berlin, Germany: Springer-Verlag, pp. 23-46.

Enders, C. K. 2006. "Analyzing Structural Equation Models with Missing Data," in *Structural Equation Modeling: A Second Course,* G. R. Hancock and R. O. Mueller (eds.), Charlotte, NC: Information Age, pp. 315-344.

Enders, C. K. 2011. "Missing Not at Random Models for Latent Growth Curve Analysis," *Psychological Methods* (16:1), pp. 1-16.

Gefen, D. 2003. "Unidimensional Validity: An Explanation and Example," *Communications of the Association for Information Systems* (12:2), pp. 23-47.

Gefen, D., Karahanna, E., and Straub, D. W. 2003. "Trust and TAM in Online Shopping: An Integrated Model," *MIS Quarterly* (27:1), pp. 51-90.

Gefen, D., Straub, D. W., and Boudreau, M.-C. 2000. "Structural Equation Modeling and Regression: Guidelines for Research Practice," *Communications of the Association for Information Systems* (4:7), pp. 1-70.

Gerbing, D. W., and Anderson, J. C. 1988. "An Updated Paradigm for Scale Development Incorporating Unidimensionality and its Assessment," *Journal of Marketing Research* (25:May), pp. 186-192.

Goodhue, D., Lewis, W., and Thompson, R. 2007. "Statistical Power in Analyzing Interaction Effects: Questioning the Advantage of PLS with Product Indicators," *Information Systems Research* (18:2), pp. 211-227.

Jöreskog, K. G., and Sörbom, D. 1994. *LISREL 8.12i and PRELIS 2.12i for Windows*, Chicago: Scientific Software International.

Kumar, N., Stern, L. W., and Anderson, J. C. 1993. "Conducting Interorganizational Research Using Key Informants," *Academy of Management Journal* (36:6), pp. 1633-1651.

MacCallum, R. C., Browne, M. W., and Sugawara, H. M. 1996. "Power Analysis and Determination of Sample Size in Covariance Structure Analysis," *Psychological Methods* (1:2), pp. 130-149.

Malhotra, N. K., Kim, S. S., and Patil, A. 2006. "Common Method Variance in IS Research: A Comparison of Alternative Approaches and a Reanalysis of Past Research," *Management Science* (52:12), pp. 1865-1883.

McDonald, R. P. 1996. "Path Analysis with Composite Variables," *Multivariate Behavioral Research* (31:2), pp. 39-70.

Muthén, B. O. 1987. *LISCOMP: Analysis of Linear Structural Relations with a Comprehensive Measurement Model* (2nd ed.), Mooresville, IN: Scientific Software.

Muthén, B. O., Asparouhov, T., Hunter, A., and Leuchter, A. forthcoming, "Growth Modeling with Non-Ignorable Dropout: Alternative Analyses of the STAR*D Antidepressant Trial," *Psychological Methods* (http://www.statmodel.com/download/FebruaryVersionWide.pdf).

Muthén, L. K., and Muthén, B. O. 2002. "How to Use a Monte Carlo Study to Decide on Sample Size and Determine Power," *Structural Equation Modeling* (9:4), pp. 599-620.

Neter, J., Wasserman, W., and Kutner, M. H. 1990. *Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Design* (3rd ed.), New York: McGraw-Hill, Inc.

Podsakoff, P. M., Lee, J. Y., and Podsakoff, N. P. 2003. "Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies," *Journal of Applied Psychology* (88:5), pp. 879-903.

Richardson, H. A., Simmering, M. J., and Sturman, M. C. 2009. "A Tale of Three Perspectives: Examining Post Hoc Statistical Techniques for Detection and Correction of Common Method Bias," *Organizational Research Methods* (12:4), pp. 762-800.

Rigdon, E. E., Ringle, C. M., and Sarstedt, M. 2010. "Structural Modeling of Heterogeneous Data with Partial Least Squares," *Review of Marketing Research* (7), pp. 257-298.

Satorra, A., and Saris, W. E. 1985. "The Power of the Likelihood Ratio Test in Covariance Structure Analysis," *Psychometrika* (50), pp. 83-90.

Segars, A. H. 1997. "Assessing the Unidimensionality of Measurement: A Paradigm and Illustration Within the Context of Information Systems Research," *Omega* (25:1), pp. 107-121.

Sivo, S. A., Saunders, C., Chang, Q., and Jiang, J. J. 2006. "How Low Should You Go? Low Response Rates and the Validity of Inference in IS Questionnaire Research," *Journal of the AIS* (7:6), pp. 361-414.

Spector, P. E. 2006. "Method Bias in Organizational Research," *Organizational Research Methods* (9:2), pp. 221-232.

Yung, Y.-F., and Bentler, P. M. 1994. "Bootstrap-Corrected ADF Test Statistics in Covariance Structure Analysis," *British Journal of Mathematical and Statistical Psychology* (47), pp. 63-84.