



Measurement Bias Detection Through Factor Analysis

M. T. Barendse , F. J. Oort , C. S. Werner , R. Ligtoet & K. Schermelleh-Engel

To cite this article: M. T. Barendse , F. J. Oort , C. S. Werner , R. Ligtoet & K. Schermelleh-Engel (2012) Measurement Bias Detection Through Factor Analysis, Structural Equation Modeling: A Multidisciplinary Journal, 19:4, 561-579, DOI: [10.1080/10705511.2012.713261](https://doi.org/10.1080/10705511.2012.713261)

To link to this article: <https://doi.org/10.1080/10705511.2012.713261>



Published online: 31 Oct 2012.



Submit your article to this journal [↗](#)



Article views: 869



View related articles [↗](#)



Citing articles: 18 View citing articles [↗](#)

Measurement Bias Detection Through Factor Analysis

M. T. Barendse,¹ F. J. Oort,¹ C. S. Werner,² R. Ligtvoet,¹ and
K. Schermelleh-Engel²

¹*University of Amsterdam*

²*Goethe University*

Measurement bias is defined as a violation of measurement invariance, which can be investigated through multigroup factor analysis (MGFA), by testing across-group differences in intercepts (uniform bias) and factor loadings (nonuniform bias). Restricted factor analysis (RFA) can also be used to detect measurement bias. To also enable nonuniform bias detection, we extend RFA with latent moderated structures (LMS) or use a random slope parameterization (RSP). In a simulation study we compare the MGFA, RFA/LMS, and RFA/RSP methods in detecting measurement bias, varying type of bias (uniform, nonuniform), type of the variable that violates measurement invariance (dichotomous, continuous), and its relationship with the trait that we want to measure (independent, dependent). For each condition, 500 sets of data are generated and analyzed with each of the three detection methods, in single run and in an iterative procedure. The RFA methods outperform MGFA when the violating variable is continuous.

Keywords: factor analysis, latent moderated structures, measurement invariance, nonuniform measurement bias, random slope parameterization

In the presence of measurement bias, observed differences in test scores do not validly represent actual differences in what the test is supposed to measure. Individual differences (including gender, age, culture, and language) and differences in treatment—as in experimental research—have an effect on response behavior and could affect the validity of the measurement. Measurement bias (or item bias, or differential item functioning) is formally defined as a violation of measurement invariance (cf. Mellenbergh, 1989):

$$f_1(X|T = t, V = v) = f_2(X|T = t), \quad (1)$$

where X is a set of observed variables (e.g., test items or questionnaire scales), T is the trait of interest measured by X , and V is a set of variables other than T , possibly violating

Correspondence should be addressed to F. J. Oort, Methods and Statistics, Department of Education, University of Amsterdam, Nieuwe Prinsengracht 130, 1018 VZ Amsterdam, The Netherlands. E-mail: f.j.oort@uva.nl

conditional independence. Function f_1 is the conditional distribution function of X given values t and v , and f_2 is the conditional distribution function of X given t . If the conditional independence does not hold, that is, if $f_1 \neq f_2$, then the measurement of T by X is said to be biased with respect to V . Variable V is then called a “violator” of measurement invariance (hence the symbol V). If V is a group membership, then measurement bias can be investigated through multigroup factor analysis (MGFA; Meredith, 1993; Schmitt & Kuljanin, 2008; Vandenberg & Lance, 2000). In MGFA, across-group differences in intercepts indicate uniform bias (i.e., constant bias) and across-group differences in factor loadings indicate nonuniform bias (i.e., the size of the bias varies with t). Simulation studies of MGFA show that uniform bias is easier to detect than nonuniform bias (Barendse, Oort, & Garst, 2010; González-Romá, Hernández, & Gómez-Benito, 2006; Hernández & González-Romá, 2003; Meade & Lautenschlager, 2004).

Oort (1992, 1998) suggested the use of restricted factor analysis (RFA) as a means to investigate bias with respect to group membership and other variables (Oort, 1991). In the RFA method of bias detection, the factor model is extended with one or more exogenous variables V that are correlated with the common factor that represents T . If covariances between V variables and observed measurements X are fully explained by the indirect relationships through T , then the X variables are measurement invariant with respect to possible violators V . However, if there is a direct effect of V on X , then X is said to be biased with respect to V . The RFA method to detect measurement bias yields results that are equivalent to the results of multiple indicator multiple cause (MIMIC) analysis, but in MIMIC models the V variables have causal effects on the T variables (Muthén, 1989).

Advantages of RFA (and MIMIC analysis) over MGFA when investigating measurement invariance are that in RFA variables V can be continuous or discrete, and observed or latent, and measurement bias can be investigated with respect to multiple variables V simultaneously. Moreover, as it is not necessary to divide the sample into subsamples by V , RFA should also have more statistical power and greater efficiency. A possible disadvantage of RFA is that it is not readily suited to detect nonuniform bias. In the RFA model, nonuniform bias appears as a nonlinear interaction effect, violating the assumption of multivariate normality. There are many ways to model interaction effects as part of structural equation models, including product indicator approaches and distribution-analytic approaches (Moosbrugger, Schermelleh-Engel, Kelava, & Klein, 2009; Schermelleh-Engel, Werner, Klein, & Moosbrugger, 2010). The analysis technique of latent moderated structural equations (LMS) belongs to the latter class. It was introduced by Klein and Moosbrugger (2000), who showed that its estimates are consistent, unbiased, and efficient. As an alternative to LMS, Muthén and Asparouhov (2003) explained how interaction effects can also be modeled as random regression coefficients in (multilevel) structural equation modeling (SEM), but the performance of this random slope parameterization (RSP) is practically unknown.

Barendse et al. (2010) showed how the RFA method with LMS, as implemented in the computer program *Mplus* (Muthén & Muthén, 2007), can detect uniform and nonuniform bias at least as well as the MGFA method. In another simulation study, Woods and Grimm (2011) also used LMS to extend their MIMIC method for the detection of bias in dichotomous and polytomous items (with five response categories). Both studies examined bias with respect to a group membership (i.e., a dichotomous violator) only. Both studies report good detection rates for biased items, but they also report concerns about inflated Type I error rates. The

proportions of false positives exceed the nominal levels of significance substantially. Earlier studies show that iterative detection procedures guard against Type I error inflation (Navas-Ara & Gómez-Benito, 2002; Oort, 1998).

In this simulation study we compare the performance of MGFA, RFA with LMS, and RFA with RSP in detecting uniform and nonuniform measurement bias, with respect to both dichotomous and continuous violators, in both single run and iterative procedures.

METHODS

Using simulated data, measurement bias is assessed by three methods: MGFA, RFA/LMS, and RFA/RSP. For data generation we vary type of bias (no bias, uniform bias, nonuniform bias, both uniform and nonuniform bias), type of violator (continuous or dichotomous violator), and relationship between trait and violator (independent or dependent). In a fully crossed design, these three data generation factors yield ($4 \times 2 \times 2 =$)16 different conditions. In each condition, 500 data sets are generated. Each data set is analyzed with each of the three methods (MGFA, RFA/LMS, and RFA/RSP), using two different procedures (single run or iteratively). The performance of each method in detecting measurement bias is evaluated by calculating proportions of true and false positives and by assessing the bias in parameter estimation (efficiency and accuracy).

Data Generation

Each data set consists of scores of 200 subjects on six items with continuous response scales. For each subject j , a vector of scores x_j on observed variables X is generated according to the linear model:

$$x_j = u + at_j + bv_j + ct_jv_j + de_j, \quad (2)$$

where t_j is subject j 's score on the latent trait T ; v_j is subject j 's score on the potential violator V ; e_j is a vector of subject j 's scores on the residual factors E of the observed variables X ; u is a vector of intercepts; and a , b , c , and d are vectors of regression coefficients.

In conditions with bias with respect to a continuous violator, subject parameters t and v are drawn from a bivariate standard normal distribution with zero correlation between T and V in conditions with independent T and V and 0.5 correlation in conditions with dependent T and V . Residual factor scores e are independently drawn from standard normal distributions. Intercepts u are set at zero, and regression coefficients a and d are set at 0.9 for all items in all conditions with continuous violators. Bias is induced in the first item only, either by setting b at 0.3 for uniform bias, or c at 0.3 for nonuniform bias, or both. As a result, the proportion of observed variance caused by bias is about 5% in items with either uniform or nonuniform bias, and about 10% in items with both biases.

In conditions with bias with respect to a dichotomous violator (group membership), we conveniently choose $V = -1$ for one group and $V = 1$ for the other group, retaining all other parameter choices, so that $u_1 = u - b$ and $a_1 = a - c$ in Group 1 and $u_2 = u + b$ and

TABLE 1
Parameter Values Under 16 Data Generation Conditions

	Condition	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	Corr <i>T, V</i>				
Continuous <i>V</i> ^a										
Independent <i>T, V</i>										
	No bias	1	0.9	0.0	0.0	0.9	0.0			
	Uniform	2	0.9	0.3	0.0	0.9	0.0			
	Nonuniform	3	0.9	0.0	0.3	0.9	0.0			
	Both biases	4	0.9	0.3	0.3	0.9	0.0			
Dependent <i>T, V</i>										
	No bias	5	0.9	0.0	0.0	0.9	0.5			
	Uniform	6	0.9	0.3	0.0	0.9	0.5			
	Nonuniform	7	0.9	0.0	0.3	0.9	0.5			
	Both biases	8	0.9	0.3	0.3	0.9	0.5			
	Condition	<i>u</i> ₁	<i>a</i> ₁	<i>d</i> ₁	Mean <i>T</i> ₁	<i>u</i> ₂	<i>a</i> ₂	<i>d</i> ₂	Mean <i>T</i> ₂	
Dichotomous <i>V</i> ^b										
Independent <i>T, V</i>										
	No bias	9	0.0	0.9	0.9	0.0	0.0	0.9	0.9	0.0
	Uniform	10	-0.3	0.9	0.9	0.0	0.3	0.9	0.9	0.0
	Nonuniform	11	0.0	0.6	0.9	0.0	0.0	1.2	0.9	0.0
	Both biases	12	-0.3	0.6	0.9	0.0	0.3	1.2	0.9	0.0
Dependent <i>T, V</i>										
	No bias	13	0.0	0.9	0.9	-0.4	0.0	0.9	0.9	0.4
	Uniform	14	-0.3	0.9	0.9	-0.4	0.3	0.9	0.9	0.4
	Nonuniform	15	0.0	0.6	0.9	-0.4	0.0	1.2	0.9	0.4
	Both biases	16	-0.3	0.6	0.9	-0.4	0.3	1.2	0.9	0.4

Note. All values concern the parameters of the first of six items; Condition 1 values are also given to the other five items in Conditions 1, 2, 3, and 4; Condition 5 values are also given to the other five items in Conditions 5, 6, 7, and 8; Condition 9 values are also given to the other five items in Conditions 9, 10, 11, and 12; Condition 13 values are also given to the other five items in Conditions 13, 14, 15, and 16.

See Equation 2: ^a*u* = 0, mean(*T*) = mean(*V*) = mean(*E*) = 0, var(*T*) = var(*V*) = var(*E*) = 1. ^bvar(*T*)₁ = var(*T*)₂ = 1, mean(*E*)₁ = mean(*E*)₂ = 0, var(*E*)₁ = var(*E*)₂ = 1.

*a*₂ = *a* + *c* in Group 2. In conditions with dependent *T* and *V*, this dependency is created by drawing 100 subject parameters *v* from a normal distribution with mean -0.4 for one group and 100 with mean 0.4 for the other group, and unity variance for both groups.

Table 1 gives an overview of 2 × 2 × 4 conditions and parameter value choices.

Measurement Bias Detection

In the RFA methods for bias detection we include the possible violator variable *V* in the model. In the MGFA method, variable *V* does not feature in the factor model but is used to divide the sample into subsamples.

MGFA Method

In the MGFA method, the maximum likelihood estimation method is used to fit a one-factor model to the variances, covariances, and means of the separate groups. If the violator is

continuous then two groups are created by dichotomizing at the mean of violator variable V first. The common factor mean and variance are fixed at zero and unity for one group and free to be estimated for the other group. Intercepts and factor loadings are constrained to be equal across groups.

Measurement bias in a particular item is detected by removing the across-group constraints on its intercept and factor loading, and comparing the fit of this model with the fit of the model with all constraints in place. The chi-square difference test with 2 df serves as a global test of significance of bias, be it uniform, nonuniform, or both. Differences in intercepts and in factor loadings indicate uniform and nonuniform bias, respectively.

RFA/LMS Method

In the RFA/LMS method, latent moderated structures are used to assess nonuniform bias. With LMS it is possible to estimate an interaction effect of two latent variables on an observed variable (Klein & Moosbrugger, 2000). If there is an interaction effect in the data, the dependent variable cannot be normally distributed even if the latent predictor variables are both normally distributed. For the maximum likelihood estimator used to analyze linear structural equation models, violation of the multivariate normality assumption can cause estimation problems. Klein and Moosbrugger (2000) solved this problem by conditioning on one of the latent variables and regarding the distribution of the observed variable as a mixture of multiple conditional distributions. LMS as implemented in *Mplus* utilizes robust maximum likelihood estimation (Muthén & Muthén, 2007; equivalent to Yuan & Bentler, 2000). As LMS is really only suited for products of latent variables, we model the violator variable V as a latent variable with a single indicator with residual variance fixed at a small value of 0.01 to overcome identification problems. In addition, to help convergence, we also fix the corresponding factor loading at the square root of the difference between observed variance and residual variance. So, in the RFA/LMS method we fit a three-factor model to the data: One factor represents latent trait T , one factor represents observed variable V , and one factor represents the product of T and V . See Appendix A for an example *Mplus* script.

Measurement bias in a particular item is detected by comparing the fit of a model with and without direct effects of the violator variable and of the product of the common factor and the violator variable. The difference between the log-likelihood values associated with the two models has a chi-square distribution with 2 df , subject to the scaling correction factors of the two models (Satorra & Bentler, 2001). Significant regressions on V and on $T \times V$ would then indicate uniform and nonuniform bias.

RFA/RSP Method

In the RFA/RSP method, random slopes are used to assess nonuniform bias. Muthén and Asparouhov (2003) explained how random slopes in multilevel SEM can be used to estimate the effect of an interaction between a latent and an observed variable, by replacing the interaction term $ct_j v_j$ in Equation 2 with $c'_j v_j$, where c'_j is a random slope. Similar to the RFA/LMS method, *Mplus* utilizes robust maximum likelihood estimation with a scaling correction, to account for the violation of distributional assumptions (Muthén & Muthén, 2007). See Appendix B for an example script.

With this RSP, measurement bias in a particular item can be detected in the same way as with the RFA/LMS method, by comparing the fit of a model with and without nonzero fixed and random regressions on the violator variable. The difference between the log-likelihood values associated with the two models has a chi-square distribution with 2 *df*, subject to the scaling correction factors of the two models (Satorra & Bentler, 2001). A significant fixed coefficient for the regression on *V* indicates uniform bias and a significant random coefficient for the regression on *V* indicates nonuniform bias.

Analysis

The computer program *Mplus* (version 5.2; Muthén & Muthén, 2007) is used to apply the MGFA, RFA/LMS, and RFA/RSP methods for bias detection to each of the 8,000 data sets (500 replications in each of the 16 conditions). Measurement bias will be tested at 5%, 1%, and 0.1% levels of significance. Each of the three methods will be applied in both a single run procedure and in an iterative procedure.

In the single run procedure, we test each item for bias once (which includes fitting seven models: the unconstrained model plus one constrained model for each item), and we count true positives and false positives. True positives are biased items that are correctly detected as biased with the 2 *df* chi-square test. False positives are unbiased items incorrectly detected as biased. Proportions of true positives and false positives are indications of power and Type I error, respectively.

In the iterative procedure, only the item associated with the largest chi-square difference in the first run is considered biased. In a second run, this bias is taken into account by allowing for different intercepts and factor loadings in the MGFA procedure or by allowing nonzero regressions on *V* and $T \times V$ in the RFA/RSP procedure, and the chi-square difference test is applied to the remaining items. The iterative procedure stops as soon as none of the remaining items is associated with a significant bias test, or as soon as half of the items are detected as biased, whichever comes first. The accuracy and efficiency of the parameter estimates is evaluated by examining the parameter estimates of the MGFA, RFA/LMS, and RFA/RSP models that account for bias in the first item (i.e., the biased item in conditions with bias). The accuracy is evaluated by calculating the differences between the parameter estimates and the chosen population values (rescaled to the metric that is used in the respective bias detection methods). The standard deviations of the parameter estimates indicate efficiency. These standard deviations can also be used to evaluate the accuracy of the estimated standard errors, by comparing them to the averages of the standard error estimates.

RESULTS

After applying each of the three methods to each of the 8,000 data sets, we find that the RFA/LMS method does not always produce valid results. With 64 data sets (0.8%) we encounter convergence problems or negative chi-square difference tests, especially in conditions with a continuous violator that is correlated with the latent trait (17 inadmissible results in Condition 6, 13 in Condition 7, and 22 in Condition 8; refer to Table 1 for condition descriptions). These

64 data sets are replaced by new data sets. The MGFA and RFA/RSP methods do not produce such problems.

Measurement Bias Detection

Table 2 gives the single run results for the MGFA method: The means and standard deviations of the chi-square difference tests, and the proportions of true positives and false positives for varying levels of significance. With a (dichotomized) continuous violator, the MGFA method detects 65% to 96% of the biased items, or 80% overall, at the 5% level of significance. With a dichotomous violator, 93% to 100% of the biased items are detected, or 97% overall. Nonuniform bias is more difficult to detect than uniform bias, unless T and V are correlated. The percentages of false positives are close to the nominal levels of significance in conditions without bias (Conditions 1, 5, 9, and 13), but higher in conditions with bias, especially if both uniform and nonuniform bias are present (Conditions 4, 8, 12, and 16). Using the MGFA method in the iterative procedure does not affect the percentages of true positives much, but the percentages of false positives are consistently closer to the nominal levels of significance; see Table 3.

The RFA/LMS and RFA/RSP single run results, given in Tables 4 and 5, are almost equivalent. At the 5% level of significance, both RFA methods detect 89% to 100% of the biased items. With a continuous violator the overall percentage of true positives is 95% and with a dichotomous violator it is 97%. As with the MGFA method, percentages of false positives are close to the nominal levels of significance in no-bias conditions, but substantially higher in conditions with bias. Table 6 shows that using the RFA/RSP method iteratively hardly affects the percentages of true positives, but it does bring the percentages of false positives close to the nominal levels of significance in all conditions. We do not report the iterative RFA/LMS results because they are almost equivalent with the iterative RFA/RSP results.

Estimation Bias

Tables 7, 8, and 9 give the accuracy and efficiency of the estimates of the parameters that are associated with uniform and nonuniform bias. In the MGFA method, uniform bias is assessed by separately estimating intercepts. Over all conditions, the average difference between the estimate and expected value is only 0.001, but average differences per condition vary from -0.022 to 0.015 (the expected values of rescaled intercepts are about 0.24, plus or minus). Nonuniform bias is assessed through the factor loadings. In conditions with a (dichotomized) continuous violator the estimation bias is much larger (averages vary from -0.040 to 0.029) than in conditions with a dichotomous violator (averages vary from -0.010 to 0.008 ; the expected values of the rescaled factor loadings vary around 0.71). In all conditions, the standard deviations of the MGFA estimates are close to the means of the standard errors, both for intercepts and for factor loadings.

For the RFA/LMS and RFA/RSP methods, uniform bias is indicated by the regression of the observed variable on the violator. On average, the difference with the expected value is -0.001 (the expected value of the rescaled parameter is about 0.21). Nonuniform bias is indicated by the regression on $T \times V$. As with the MGFA method, the accuracy of this regression coefficient

TABLE 2
Single Run Results of Measurement Bias Detection With the Multigroup Factor Analysis Method

Condition	Biased Items ^a				Unbiased Items ^b					
	χ^2 Test (df = 2)		Proportions of True Positives		χ^2 Test (df = 2)		Proportions of False Positives			
	M	SD	$\alpha = .05$	$\alpha = .01$	$\alpha = .001$	M	SD	$\alpha = .05$	$\alpha = .01$	$\alpha = .001$
Continuous V										
Independent T, V	Overall		0.796	0.614	0.384			0.081	0.022	0.003
No bias	1	—	—	—	—	2.063	2.054	0.057	0.011	0.001
Uniform	2	12.497	6.516	0.848	0.644	2.380	2.288	0.078	0.020	0.002
Nonuniform	3	9.261	6.060	0.656	0.416	2.353	2.319	0.076	0.021	0.002
Both biases	4	19.253	8.739	0.962	0.902	2.669	2.626	0.109	0.029	0.005
Dependent T, V										
No bias	5	—	—	—	—	2.046	2.063	0.058	0.013	0.001
Uniform	6	9.003	5.762	0.646	0.406	2.362	2.278	0.074	0.020	0.000
Nonuniform	7	10.248	6.038	0.738	0.508	2.324	2.300	0.075	0.019	0.003
Both biases	8	16.464	8.193	0.924	0.806	2.861	2.894	0.124	0.039	0.007
Dichotomous V	Overall		0.966	0.904	0.739			0.110	0.032	0.005
Independent T, V										
No bias	9	—	—	—	—	2.106	2.148	0.060	0.015	0.001
Uniform	10	20.137	8.342	0.980	0.930	2.801	2.687	0.115	0.036	0.004
Nonuniform	11	15.591	7.222	0.932	0.816	2.546	2.475	0.091	0.025	0.004
Both biases	12	31.285	10.864	0.998	0.998	3.079	3.006	0.140	0.049	0.010
Dependent T, V										
No bias	13	—	—	—	—	2.113	2.081	0.061	0.011	0.000
Uniform	14	16.511	7.347	0.946	0.850	2.862	2.753	0.120	0.037	0.006
Nonuniform	15	16.245	7.357	0.944	0.842	2.582	2.524	0.095	0.027	0.006
Both biases	16	30.587	10.917	0.994	0.986	3.545	3.033	0.196	0.056	0.006

Note. Overall proportions are calculated over eight conditions.

^aMeans, standard deviations, and proportions are calculated over 500 observations. ^bMeans, standard deviations, and proportions are calculated over 500 × 6 = 3,000 observations in Conditions 1, 5, 9, and 13, and over 500 × 5 = 2,500 observations in all other conditions.

TABLE 3
Iterative Procedure Results of Measurement Bias Detection With the Multigroup Factor Analysis Method

Condition		Biased Items ^a Proportions of True Positives			Unbiased Items ^b Proportions of False Positives		
		$\alpha = .05$	$\alpha = .01$	$\alpha = .001$	$\alpha = .05$	$\alpha = .01$	$\alpha = .001$
Continuous <i>V</i>	Overall	0.779	0.584	0.390	0.053	0.014	0.002
Independent <i>T</i> , <i>V</i>							
	No bias	—	—	—	0.053	0.010	0.001
	Uniform	0.854	0.686	0.424	0.055	0.016	0.001
	Nonuniform	0.636	0.410	0.192	0.054	0.016	0.002
	Both biases	0.948	0.706	0.706	0.051	0.016	0.002
Dependent <i>T</i> , <i>V</i>							
	No bias	—	—	—	0.053	0.013	0.001
	Uniform	0.614	0.398	0.200	0.049	0.012	0.000
	Nonuniform	0.712	0.502	0.230	0.053	0.012	0.002
	Both biases	0.912	0.800	0.590	0.059	0.016	0.003
Dichotomous <i>V</i>	Overall	0.955	0.898	0.739	0.053	0.014	0.002
Independent <i>T</i> , <i>V</i>							
	No bias	—	—	—	0.054	0.014	0.001
	Uniform	0.976	0.924	0.746	0.056	0.012	0.002
	Nonuniform	0.912	0.804	0.566	0.047	0.015	0.002
	Both biases	0.996	0.996	0.970	0.055	0.012	0.002
Dependent <i>T</i> , <i>V</i>							
	No bias	—	—	—	0.057	0.011	0.000
	Uniform	0.924	0.846	0.600	0.057	0.014	0.002
	Nonuniform	0.934	0.834	0.592	0.048	0.016	0.002
	Both biases	0.990	0.984	0.958	0.050	0.014	0.002

Note. Overall proportions are calculated over eight conditions.

^aProportions are calculated over 500 observations. ^bProportions are calculated over $500 \times 6 = 3,000$ observations in Conditions 1, 5, 9, and 13, and over $500 \times 5 = 2,500$ observations in all other conditions.

in the conditions with a continuous violator is not as good (averages vary from -0.039 to 0.004) as in conditions with a dichotomous violator (averages vary from -0.006 to 0.004 ; the expected value of the rescaled parameter is about 0.21). In conditions with nonuniform bias with respect to a continuous violator, the bias is underestimated. Standard deviations of parameter estimates are consistently close to the means of the standard errors, and much smaller than with the MGFA method. The better efficiency of the RFA methods is due to the smaller ratio of numbers of parameters and sample size.

DISCUSSION

The bias detection rates of all three factor analysis methods are generally satisfactory. Still, the RFA methods clearly outperform the MGFA method in conditions with a continuous violator and also perform slightly better in conditions with a dichotomous violator (i.e., group

TABLE 4
Single Run Results of Measurement Bias Detection With the Restricted Factor Analysis/Latent Moderated Structures Method

Condition	Biased Items ^a				Unbiased Items ^b					
	χ^2 Test (<i>df</i> = 2)		Proportions of True Positives		χ^2 Test (<i>df</i> = 2)		Proportions of False Positives			
	<i>M</i>	<i>SD</i>	$\alpha = .05$	$\alpha = .01$	$\alpha = .001$	<i>M</i>	<i>SD</i>	$\alpha = .05$	$\alpha = .01$	$\alpha = .001$
Continuous <i>V</i>	Overall		0.954	0.883	0.748			0.108	0.035	0.008
Independent <i>T</i> , <i>V</i>										
No bias	1	—	—	—	—	2.209	2.237	0.070	0.017	0.002
Uniform	2	22.344	10.069	0.976	0.926	2.905	2.780	0.126	0.039	0.005
Nonuniform	3	17.891	11.761 ^c	0.892	0.768	2.423	2.612	0.082	0.027	0.006
Both biases	4	39.311	21.964	1.000	0.984	3.040	3.149	0.136	0.051	0.013
Dependent <i>T</i> , <i>V</i>										
No bias	5	—	—	—	—	2.230	2.348	0.070	0.015	0.004
Uniform	6	15.702	9.561	0.890	0.730	2.928	2.893	0.132	0.044	0.008
Nonuniform	7	24.455	16.406 ^c	0.968	0.900	2.547	2.655	0.091	0.029	0.006
Both biases	8	38.349	25.964 ^c	0.996	0.988	3.258	3.338	0.158	0.058	0.018
Dichotomous <i>V</i>	Overall		0.973	0.927	0.790			0.100	0.033	0.006
Independent <i>T</i> , <i>V</i>										
No bias	9	—	—	—	—	2.152	2.246	0.064	0.040	0.009
Uniform	10	21.036	9.121	0.982	0.932	2.849	2.756	0.124	0.038	0.005
Nonuniform	11	19.498	9.434	0.956	0.898	2.343	2.353	0.074	0.020	0.004
Both biases	12	37.846	14.976	1.000	0.996	2.880	2.893	0.123	0.042	0.009
Dependent <i>T</i> , <i>V</i>										
No bias	13	—	—	—	—	2.164	2.177	0.062	0.016	0.001
Uniform	14	17.060	8.091	0.942	0.848	2.895	2.811	0.126	0.042	0.006
Nonuniform	15	19.995	9.624	0.964	0.898	2.364	2.397	0.076	0.020	0.004
Both biases	16	35.566	13.919	0.996	0.990	3.154	2.877	0.151	0.042	0.006

Note. Overall proportions are calculated over eight conditions.

^aMeans, standard deviations, and proportions are calculated over 500 observations. ^bMeans, standard deviations, and proportions are calculated over 500 × 6 = 3,000 observations in Conditions 1, 5, 9, and 13, and over 500 × 5 = 2,500 observations in all other conditions. ^cAfter removal of 2, 1, and 2 outliers in Conditions 3, 7, and 8; before removing outliers, means (*M*) and standard deviations (*SD*) are *M* = 18.559 and *SD* = 16.506 in Condition 3, *M* = 25.740 and *SD* = 26.262 in Condition 7, and *M* = 42.599 and *SD* = 66.751 in Condition 8.

TABLE 5
Single Run Results of Measurement Bias Detection With the Restricted Factor Analysis/Random Slope Parameterization Method

Condition	Biased Items ^a						Unbiased Items ^b					
	χ^2 Test (df = 2)			Proportions of True Positives			χ^2 Test (df = 2)			Proportions of False Positives		
	M	SD	$\alpha = .05$	$\alpha = .01$	$\alpha = .001$	M	SD	$\alpha = .05$	$\alpha = .01$	$\alpha = .001$		
Continuous V	Overall											
Independent T, V												
No bias	—	—	0.954	0.884	0.748	—	—	0.108	0.035	0.008		
Uniform	22.339	10.065	—	—	—	2.208	2.237	0.070	0.017	0.002		
Nonuniform	17.858	11.725 ^c	0.976	0.926	0.794	2.905	2.780	0.126	0.039	0.005		
Both biases	39.227	21.804	1.000	0.984	0.962	2.422	2.611	0.082	0.027	0.006		
Dependent T, V												
No bias	—	—	—	—	—	2.231	2.346	0.071	0.015	0.004		
Uniform	15.819	9.592	0.892	0.738	0.504	2.915	2.884	0.132	0.042	0.008		
Nonuniform	24.407	16.337 ^c	0.968	0.900	0.742	2.549	2.656	0.090	0.029	0.006		
Both biases	38.419	25.924 ^c	0.996	0.988	0.942	3.245	3.331	0.156	0.057	0.018		
Dichotomous V	Overall											
Independent T, V												
No bias	—	—	0.973	0.927	0.791	—	—	0.092	0.027	0.004		
Uniform	21.033	9.114	—	—	—	2.152	2.245	0.064	0.016	0.003		
Nonuniform	19.465	9.407	0.982	0.932	0.768	2.173	2.211	0.064	0.017	0.002		
Both biases	37.769	14.917	1.000	0.996	0.984	2.342	2.352	0.074	0.020	0.004		
Dependent T, V												
No bias	—	—	—	—	—	2.163	2.176	0.062	0.016	0.001		
Uniform	17.065	8.086	0.942	0.848	0.608	2.893	2.809	0.126	0.042	0.006		
Nonuniform	19.969	9.606	0.962	0.900	0.718	2.363	2.395	0.075	0.020	0.004		
Both biases	35.512	13.879	0.996	0.990	0.970	3.152	2.873	0.151	0.043	0.006		

Note. Overall proportions are calculated over eight conditions.

^aMeans, standard deviations, and proportions are calculated over 500 observations. ^bMeans, standard deviations, and proportions are calculated over 500 × 6 = 3,000 observations in Conditions 1, 5, 9, and 13, and over 500 × 5 = 2,500 observations in all other conditions. ^cAfter removal of 2, 2, and 3 outliers in Conditions 3, 7, and 8; before removing outliers, means (M) and standard deviations (SD) are M = 18.518 and SD = 16.399 in Condition 3, M = 25.618 and SD = 25.245 in Condition 7, and M = 42.481 and SD = 65.462 in Condition 8.

TABLE 6
 Iterative Procedure Results of Measurement Bias Detection With the Restricted Factor
 Analysis/Random Slope Parameterization Method

Condition		Biased Items ^a Proportions of True Positives			Unbiased Items ^b Proportions of False Positives		
		$\alpha = .05$	$\alpha = .01$	$\alpha = .001$	$\alpha = .05$	$\alpha = .01$	$\alpha = .001$
Continuous <i>V</i>	Overall	0.945	0.875	0.747	0.068	0.020	0.004
Independent <i>T, V</i>							
No bias	1	—	—	—	0.064	0.018	0.003
Uniform	2	0.970	0.920	0.794	0.064	0.018	0.005
Nonuniform	3	0.886	0.768	0.544	0.061	0.020	0.004
Both biases	4	0.996	0.960	0.960	0.070	0.020	0.005
Dependent <i>T, V</i>							
No bias	5	—	—	—	0.069	0.016	0.004
Uniform	6	0.860	0.720	0.500	0.070	0.024	0.005
Nonuniform	7	0.964	0.898	0.740	0.076	0.019	0.003
Both biases	8	0.996	0.982	0.942	0.069	0.026	0.004
Dichotomous <i>V</i>	Overall	0.966	0.922	0.790	0.058	0.015	0.003
Independent <i>T, V</i>							
No bias	9	—	—	—	0.059	0.016	0.003
Uniform	10	0.974	0.924	0.766	0.061	0.012	0.003
Nonuniform	11	0.952	0.896	0.698	0.057	0.014	0.002
Both biases	12	0.998	0.992	0.984	0.057	0.016	0.002
Dependent <i>T, V</i>							
No bias	13	—	—	—	0.059	0.016	0.001
Uniform	14	0.916	0.836	0.606	0.061	0.018	0.002
Nonuniform	15	0.958	0.896	0.718	0.054	0.015	0.004
Both biases	16	0.996	0.990	0.970	0.056	0.012	0.003

Note. Overall proportions are calculated over eight conditions.

^aProportions are calculated over 500 observations. ^bProportions are calculated over $500 \times 6 = 3,000$ observations in Conditions 1, 5, 9, and 13, and over $500 \times 5 = 2,500$ observations in all other conditions.

membership). In conditions with either uniform or nonuniform bias, the size of simulated bias is modest, as it only accounts for about 5% of the observed variance. This percentage can be considered halfway between “small” and “medium” effect sizes (Cohen, 1988). At the 5% level of significance, overall detection rates are above 95% in all conditions, except for the MGFA method in the conditions with a continuous *V*. Here we observe the negative consequences of categorizing variables through a median-split: loss of information and reduction of effect size (MacCallum, Zhang, Preacher, & Rucker, 2002).

All three methods generally detect nonuniform bias about as well as uniform bias. Still, with independent *T* and *V* nonuniform bias appears harder to detect than uniform bias, whereas with correlated *T* and *V* it is the other way around. The probable cause is that the positive *T*–*V* correlation and the nonuniform bias both favor the same group (e.g., higher scoring subjects). In other conditions, the dependency between *T* and *V* does seem to render bias detection somewhat more difficult.

As in previous studies (Barendse et al., 2010; Woods & Grimm, 2011), we observe inflated Type I error rates when we use a single run procedure. However, the iterative procedure reduces

TABLE 7
 Estimation Bias in the Multigroup Factor Analysis Method

Condition	Intercepts (Intercept Differences Indicate Uniform Bias)						Factor Loadings (Factor Loading Differences Indicate Nonuniform Bias)					
	Group 1			Group 2			Group 1			Group 2		
	M Difference	SD Difference	M SE	M Difference	SD Difference	M SE	M Difference	SD Difference	M SE	M Difference	SD Difference	M SE
Continuous V	Overall	0.121	0.119	0.002	0.170	0.167	0.003	0.117	0.115	-0.011	0.150	0.146
Independent T, V	No bias	0.004	0.127	0.126	0.002	0.134	0.002	0.134	0.140	-0.008	0.136	0.135
	Uniform	-0.008	0.132	0.127	-0.002	0.141	-0.008	0.120	0.119	-0.009	0.136	0.135
	Nonuniform	0.001	0.112	0.113	-0.012	0.169	0.015	0.112	0.110	-0.040	0.162	0.150
	Both biases	0.005	0.114	0.115	0.004	0.160	0.028	0.120	0.112	-0.028	0.161	0.153
Dependent T, V	No bias	0.005	0.123	0.121	0.003	0.170	-0.005	0.116	0.116	-0.003	0.141	0.135
	Uniform	0.015	0.130	0.127	-0.001	0.182	0.000	0.126	0.120	0.005	0.144	0.142
	Nonuniform	-0.013	0.108	0.109	0.009	0.192	-0.025	0.112	0.110	0.023	0.158	0.154
	Both biases	-0.018	0.118	0.112	0.015	0.214	0.029	0.109	0.112	-0.035	0.163	0.160
Dichotomous V	Overall	0.001	0.119	0.117	-0.003	0.167	0.000	0.114	0.168	-0.001	0.150	0.146
Independent T, V	No bias	-0.006	0.121	0.126	0.001	0.142	0.004	0.116	0.117	0.000	0.141	0.134
	Uniform	0.004	0.135	0.126	0.007	0.147	-0.006	0.125	0.117	-0.008	0.144	0.134
	Nonuniform	-0.009	0.104	0.107	0.001	0.164	0.000	0.099	0.107	-0.007	0.159	0.157
	Both biases	0.009	0.114	0.108	0.000	0.160	0.003	0.110	0.107	-0.010	0.162	0.156
Dependent T, V	No bias	0.000	0.130	0.127	-0.014	0.172	-0.001	0.121	0.118	0.003	0.135	0.134
	Uniform	0.003	0.128	0.126	0.007	0.158	0.006	0.122	0.117	0.000	0.141	0.134
	Nonuniform	0.000	0.108	0.108	-0.022	0.204	-0.003	0.108	0.107	0.008	0.159	0.159
	Both biases	0.007	0.110	0.107	-0.002	0.192	0.000	0.110	0.107	0.007	0.158	0.158

Note. All means and standard deviations are calculated over 500 observations (first item only). Overall means and standard deviations are calculated over eight conditions.

TABLE 8
 Estimation Bias in the Restricted Factor Analysis/Latent Moderated Structures Method

Condition		Parameter <i>b</i> Indicating Uniform Bias			Parameter <i>c</i> Indicating Nonuniform Bias			
		<i>M</i>	<i>SD</i>	<i>M SE</i>	<i>M</i>	<i>SD</i>	<i>M SE</i>	
		Difference	Difference		Difference	Difference		
Continuous <i>V</i>	Overall	-0.001	0.080	0.077	-0.017	0.072	0.068	
Independent <i>T, V</i>								
	No bias	1	0.000	0.072	0.069	0.004	0.076	0.072
	Uniform	2	0.005	0.072	0.069	-0.002	0.076	0.072
	Nonuniform	3	-0.003	0.075	0.072	-0.036	0.081	0.073
	Both biases	4	-0.002	0.078	0.071	-0.036	0.076	0.073
Dependent <i>T, V</i>								
	No bias	5	0.000	0.086	0.084	0.004	0.069	0.063
	Uniform	6	-0.004	0.085	0.084	0.001	0.069	0.063
	Nonuniform	7	-0.003	0.087	0.085	-0.031	0.064	0.065
	Both biases	8	0.003	0.082	0.084	-0.039	0.067	0.065
Dichotomous <i>V</i>	Overall	-0.001	0.076	0.075	0.000	0.079	0.076	
Independent <i>T, V</i>								
	No bias	9	0.003	0.073	0.070	0.001	0.075	0.073
	Uniform	10	0.003	0.072	0.070	-0.001	0.080	0.073
	Nonuniform	11	0.004	0.074	0.073	-0.003	0.071	0.074
	Both biases	12	-0.005	0.073	0.073	-0.004	0.075	0.074
Dependent <i>T, V</i>								
	No bias	13	-0.005	0.080	0.077	0.002	0.080	0.078
	Uniform	14	-0.002	0.074	0.076	0.000	0.082	0.078
	Nonuniform	15	-0.006	0.078	0.080	0.004	0.081	0.080
	Both biases	16	0.001	0.084	0.079	0.000	0.084	0.079

Note. All means and standard deviations are calculated over 500 observations (first item only). Overall means and standard deviations are calculated over eight conditions.

the proportions of false positives to values that are close to the nominal levels of significance. This is true for both the MGFA and the RFA methods.

Most remarkable is the equivalence of the results that are found with RFA/LMS and RFA/RSP methods. Comparison of Tables 4 and 5 shows that the detection rates of the RFA/LMS and RFA/RSP methods are almost identical, for true positives as well as for false positives. Moreover, comparison of Tables 8 and 9 shows that the RFA methods' accuracy and efficiency of the parameter estimates are almost identical as well. A possible explanation for these unexpected findings is that the two RFA methods, as different as they are conceptually, do utilize the same information in the same robust maximum likelihood estimation method.

The equivalence of RFA/LMS and RFA/RSP results is one reason why we do not report the iterative procedure results for the RFA/LMS method. Another reason is that we encounter convergence problems with the RFA/LMS method, apparent from inadmissible solutions, negative chi-square differences, or no convergence at all. Only sometimes convergence problems are solved by trying different starting values, and only sometimes the negative chi-square differences disappear with the strictly positive robust chi-square test described by Satorra

TABLE 9
 Estimation Bias in the Restricted Factor Analysis/Random Slope Parameterization Method

Condition		Parameter <i>b</i> Indicating Uniform Bias			Parameter <i>c</i> Indicating Nonuniform Bias			
		<i>M</i>	<i>SD</i>	<i>M SE</i>	<i>M</i>	<i>SD</i>	<i>M SE</i>	
		Difference	Difference		Difference	Difference		
Continuous <i>V</i>	Overall	-0.001	0.079	0.077	-0.017	0.072	0.068	
	Independent <i>T, V</i>							
	No bias	1	0.000	0.072	0.069	0.004	0.076	0.072
	Uniform	2	0.004	0.071	0.069	-0.002	0.075	0.072
	Nonuniform	3	-0.003	0.075	0.071	-0.036	0.081	0.073
	Both biases	4	0.003	0.078	0.071	-0.036	0.076	0.073
	Dependent <i>T, V</i>							
	No bias	5	0.000	0.085	0.083	0.004	0.069	0.062
	Uniform	6	-0.006	0.084	0.083	0.001	0.069	0.063
	Nonuniform	7	-0.003	0.086	0.084	-0.031	0.064	0.065
	Both biases	8	0.000	0.082	0.084	-0.039	0.067	0.065
Dichotomous <i>V</i>	Overall	-0.002	0.076	0.074	0.000	0.078	0.076	
	Independent <i>T, V</i>							
	No bias	9	0.003	0.073	0.069	0.001	0.075	0.073
	Uniform	10	0.002	0.071	0.069	-0.001	0.079	0.073
	Nonuniform	11	0.004	0.074	0.073	-0.004	0.071	0.074
	Both biases	12	-0.006	0.073	0.072	-0.004	0.075	0.074
	Dependent <i>T, V</i>							
	No bias	13	-0.005	0.080	0.076	0.002	0.080	0.078
	Uniform	14	-0.004	0.074	0.076	0.000	0.082	0.078
	Nonuniform	15	-0.006	0.078	0.080	0.004	0.081	0.080
	Both biases	16	-0.001	0.083	0.079	0.000	0.084	0.079

Note. All means and standard deviations are calculated over 500 observations (first item only). Overall means and standard deviations are calculated over eight conditions.

and Bentler (2010; also see Asparouhov & Muthén, 2010). We presume that the problems with the RFA/LMS method could be caused by our (unjustified) application of the method to the interaction between a latent variable and an observed variable, even though LMS assumes interaction between two latent variables. Perhaps we should consider it remarkable that the RFA/LMS performs flawlessly with 99.2% of our 8,000 data sets in the single run procedure.

Comparison of the accuracy and efficiency of the RFA methods (Tables 8 and 9) with the MGFA method (Table 7) shows less estimation bias and, especially, better efficiency of the RFA methods in all conditions. This seems to confirm the advantage of RFA having a smaller ratio of number of parameters and sample size, although the better performance might also be due to the robust maximum likelihood estimation method of RFA/LMS and RFA/RSP that utilizes more information than the usual maximum likelihood estimation method of MGFA.

In conclusion, we prefer the RFA methods to the MGFA method because of their better bias detection rates, especially with respect to continuous violators, and because of their higher accuracy and efficiency of parameter estimation. Although laborious, we further prefer the

iterative procedure to the single run procedure because of the lower Type I error rates. Finally, we cautiously prefer the RFA/RSP method to the RFA/LMS method, because in this study it is more appropriate and runs trouble-free. However, we cannot rule out the possibility that the RFA/LMS method does better in other situations, not simulated in this study, with measurement bias with respect to latent violator variables. Moreover, in future research we would also like to consider alternative estimation methods for RFA/LMS, such as Klein's (2007) quasi-maximum likelihood (Klein & Muthén, 2007).

Surely, the concerted RFA methods have the advantages over MGFA that bias can be investigated with respect to any possible violator, observed or latent, discrete or continuous, that bias can be investigated without the necessity of categorizing continuous violators to divide samples into subsamples, and that bias can thus be investigated with respect to multiple possible violators simultaneously.

REFERENCES

- Asparouhov, T., & Muthén, B. O. (2010). Computing the strictly positive Satorra–Bentler chi-square test in *Mplus* (*Mplus* Web Notes No. 12). Retrieved September 1, 2010, from <http://www.statmodel.com/examples/webnotes/webnote12.pdf>
- Barendse, M. T., Oort, F. J., & Garst, G. J. A. (2010). Using restricted factor analysis with latent moderated structures to detect uniform and nonuniform measurement bias: A simulation study. *Advances in Statistical Analysis, 94*, 117–127.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- González-Romá, V., Hernández, A., & Gómez -Benito, J. (2006). Power and Type I error of the mean and covariance structure analysis model for detecting differential item functioning in graded response items. *Multivariate Behavioral Research, 41*, 29–53.
- Hernández, A., & González-Romá, V. (2003). Evaluating the multiple-group mean and covariance structure model for the detection of differential item functioning in polytomous ordered items. *Psicothema, 15*, 322–327.
- Klein, A. G. (2007). *QuasiML 3.10—Quick reference manual*. Unpublished manuscript, University of Western Ontario, Waterloo, Canada.
- Klein, A. G., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika, 65*, 457–474.
- Klein, A. G., & Muthén, B. O. (2007). Quasi maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects. *Multivariate Behavioral Research, 42*, 647–673.
- MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*, 19–40.
- Meade, A. W., & Lautenschlager, G. J. (2004). A Monte Carlo study of confirmatory factor analytic tests of measurement equivalence/invariance. *Structural Equation Modeling, 11*, 60–72.
- Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research, 13*, 127–143.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525–543.
- Moosbrugger, H., Schermelleh-Engel, K., Kelava, A., & Klein, A. G. (2009). Testing multiple nonlinear effects in structural equation modelling: A comparison of alternative estimation approaches. In T. Teo & M. S. Khine (Eds.), *Structural equation modeling in educational research: Concepts and applications* (pp. 103–136). Rotterdam, Netherlands: Sense Publishers.
- Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika, 54*, 557–585.
- Muthén, B. O., & Asparouhov, T. (2003). Modeling interactions between latent and observed continuous variables using maximum-likelihood estimation in *Mplus* (*Mplus* Web Notes No. 6). Retrieved August 30, 2010, from <http://www.statmodel.com/download/webnotes/webnote6.pdf>

- Muthén, B. O., & Muthén, L. K. (2007). *Mplus user's guide: Statistical analysis with latent variables*. Los Angeles, CA: Muthén & Muthén.
- Navas-Ara, M. J., & Gómez-Benito, J. (2002). Effects of ability scale purification on identification of DIF. *European Journal of Psychological Assessment, 18*, 9–15.
- Oort, F. J. (1991). Theory of violators: Assessing unidimensionality of psychological measures. In R. Steyer, K. F. Wender, & K. F. Widaman (Eds.), *Psychometric methodology* (pp. 377–381). Stuttgart, Germany: Gustav Fischer.
- Oort, F. J. (1992). Using restricted factor analysis to detect item bias. *Methodika, 6*, 150–166.
- Oort, F. J. (1998). Simulation study of item bias detection with restricted factor analysis. *Structural Equation Modeling, 5*, 107–124.
- Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*, 507–514.
- Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika, 75*, 243–248.
- Schermelleh-Engel, K., Werner, C. S., Klein, A. G., & Moosbrugger, H. (2010). Nonlinear structural equation modeling: Is partial least squares an alternative? *Advances in Statistical Analysis, 9*, 157–166.
- Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review, 18*, 210–222.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices and recommendations for organizational research. *Organizational Research Methods, 3*, 4–70.
- Woods, C. M., & Grimm, K. J. (2011). Testing for nonuniform differential item functioning with multiple indicator multiple cause models. *Applied Psychological Measurement, 35*, 536–556.
- Yuan, K. H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. In M. E. Sobel & M. P. Becker (Eds.), *Sociological methodology 2000* (pp. 165–200). Washington, DC: ASA.

APPENDIX A
EXAMPLE MPLUS SCRIPT FOR FITTING THE RFA/LMS MODEL

```

title: detection of uniform and nonuniform bias in the first item;
data: file is cond2r3.dat ;
variable: names are x1 x2 x3 x4 x5 x6 v1;
analysis: type = random;
algorithm = integration;
! adding MCONVERGENCE improves estimation in a single data set
model:
ability by x1*0.5 x2*.5 x3*.5 x4*.5 x5*.5 x6*.5;
violat by v1 @ .9949874; !v1 is indicator for violator variable
ability with violat *0.26337; !estimate correlation between trait and violator
ability @1;
violat @1;
[ability @ 0];
[violat @ 0];
[v1 @ 0];
!alt option: set free to be estimated
v1 @ .01; !fix residuals of the violator at .01
abxvio | ability xwith violat; !introduce interaction term
x1 0N violat conXvio; !estimate effects to detect bias
x2 0N violat@0 abxvio @0; !alt option: not mention these effects
x3 0N violat@0 abxvio @0; !alt option: not mention these effects
x4 0N violat@0 abxvio @0; !alt option: not mention these effects
x5 0N violat@0 abxvio @0; !alt option: not mention these effects
x6 0N violat@0 abxvio @0; !alt option: not mention these effects
output: tech1 tech8 tech9;
savedata:
results are cond2r3.res; !save the results

```

APPENDIX B
EXAMPLE MPLUS SCRIPT FOR FITTING THE RFA/RSP MODEL

```

title: detection of uniform and nonuniform bias in the first item;
data: file is cond2r3.dat ;
variable: names are x1 x2 x3 x4 x5 x6 v1;
usev = x1 x2 x3 x4 x5 x6 v1 vio;
define: violat = v1;
analysis: type = random;
model:
  ability by x1*.5 x2*.5 x3*.5 x4*.5 x5*.5 x6*.5 ;
  ability @1 ;
  [ability @0];
  ability with violat *.26337;    !estimate correlation between trait and violator
  [violat @ 0] ;
  violat @ 1 ;
  x1 on v1; s1 | x1 on v1; s1 on ability; s1@0; [s1@0];    !estimate effects to detect bias
  x2 on v1 @0; s2 | x2 on v1; s2 on ability @0; s2@0; [s2@0] ;    !alt option: not mention these effects
  x3 on v1 @0; s3 | x3 on v1; s3 on ability @0; s3@0; [s3@0] ;    !alt option: not mention these effects
  x4 on v1 @0; s4 | x4 on v1; s4 on ability @0; s4@0; [s4@0] ;    !alt option: not mention these effects
  x5 on v1 @0; s5 | x5 on v1; s5 on ability @0; s5@0; [s5@0] ;    !alt option: not mention these effects
  x6 on v1 @0; s6 | x6 on v1; s6 on ability @0; s6@0; [s6@0] ;    !alt option: not mention these effects
output: tech1 tech8 tech9;
savedata:
  results are cond2r3.res;    !save the results

```