Paper 11381-2016

Multiple-Group Multiple Imputation: An Empirical Comparison of Parameter Recovery in the Presence of Missing Data using SAS®

Patricia Rodríguez de Gil and Jeffrey D. Kromrey
University of South Florida

ABSTRACT

It is well documented in the literature the impact of missing data: data loss (Kim & Curry, 1977) and consequently loss of power (Raaijmakers, 1999), and biased estimates (Roth, Switzer, & Switzer, 1999). If left untreated, missing data are a threat to the validity of inferences made from research results. However, the application of a missing data treatment does not prevent researchers from reaching spurious conclusions. In addition to considering the research situation at hand and the type of missing data present, another important factor that researchers should also consider when selecting and implementing a missing data treatment is the structure of the data. In the context of educational research, multiple-group structured data are not uncommon. Assessment data gathered from distinct subgroups of students according to relevant variables (e.g., socio-economic status and demographics) might not be independent. Thus, when comparing the test performance of subgroups of students in the presence of missing data, it is important to preserve any underlying group effect by applying separate multiple imputation with multiple groups before any data analysis. Using Likert-type data from the Civics Education Study (1999), this simulation study evaluates the performance of multiple-group multiple imputation and total-group multiple imputation in terms of item parameter invariance within Structural Equation Modeling and Item Response Theory.

Keywords: Measurement invariance, structural equation modeling, item response theory, PROC CALIS, PROC IRT, MSTRUCT model, polytomous data

INTRODUCTION

A fundamental concern in statistical analysis is the validity of inferences made from groups' comparisons; that is, ensuring measurement invariance or the comparability of measures from different groups. When the assessment of measurement invariance involves latent variables (variables not observed directly), multiple-group structural equation modeling (MG-SEM) and item response theory (IRT) are appropriate methodologies for evaluating the invariance property in factors and items across groups (See Jöreskog, 1971; Muthén, 1984; and sörbom, 1974 for early work on SEM analysis with multiple-group data, and Takane & de Leew, 1987; and Wirth and Edwards, 2007, for work on the relationship between FA and IRT). Data from behavioral and social sciences are often ordered categorical data generated from Likert-type items (Christoffersson, 1975; Muthén, 1978; Shi & Lee, 2000). Both MG-SEM and IRT can handle this type of data. However, missing data are a pervasive problem and before conducting any MG-SEM or IRT analysis, missing data should be treated. In the context of multiple-group structured data, it is relevant that the application of any missing data treatment (MDT) preserves any underlying group effect. Thus, the main goal of this study is to compare the effectiveness of Multiple Imputation (MI) for preserving group differences by applying this method to the total sample and to each group separately. Complete data analysis and Listwise deletion are also applied for comparison. The effect of these MDT on parameter recovery was examined by estimating unknown structural parameters and unknown parameter thresholds using PROC CALIS and PROC IRT.

SEM APPROACH TO MODEL SPECIFICATION, PARAMETER IDENTIFICATION, AND ESTIMATION WITH ORDINAL SCALES

Basic to the implementation of SEM is the specification of the relationship between latent and observed variables based on theory. Figure 1 shows Schulz's (2004) two-factor model specification (how the latent and observed variables are interrelated to identify the model that explains the data). This single-group confirmatory factor model has two factors (CON and SOC). Single-headed arrows represent the **functional** relationship between the factors and their corresponding observed variables while doubleheaded arrows represent factors's and observed variables' error variance. Unlabeled arrows are free parameters.

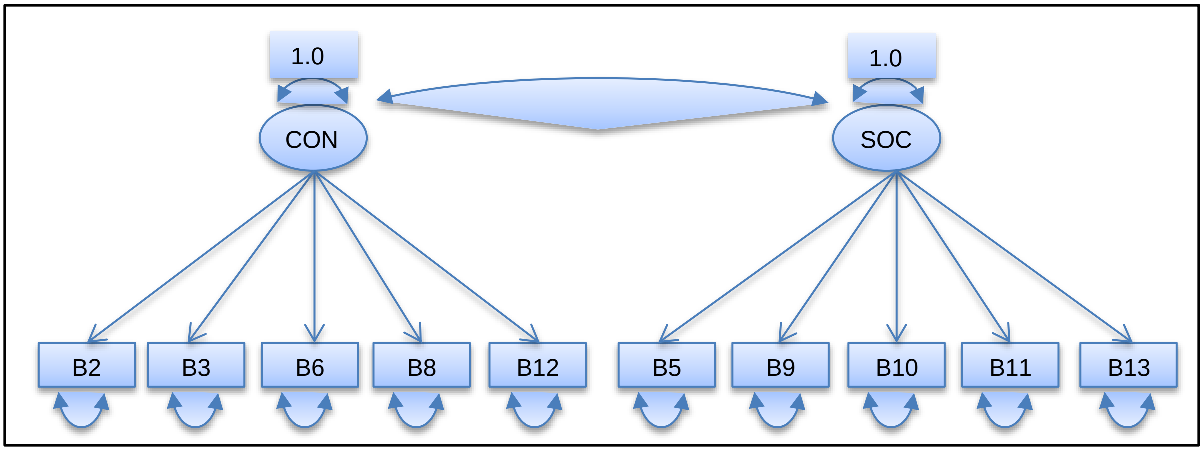

Figure1. Path Diagram for the Confirmatory Factor Model for the Scale B Data

The functional or measurement model model in Figure 1 was specified using the FACTOR modeling language in PROC CALIS using the following statements:

```
ods graphics on; ods html style=statistical sge=on;
proc calis data=samples_b plot=pathdiagram;
   factor
      CON ===> B2 B3 B6 B8 B12,
      SOC ===> B5 B9 B10 B11 B13;
   pvar
     CON = 1.,
     SOC = 1.;
run;
ods _all_ close; ods listing;
```

Each factor is specified in each entry of the FACTOR statement and is followed by the list of observed variables functionally related to the factor. Figure 2 shows the path diagram created by enabling ODS graphics (the sge = on option allows the editing of the diagram using the graphics editor).

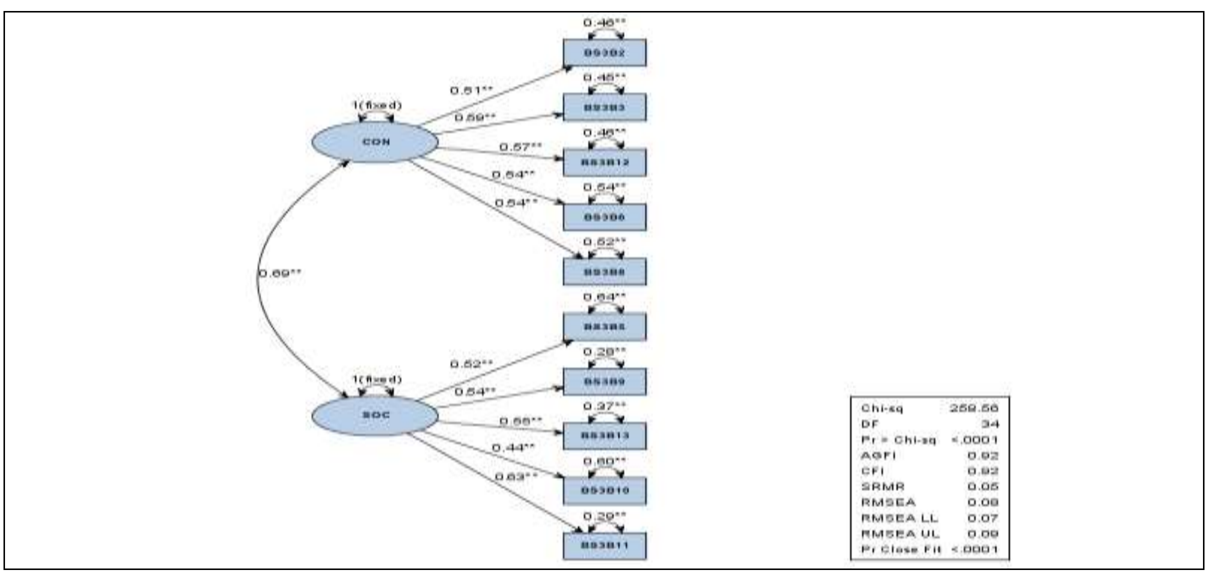

Figure 2. Confirmatory Factor Model with Correlated Factors (Unstandardized Solution)

In this confirmatory two-factor (Figure 2), showing an ideal process-flow pattern, the factor "conventional" (CON) is loading on items regarding the desirability of being politically active in conventional forms of participation while the factor "social movement" (SOC) is loading on items indicating the desirability of having an active political participation with new forms of social movements. The model identification was established by setting the factors' variance to1. Doing so, the scale of measurement has also been stablished. The fit statistics in the Figure 2 inset show that the model has an acceptable fit ($\chi^2_{(34)}$ = 258.56, $p < .0001$). Significant estimates at α =.01 are flagged by asterisks.

However, in many research situations, the primary interest is the between-group variation or differential performance by groups. For such situations, MG-SEM is an appropriate methodology for evaluating whether the constructs represented by the model factors are comparably measured across groups, which requires the evaluation of measurement invariance. If it holds, then differences between groups can be meaningfully interpreted. The confirmatory model in Figure 1 is a measurement or functional model that models the relationship between the observed measures and the factors. For evaluating invariance, it is needed to include means and covariances in the model (i.e., MACS models).

Estimating the means and covariances simultaneously is requested by adding the MEANSTR option to the previous SAS code:

```
proc calis data=samples_b meanstr plot=pathdiagram;
```

Output 1 shows the modeling information. The MEANSTR option in the PRO CALIS statement requests the mean structure analysis, in addition to the default covariance structure analysis:

```
                          The CALIS Procedure
         Mean and Covariance Structures: Model and Initial Values

                         Modeling Information
                      Maximum Likelihood Estimation

                     Data Set            WORK.SAMPLES_B
                     N Records Read      1000
                     N Records Used      1000
                     N Obs               1000
                     Model Type          FACTOR
                     Analysis            Means and Covariance
                           Variables in the Model

        Variables     BS3B2   BS3B3   BS3B5   BS3B6   BS3B8   BS3B9   BS3B10
                            BS3B11   BS3B12   BS3B13
                             Factors      CON   SOC

                          Number of Variables = 10
                          Number of Factors   = 2
```

Output 1. Model Information of Saturated Means and Covariance Structure Model

However, our interest is now in applying the mean and covariance structures analysis to groups. In the methods section, the SAS routines for preparing the covariance matrix as permanent dataset for input by groups are provided along the settings for conducting the multiple group analysis for measuring invariance using equality constrains in mean and covariance structures.

As stated before, data are gathered not only from continuous measurements. There are many research contexts in which item-level indicators generate ordered categorical response data (i.e., not important, somewhat unimportant, somewhat important, and very important). Item Response Theory (IRT) provides methods for modeling latent variables.

ITEM RESPONSE THEORY APPROACH

In attitude assessment, polytomous, Likert-type items with m ordered response categories require examinees to indicate, for example, the degree of intensity with which they agree or disagree with statements. This degree of intensity is underlined by the kth item category, whose label (e.g., Strongly Disagree, Disagree, Agree, Strongly Agree) not only underlines the degree of intensity of the response but also the direction or order of the categories. Samejima's model for graded responses, the graded response model (GRM; Samejima, 1969, 2010) is an appropriate model for this type of items (Steinberg, 2001). The next section introduces this model and provides an overview of its assumptions.

GRADED RESPONSE MODEL (GRM)

Samejima's (1969, 2010) graded response model is estimated,

$$P(k) = \frac{1}{1+\exp[-a(\theta - b_{k-1})]} - \frac{1}{1+\exp[-a(\theta - b_k)]} = P^*(k) - P^*(k+1).$$

where

a = discrimination parameter or the slope of the option response function quantifying the relationship of the item to the latent variable

b_{k-1} = the k^{th} threshold parameter representing the response endorsement, or the point in in the scale at which the probability of responding to category k passes .5

$P^*(k)$ = probability of selecting category *k*

The following PROC IRT statements request the scale B items to be fitted by the GRM:

```
ods graphics on; ods html style=statistical sge=on;
proc irt data=samples_b pinitial itemstat polychoric
   plots = (icc polychoric);
   var BS3B2 BS3B3 BS3B6 BS3B8 BS3B12 BS3B5 BS3B9 BS3B10 BS3B11 BS3B13;
   model BS3B2 BS3B3 BS3B6 BS3B8 BS3B12 BS3B5 BS3B9 BS3B10 BS3B11 BS3B13
          /resfunc=graded;
run;
ods _all_ close; ods listing;
```

The specification in the model statement fits the items using the GRM. The combined Table 1 shows item statistics and Eigenvalues:

		Item-Total Correlations		Eigenvalues of the Polychoric Correlation Matrix			
Item	Mean	Unadjusted	Adjusted	Eigenvalue	Difference	Proportion	Cumulative
BS3B2	2.134	0.610	0.498	**4.524**	3.238	0.452	0.452
BS3B3	1.349	0.644	0.533	**1.287**	0.517	0.129	0.581
BS3B6	1.933	0.611	0.489	0.770	0.098	0.077	0.658
BS3B8	1.699	0.611	0.491	0.672	0.013	0.067	0.725
BS3B12	1.584	0.689	0.588	0.658	0.090	0.066	0.791
BS3B5	1.899	0.614	0.486	0.568	0.081	0.057	0.848
BS3B9	2.309	0.650	0.559	0.488	0.070	0.049	0.897
BS3B10	2.065	0.586	0.463	0.418	0.061	0.042	0.939
BS3B11	2.165	0.693	0.601	0.357	0.098	0.036	0.974
BS3B13	2.220	0.602	0.492	0.259		0.026	1.000

Table 1. Item statistics and Eigen values of the Polychoric Correlation Matrix. The two bolded eigenvalues have values greater than 1, which reasonably suggests a two-factor model

ODS GRAPHICS is enabled so the item characteristic curves plots are displayed . Figure 3 shows the ICC plots for selected items:

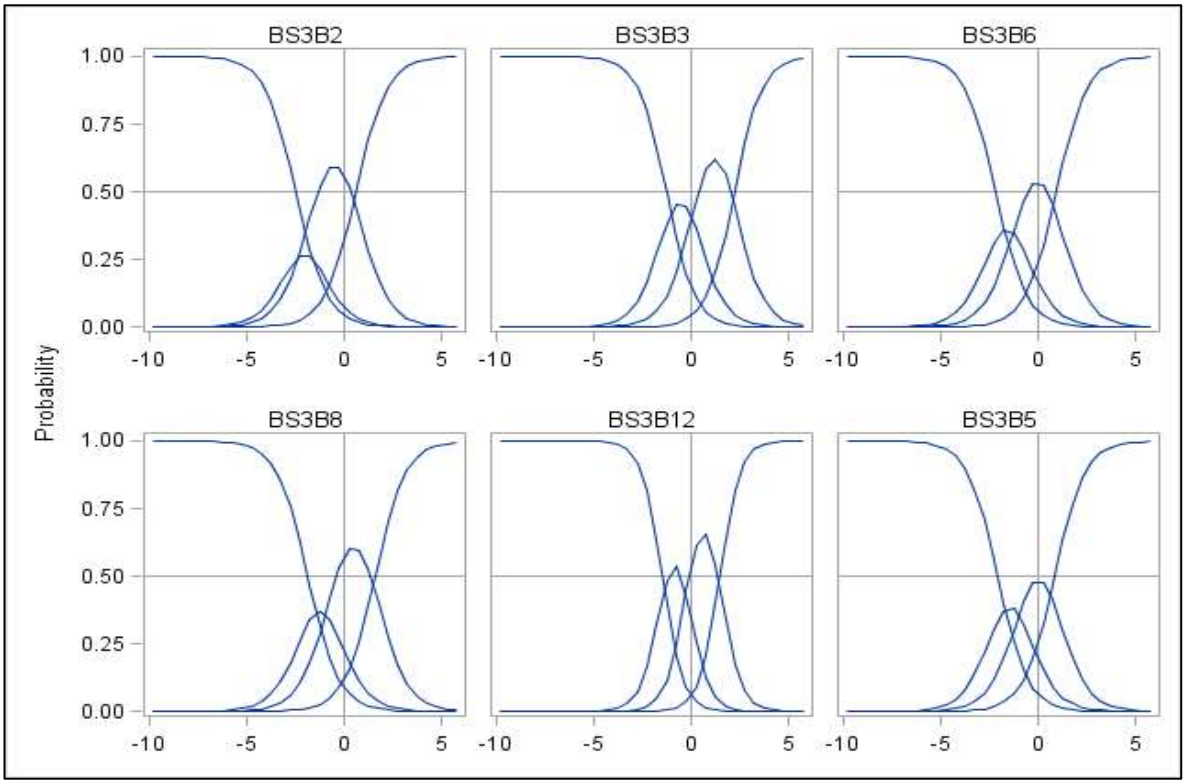

Figure3. Item Characteristic Curve (ICC)

By default, item plots are displayed in panels. Individual ICC can be displayed using the UNPACK suboption in the PLOTS statement .In the Methods section is shown how to use PROC IRT to set up a Multiple-Group analysis.

Missing data in the context of structural equation modeling and item response theory are very common and data analysts should decide how to proceed in the presence of missing data.

ANALYSIS WITH INCOMPLETE DATA

The effect of missing data on structural Equation models has received previous attention (Muthen, Kaplan, & Hollis, 1987). However, because the properties of the observed data and the missing data might be different, when applying a MDT, it is important to consider the underlying process or mechanism that causes missingness. The incorporation of the missingness mechanism into the statistical model can lead to efficient data modeling. Rubin’s (1976) theoretical framework is the most widely accepted for missing data research.

RUBIN’S MISSING DATA FRAMEWORK

Missing data in applied research are ubiquitous and it is not the exception in structural equation modeling and IRT, in the context of educational research. In educational testing, missing data can occur when examinees do not respond to an item (i.e., item nonresponse). Item nonresponse generates patterns of missing data that can be a threat to the validity of inferences made from test scores if left untreated. Rubin’s (1976) and Little and Rubin’s (1987) classification of missing data according to the relationship of variables with the probability of missing data provide the most accepted theoretical framework for the study of missing data.

Let Y = (Y1, Y2,…,Yp) be a random vector variable for a p-dimensional multivariate data (y x p matrix) in which Yobs and Ymiss are the observed and missing data elements of the Y random vector. R is a missingness dummy response indicator; that is, r =1 if Ymiss and r = 0 if Yobs. Rubin's missing data framework then provides the following assumptions about the mechanisms by which data are missing, whether random or not:

MISSING COMPLETELY AT RANDOM (MCAR)

Data are missing completely at random (MCAR) if the probability of missing data Z=1 depends neither on Yobs nor on Ymiss . That is, the reason for missingness in a given variable is not related to the variable itself or to any other variable, (R | X, Y=R),

$$P(R = 1 \mid Y_{\text{obs}}, Y_{\text{miss}}) = P(R = 1)$$

When missing data are MCAR, missingness on a variable is "unsystematic" (Baraldi & Enders, 2010; p. 7), unrelated to examinees' responses to other variables; thus, Yobs can be thought of as a random subsample of the population and can used for inferences on Ymiss (Baraldi & Enders, 2010; Tsikriktsis, 2005).

MISSING AT RANDOM (MAR)

Data are missing at random (MAR) if the probability of missing data on a given variable Yi depends on Yobs, not on Ymiss. That is, the probability of data missing at random depends on data that are observed in other variables but not on the data that are not observed for a given variable (Schafer & Graham, 2001).

$$P(R = 1 \mid Y_{\text{obs}}, Y_{\text{miss}}) = P(R = 1 \mid Y_{\text{obs}})$$

Tsikriktsis (2005) explains that MAR means that "the probability of a missing value on some variables is independent of the respondents' true status on that variable" (p. 55). Thus, examinees having missing data differ from examinees with complete data only by chance. Rubin (1976) considers MAR and MCAR to be related to the data, thus ignorable for likelihood based inferences.

MISSING NOT AT RANDOM (MNAR)

Data are missing not at random (MNAR) if the probability of Ymiss depends on Yi. When the process or mechanism that generates missingness is not random, there is a relationship between the variable(s) with missing data and examinees with data on the variable. See Schafer and Graham (2002) for examples of MAR, MCAR, and MNAR missing data mechanisms in the context of blood pressure data.

In addition to considering the types of missing data (MAR, MCAR, and MNAR), researchers need to consider the effect that a given MDT can have on the research outcomes. An extensive number of MDT has been developed (see Baraldi & Enders, 2010; Schafer & Graham, 2002 for introductions to missing data analysis) and include deletion, imputation (single and multiple), and model-based procedures. The next section offers a brief description of the MDT that were implemented in this study.

MISSING DATA METHODS

Listwise Deletion (LD).

Also called complete-case analysis, this MDT omits observations that have missing values. It is the default option in most statistical software packages (e.g., SAS, SPSS). Because of the potential of large amounts of data being discarded, this method's estimates are very conservative, as a result of loss in statistical power.

Multiple Imputation (MI)

For the multiple imputation (MI) procedure, missing values are replaced with a number of plausible values determined by the researcher; this means that several multiple data sets with identical observed values across data sets are generated, but the imputed values vary. This variability in missing data allows the MI procedure to estimate the uncertainty and relationships when imputing missing values. Due to its excellent properties, MI is one of the most accepted methods for the treatment of missing data (See excellent reviews of this method by Graham & Hofer, 2000; Schafer & Graham, 2000; Schafer & Olsen, 1998).

However, in most applications MI is applied to the total dataset. Considering the relevance of preserving group characteristics when conducting multiple group analyses (Enders & Gottschall, 2011), the main goal of this study was to compare the effectiveness of MI under total group imputation and separate group imputation. Because missing data are best studied within a specific research context, this study evaluated the performance of MI in the context of educational research, using ordered categorical data from an attitudinal assessment (affective domain) within the Structural Equation Modeling (SEM) and Item Response Theory (IRT) framework, using PROC CALIS and PROC IRT. Complete data and Listwise deletion were also applied for comparison purposes.

METHOD

This study examined two approaches to the implementation of Multiple Imputation (MI): total group imputation and separate group imputation. Additionally, complete data and Listwise deletion were also implemented for comparison. Factors of the study were sample size (500, 1000), proportion of missing observations (.10, .20, .30), and proportion of missing items (.10, .20, .30). Using SAS 9.4, replicated samples were bootstrapped from the scale B of the IEA Civics Education (CIVED) study of 1999 (N=1725), US sample. Replicated sampling was implemented using PROC SURVEY SELECT to select multiple stratified samples:

```
proc sort data=CIVED.student_b; by GENDER; run;

proc surveyselect data=CIVED.student_b
   out=samples_b seed=12345
   method=URS N=250 reps = 1000;
   strata GENDER;
run;
```

This sampling frame was stratified by gender and included 1000 replications or samples. Each stratified sample (N=250) was selected by sequential random sampling with equal probability within STRATA. That is, N determined the stratum sample size, selected in the order the strata appears in the sorted dataset. Thus, G1=250 and G2=250, where 1=female and 2=male. The method=URS, or unrestricted random sampling option allowed observations to be selected with replacement. These probability samples were stored in the samples_b output dataset. This same process was followed for *N*=500.

DATA DESCRIPTION AND PRELIMINARY ANALYSIS

The CiVED study was conducted in the US in 1999 to assess 9th grade students' concepts in the domain of citizenship. The subscale B of this test consisted on 10 non-cognitive Likert-type items. Using the MSTRUCT model specification in PROC CALIS, mean and covariance estimates were obtained simultaneously:

```
proc calis data=samples_b pcorr;
   mstruct;
     by BSGGEND;
   var BS3B2 BS3B3 BS3B5 BS3B6 BS3B8 BS3B9 BS3B10 BS3B11 BS3B12 BS3B13;
run;
```

Multiple Group analyses in PROC CALIS require the input of data as covariance matrices. Table 2 and Table 3 present basic statistics by gender (mean and SD) and the fitted covariance matrix for the 10 non-cognitive items from the subscale B, measuring students' concepts of citizenship.

	Girls (*n*=250)		**Boys (*n*=250)**		**Covariance Matrix**									
Item	*M*	*SD*	*M*	*SD*	1	2	3	4	5	6	7	8	9	10
B2	3.14	0.82	3.30	0.79	--	.20	.25	.19	.23	.18	.22	.24	.21	.17
B3	2.39	0.90	2.46	0.91	.39	--	.22	.22	.31	.18	.27	.22	.39	.14
B6	2.98	0.92	2.94	0.95	.10	.21	--	.19	.25	.24	.11	.31	.20	.28
B8	2.85	0.94	3.05	0.92	.21	.28	.22	---	.35	.22	.25	.27	.39	.24
B12	2.73	0.86	2.81	0.89	.21	.17	.14	.37	--	.25	.19	.25	.34	.13
B5	3.42	0.69	3.26	0.77	.12	.17	.19	.18	.19	--	.18	.37	.14	.35
B9	3.18	0.81	3.14	0.89	.16	.18	.11	.21	.23	.22	--	.24	.31	.22
B10	3.29	0.74	3.16	0.83	.12	.17	.24	.13	.14	.26	.17	--	.28	.41
B11	2.57	0.84	2.62	0.89	.14	.25	.22	.15	.22	.12	.21	.28	--	.18
B13	3.28	0.77	3.21	0.90	.14	.17	.23	.13	.05	.27	.14	.30	.19	--

Table 2. Means, Standard Deviations, and Covariance Matrix (Total *N* = 500)

	Girls (n=500)		**Boys (n=500)**											
Item	*M*	*SD*	*M*	*SD*	1	2	3	4	5	6	7	8	9	10
B2	3.17	0.79	3.10	0.90	--	.42	.32	.33	.39	.25	.29	.28	.31	.20
B3	2.37	0.87	2.33	0.91	.31	--	.29	.36	.41	.25	.27	.29	.42	.21
B6	3.02	0.91	2.78	0.98	.11	.20	--	.27	.28	.31	.28	.39	.36	.34
B8	2.91	0.89	2.95	0.94	.21	.26	.18	--	.41	.27	.21	.33	.41	.17
B12	2.72	0.86	2.67	0.93	.20	.19	.17	.28	--	.25	.27	.24	.41	.20
B5	3.42	0.67	3.19	0.82	.10	.14	.17	.15	.13	--	.25	.41	.28	.40
B9	3.15	0.82	2.98	0.94	.12	.16	.12	.16	.18	.19	--	.34	.39	.26
B10	3.30	0.71	3.03	0.91	.10	.14	.22	.14	.10	.23	.17	--	.43	.45
B11	2.60	0.84	2.57	0.93	.14	.24	.24	.18	.19	.11	.24	.28	--	.27
B13	3.27	0.75	3.17	0.88	.11	.12	.23	.15	.07	.23	.13	.26	.19	--

Table 3. Item Means, Standard Deviations, and Covariance Matrix (Total *N* = 1000)

Items were graded on 4-point Likert-type categories: not important, somewhat unimportant, somewhat important, and very important. Covariance among items for total samples are displayed by group with girls' items covariance located under the diagonal and covariance among items for the boys sample are above the diagonal. All covariance estimates were significant at .05.

SEM MODEL

To evaluate the measurement invariance of the 10-item scale across groups (i.e., gender), the multiple group CFA model is (Reise et al., 1993),

$$S_g \simeq \hat{\Lambda}_g \hat{\Phi}_g \hat{\Lambda}_g{}' + \hat{\Psi}_g = \hat{\Sigma}_g,$$

Where

g denotes group membership (i.e., matrices are derived from the gth sample)
S = sample covariance matrix
Λ = (*p* x *r*) matrix of loadings on *p* measured variables (items) on the latent factors
Φ = (r x r) covariance matrix among measured latent variables
Σ = (*n* x *n*) population covariance matrix among measured variables (items)
Ψ = (r x r) covariance matrix among the measurement residuals

When the measurement or indicator variables are ordinal or categorical (i.e., Likert-type items), the CFA model specify the probability of each response category as a function of the latent variable ξ

$$Pr(x_1 = a_1, x_2 = a_2, \ldots x_p = a_p) \mid \xi_1, \xi_2, \ldots \xi_p) = f(\xi_1, \xi_2, \ldots \xi_k),$$

where

X = vector of measured variables or indicator variables x1, x2,…xp
ξ = vector of n underlying factors
$a_1, a_2, \ldots a_p$ = different response categories for x1, x2,…xp

Group parameters were obtained by re-specifying the 10-item subscale dimensions across gender. Figure 4 shows this the re-specification of items by gender:

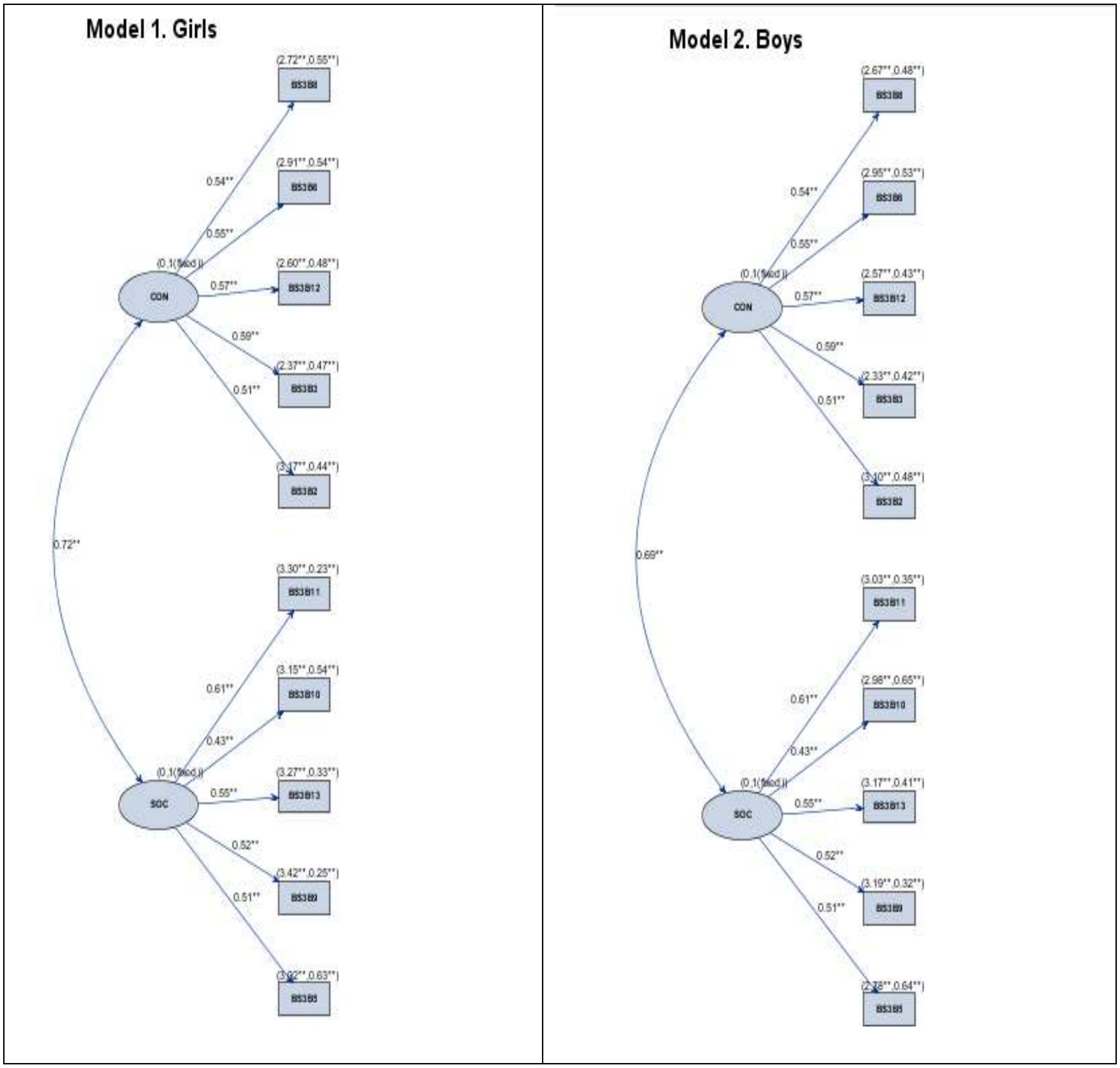

Figure 4. Model 1 and Model 2 Identification for Girls and Boys Respectively

To input group covariance matrices rather than raw data, the correlation matrix for each group was estimated from the bootstrapped samples created with PROC SURVEYSELECT using PROC CORR:

```
**Covariance Matrix by group;

   proc corr data=samples_b cov nocorr out=outcov(type=cov);
      var BS3B2 BS3B3 BS3B5 BS3B6 BS3B8 BS3B9 BS3B10 BS3B11 BS3B12 BS3B13;
      by BSGGEND;
   where replicate=1;
   run;
```

Then, the covariance matrices were saved as permanent files (G1 and G2):

```
data g1 (type=cov); set outcov; drop BSGGEND; where BSGGEND=1; run;
data g2 (type=cov); set outcov; drop BSGGEND; where BSGGEND=2; run;
```

Now, the re-specification by group was obtained as follows

```
ods graphics on; ods html style=statistical sge=on;
proc calis plots=pathdiagram;
   group 1 / data=G1;
   group 2 / data=G2;
Model 1 / group=1 label="Girls";
   factor
      CON ===> BS3B2 BS3B3 BS3B6  BS3B8  BS3B12,
      SOC ===> BS3B5 BS3B9 BS3B10 BS3B11 BS3B13;
   pvar
      CON = 1.0,
      SOC = 1.0;
model 2 /group=2 label="Boys";
refmodel 1;
run;
ods _all_ close; ods listing;
```

To evaluate the equality of mean vectors and variances matrices between groups, the COVPATTERN=EQCOVMAT and MEANPATTERN=EQMEANVEC options provided in PROC CALIS were used to invoke built-in covariance and mean patterns:

```
proc calis covpattern=eqcovmat meanpattern=eqmeanvec;
   outfiles outmodel=outmodel [model=1,2]
            outstat=outstat [group=1,2];
   var BS3B2 BS3B3 BS3B5 BS3B6 BS3B8 BS3B9 BS3B10 BS3B11 BS3B12 BS3B13;
   group 1 / data=g1;
   group 2 / data=g2;
   fitindex NoIndexType On(only)=[chisq df probchi rmsea agfi];
run;
```

The $\chi 2$ statistic, degrees of freedom, and p-value were used to reject or fail to reject the hypothesis of equal means and covariance structures between groups. These values were averaged across the 1000 replications, with the exception of the p-value. A significant *p*-value for each replication was coded 1 if it was significant at α = .05 and averaged across replications. Statistical power was computed for each condition. In addition to testing the equality of mean vectors and equality of covariance matrices, it was tested whether the two groups have equal structural models. To do so, simultaneous tests for estimating differences between intercepts are specified explicitly (i.e., not by default by PROC CALIS):

```
simtests
   MeasurementDiff =
   (BS3B2Diff BS3B3Diff BS3B6Diff  BS3B8Diff  BS3B12Diff
   BS3B5Diff BS3B9Diff BS3B10Diff BS3B11Diff BS3B13Diff);
   BS3B2Diff  = G2_Int_BS3B2  - G1_Int_BS3B2;
   BS3B3Diff  = G2_Int_BS3B3  - G1_Int_BS3B3;
   BS3B6Diff  = G2_Int_BS3B6  - G1_Int_BS3B6;
   BS3B8Diff  = G2_Int_BS3B8  - G1_Int_BS3B8;
   BS3B12Diff = G2_Int_BS3B12 - G1_Int_BS3B12;
   BS3B5Diff  = G2_Int_BS3B5  - G1_Int_BS3B5;
   BS3B9Diff  = G2_Int_BS3B9  - G1_Int_BS3B9;
   BS3B10Diff = G2_Int_BS3B10 - G1_Int_BS3B10;
   BS3B11Diff = G2_Int_BS3B11 - G1_Int_BS3B11;
   BS3B13Diff = G2_Int_BS3B13 - G1_Int_BS3B13;
   run;
```

The discrepancy per item (i.e., Function Value) was plotted using bar graphs and moderate to large effect size (η2) are reported.

IRT APPROACH

Using PROC IRT, researchers can perform multiple group analysis using the GRM with fixed values and equality constraints within and between groups to evaluate measurement invariance in items and tests. However, because observed frequency for some response patterns was 0, tests of fit were not conducted. Only MG-SEM results are presented

RESULTS

The test of the hypothesis of equal means and covariance structures between groups for each condition was conducted using fit indices values. Fit summaries are presented in Table 4:

Method	Sample Size	Proportion of Missing Observations	Proportion of Missing Covariates	Adjusted GFI	Chi Square	Pr Chi_Square	RMSEA
Complete Data	500	0.1	0.1	0.98	133.78	0.00	0.06
	500	0.1	0.2	0.98	132.69	0.00	0.06
	500	0.1	0.3	0.98	133.89	0.00	0.06
	500	0.2	0.1	0.98	133.57	0.00	0.06
	500	0.2	0.2	0.98	132.05	0.00	0.06
	500	0.2	0.3	0.98	132.94	0.00	0.06
	500	0.3	0.1	0.98	134.58	0.00	0.06
	500	0.3	0.2	0.98	132.64	0.00	0.06
	500	0.3	0.3	0.98	133.51	0.00	0.06
	1000	0.1	0.1	0.99	188.40	0.00	0.06
	1000	0.1	0.2	0.98	190.53	0.00	0.06
	1000	0.1	0.3	0.98	190.29	0.00	0.06
	1000	0.2	0.1	0.99	189.72	0.00	0.06
	1000	0.2	0.2	0.99	187.85	0.00	0.06
	1000	0.2	0.3	0.99	189.38	0.00	0.06
	1000	0.3	0.1	0.99	190.49	0.00	0.06
	1000	0.3	0.2	0.99	189.84	0.00	0.06
	1000	0.3	0.3	0.99	187.07	0.00	0.06
Listwise	500	0.1	0.1	0.98	132.06	0.00	0.06
	500	0.1	0.2	0.98	131.51	0.00	0.06
	500	0.1	0.3	0.98	131.58	0.00	0.06
	500	0.2	0.1	0.98	132.90	0.00	0.06
	500	0.2	0.2	0.98	132.58	0.00	0.07
	500	0.2	0.3	0.98	131.93	0.00	0.07
	500	0.3	0.1	0.98	134.14	0.00	0.07
	500	0.3	0.2	0.98	131.32	0.00	0.07
	500	0.3	0.3	0.98	130.79	0.00	0.07
	1000	0.1	0.1	0.98	187.53	0.00	0.06
	1000	0.1	0.2	0.98	187.09	0.00	0.06
	1000	0.1	0.3	0.98	185.36	0.00	0.06
	1000	0.2	0.1	0.98	189.75	0.00	0.06
	1000	0.2	0.2	0.98	186.28	0.00	0.06
	1000	0.2	0.3	0.98	187.95	0.00	0.06
	1000	0.3	0.1	0.98	188.74	0.00	0.06
	1000	0.3	0.2	0.98	184.83	0.00	0.06
	1000	0.3	0.3	0.98	181.89	0.00	0.06

					Table 4 Cont'		
MI	500	0.1	0.1	0.98	132.52	0.00	0.06
	500	0.1	0.2	0.98	130.77	0.00	0.06
	500	0.1	0.3	0.98	131.03	0.00	0.06
	500	0.2	0.1	0.98	131.18	0.00	0.06
	500	0.2	0.2	0.98	128.92	0.00	0.06
	500	0.2	0.3	0.98	127.87	0.01	0.06
	500	0.3	0.1	0.98	131.50	0.00	0.06
	500	0.3	0.2	0.98	126.00	0.01	0.06
	500	0.3	0.3	0.98	123.94	0.01	0.06
	1000	0.1	0.1	0.99	187.04	0.00	0.06
	1000	0.1	0.2	0.99	186.81	0.00	0.06
	1000	0.1	0.3	0.99	185.73	0.00	0.06
	1000	0.2	0.1	0.99	186.99	0.00	0.06
	1000	0.2	0.2	0.99	182.00	0.00	0.06
	1000	0.2	0.3	0.99	181.09	0.00	0.06
	1000	0.3	0.1	0.99	184.77	0.00	0.06
	1000	0.3	0.2	0.99	178.23	0.00	0.06
	1000	0.3	0.3	0.99	170.64	0.00	0.06
MG-MI	500	0.1	0.1	0.98	134.97	0.00	0.06
	500	0.1	0.2	0.98	135.61	0.00	0.06
	500	0.1	0.3	0.98	138.23	0.00	0.07
	500	0.2	0.1	0.98	136.27	0.00	0.07
	500	0.2	0.2	0.98	138.72	0.00	0.07
	500	0.2	0.3	0.98	143.13	0.00	0.07
	500	0.3	0.1	0.98	139.35	0.00	0.07
	500	0.3	0.2	0.98	141.27	0.00	0.07
	500	0.3	0.3	0.98	147.01	0.00	0.07
	1000	0.1	0.1	0.99	189.84	0.00	0.06
	1000	0.1	0.2	0.98	193.07	0.00	0.06
	1000	0.1	0.3	0.98	194.72	0.00	0.06
	1000	0.2	0.1	0.98	193.19	0.00	0.06
	1000	0.2	0.2	0.98	194.01	0.00	0.06
	1000	0.2	0.3	0.98	199.39	0.00	0.06
	1000	0.3	0.1	0.98	194.64	0.00	0.06
	1000	0.3	0.2	0.98	197.22	0.00	0.06
	1000	0.3	0.3	0.98	197.80	0.00	0.06

Table 4. Fit Summary. Each condition was replicated 1000.

As can be observed in Table 4, extremely χ^2 values were observed. With df = 65 and p-values less than 0.05, the hypothesis of equal means and covariance structures was rejected. However, the adjusted GFI index indicates a good model fit.

For the results of the simultaneous tests of measurement differences of the intercepts, for each individual item are presented in Figure 5:

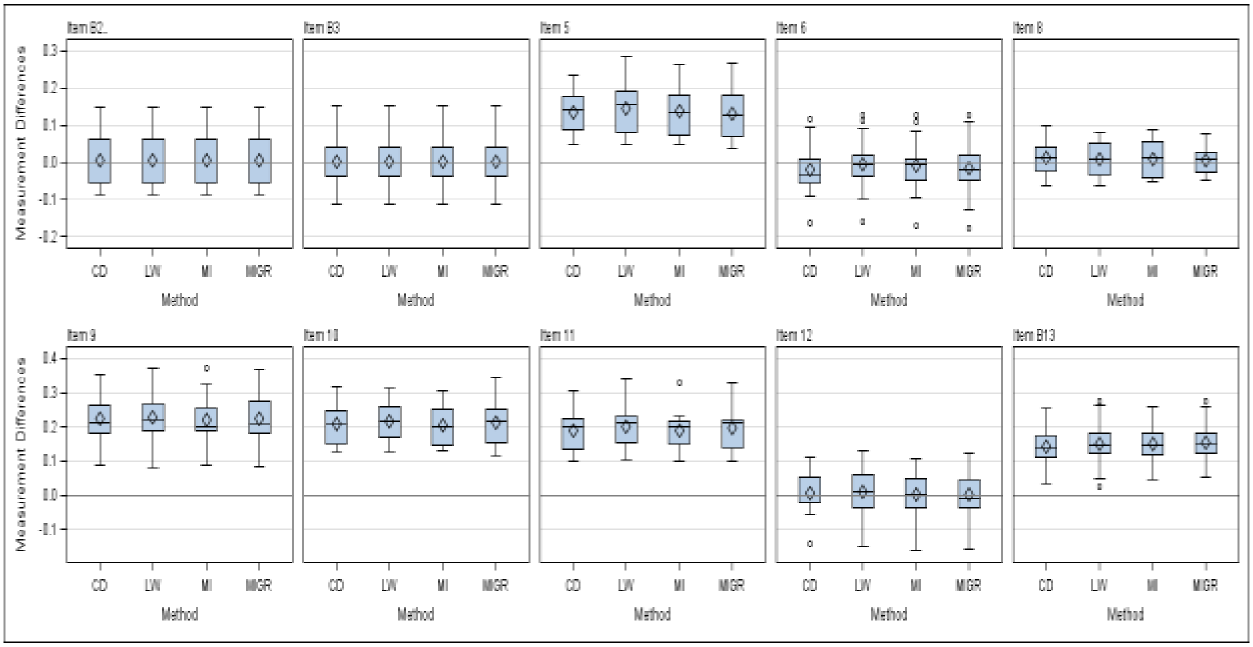

Figure 5. Simultaneous Tests of the Intercepts in the Measurement Model by Item and Method

The simultaneous tests of the intercepts in the measurement model suggests that the two groups (i.e., G1 = Girls, G2 = Boys) differ significantly in the means of some of the measured variables. Group 1 has significantly higher means in items 5, 9, 10, 11, and 13, which indicates that girls endorsed these items The means for items 2, 3, 6, 8, and 12 are similar in both groups. This indicates that Girls endorsed the factor "conventional" (CON), regarding the desirability of being politically active in conventional forms of participation while both groups endorsed about the same the items for the factor "social movement" (SOC) indicating the desirability of having an active political participation with new forms of social movements.

As can be observed in Figure 5, variability in measurement differences was very consistent across all items and methods. An analysis of effect size for MI and MG-MI across all items showed that MI and MG-MI had the same effect size for item B2 and B8 ($\eta^2 = 0.12$) for the interaction of sample size and proportion of missing observations. Both MI and MG-MI had large effect size for item B5 ($\eta^2 = 0.21$ and $\eta^2 = 0.16$ respectively) for the interaction of sample size and the proportion of missing covariates. For item B6, MG-MI had moderate effect size ($\eta^2 = 0.08$) while MI had a large effect size ($\eta^2 = 0.14$) for the interaction of sample size and proportion of missing covariates. For items B10 and B11, MG-MI has a slightly smaller effect sizes than MI ($\eta^2 = 0.08$ - $\eta^2 = 0.10$ and $\eta^2 = 0.07$ $\eta^2 = 0.08$ respectively). For item 12, both MI and MG-MI had no effect on the variability of the measurement difference for the item while for MG-MI had a large effect size ($\eta^2 = 0.09$) for item 9 when sample size and the proportion of missing covariates interacted. Finally, for item 13, MI had a moderate effect size for the interaction of sample size and the proportion of missing covariates while MG-MI had a large effect ($\eta^2 = 0.10$) for the same interaction, in addition to having also a moderate effect size ($\eta^2 = 0.07$) for the interaction of the proportion of missing observations and the proportion of missing covariates.

Overall, both MI and MG-MI had moderate to large effect sizes over most items but MI had larger effect sizes than MG-MI only in two items. Thus, under the MI approach, smaller measurement differences between groups across items were obtained.

DISCUSSION AND CONCLUSIONS

In terms of population inferences, there is no single best method. However, the use of SEM for the analysis of latent variable models has the advantage of being more accurate for estimating latent constructs because latent factors estimates in SEM are free from measurement error. Specifically, MG-SEM allows the simultaneous examination of mean and covariance structures for evaluating the similarities and differences of these structures among two or more groups. IRT has become the most widely implemented methodology in the field of psychometrics for item calibration and subject scoring. Until recently, IRT analyses were performed only in very specialized software and PROC IRT is a great addition to the SAS tools for statistical analysis. While both SEM and IRT modeling are heavy, computer extensive methods, they are important methodologies that provide optimal solutions over other estimation methods for example Full Information Maximum Likelihood (FIML).

Multiple Imputation is theoretically superior to Listwise deletion; however, as shown in Figure, both methods performed similarly across items. Thus, the application of Listwise deletion to treat missing data might be a sound option in some situations (e.g., MAR data, small percentage of missing items and variables) because MI is computational expensive in terms of execution time.

REFERENCES

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103*(3), 411-423.

Beaujean, A. A. (2014). *Latent Variable Modeling using R: A step-by-step approach.* New York, NY; Taylor & Francis.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 8*(3), 430-457.

Enders, C. K., & Gottschall, A. C. (2011). Multiple imputation strategies for multiple group structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 18*(1), 35-54.

Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analysis. *Journal of School Psychology, 48*, 5-37.

Bentler, P. M. (1983). Some contributions to efficient statistics for structural models: Specification and estimation of moment structures*. Psychometrika, 48*, 493-517.

Browne, M. W. (1984). Asymptotically distribution-free methods in the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*(1), 62-83.

Cristoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika*, *40*, 5-32.

Graham, J. W., & Hofer, S. M. (2000). Multiple imputation in multivariate research. In T. D. Little, K. U. Schanabel, & J. Baumert (Eds.). *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 201-218). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research, 33*(4), 545-571.

Kalton, G. (1983), *Introduction to Survey Sampling*, Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-035, Beverly Hills, CA: Sage Publications.

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

Muthen, B. (1978). Contributions to factor analysis of dichotomous variables, *Psychometrika*, *43*, 551-560.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*(3), 581-592

SAS Institute Inc. (2012). *SAS* (version 9.4) *[Computer Software].* Cary, NC: SAS Institute Inc.

SAS Institute Inc. (2013). *SAS/STAT® 13.2 User's Guide*. Cary, NC: SAS Institute, Inc.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph No. 17*, *34*(4, Pt. 2).

Samejima, F. (2010). The general graded response model. In M. L. Nering & R. Ostini (Eds.), Handbook of polytomous item response theory (pp. 77-107), New York, NY: Routledge, Taylor & Francis Group.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*(2), 147-177.

Schulz, W. (2004). Scaling procedures for Likert-type items on students' concepts, attitudes, and actions. In W. Schulz & H. Sibberns (Eds.), IEA Civic Education Study technical report (pp. 93–126). Amsterdam, The Netherlands: Paula Wagemaker Editorial Services.

Shi, J.Q, & Lee, S.Y. (2000). Latent variables models with mixed continuous and polytomous data. *Journal of the Royal Statistical Society: Series (Statistical Methodology)*,*62*(1), 77-87.

Takane, Y., & de Leew, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, *52*(3). 393-408.

Tsikriktsis, N. (2005). A review of techniques for treating missing data in OM survey research. *Journal of Operations Management, 24,* 53-62.

Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods, 12*(1), 58-79.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Patricia Rodriguez de Gil
University of South Florida
prodrig6@mail.usf.edu