Random permutation tests of nonuniform differential item functioning in multigroup item factor analysis

Kite, B.A.; Jorgensen, T.D.; Chen, P.-Y.

DOI
10.1007/978-3-319-77249-3_24

Publication date
2018

Document Version
Final published version

Published in
Quantitative Psychology

Link to publication

Citation for published version (APA):
Kite, B. A., Jorgensen, T. D., & Chen, P-Y. (2018). Random permutation tests of nonuniform differential item functioning in multigroup item factor analysis. In M. Wiberg, S. Culpepper, R. Janssen, J. González, & D. Molenaar (Eds.), *Quantitative Psychology: The 82nd Annual Meeting of the Psychometric Society, Zurich, Switzerland, 2017* (pp. 287-296). (Springer Proceedings in Mathematics & Statistics; Vol. 233). Springer. https://doi.org/10.1007/978-3-319-77249-3_24

Random Permutation Tests of Nonuniform Differential Item Functioning in Multigroup Item Factor Analysis

Benjamin A. Kite, Terrence D. Jorgensen and Po-Yi Chen

Abstract The purpose of the present research was to introduce and evaluate random permutation testing applied to measurement invariance testing with ordered-categorical data. The random permutation test builds a reference distribution from the observed data that is used to calculate a p value for the observed $(\Delta)\chi^2$ statistic. The reference distribution is built by repeatedly shuffling the grouping variable and then saving the $\Delta\chi^2$ statistic between the two models fitted to the resulting data. The present research consisted of two Monte Carlo simulations. The first simulation was designed to evaluate random permutation testing across a variety of conditions with scalar invariance testing in comparison to an existing analytical solution: the robust mean- and variance-adjusted $\Delta\chi^2$ test. The second simulation was designed to evaluate the random permutation test applied to testing configural invariance by evaluating overall model fit (the χ^2 fit statistic). Simulation results and suggestions for the use of the random permutation test are provided.

Keywords Measurement invariance · Differential item functioning
Ordered-categorical data · Permutation · Multiple group confirmatory
factor analysis

B. A. Kite · P.-Y. Chen
University of Kansas, Lawrence, KS, USA
e-mail: bakite@ku.edu

P.-Y. Chen
e-mail: p090c021@ku.edu

T. D. Jorgensen (✉)
University of Amsterdam, Nieuwe Achtergracht, 127 (Room D7.17),
1018 WS Amsterdam, The Netherlands
e-mail: T.D.Jorgensen@uva.nl

M. Wiberg et al. (eds.), *Quantitative Psychology*, Springer Proceedings
in Mathematics & Statistics 233, https://doi.org/10.1007/978-3-319-77249-3_24

1 Introduction

Behavioral researchers often use multiple-group confirmatory analysis (MG-CFA) to test measurement invariance (MI) with indicator variables on a Likert-type scale. The procedure of testing MI can be seen as a procedure of finding items with differential item functioning (DIF). In a MG-CFA framework, testing MI with ordinal data usually involves comparing nested invariance models. To test hypotheses about different levels of invariance, researchers could first use the ordinal estimators based on polychoric correlations from software such as M*plus* and `lavaan`, which employ diagonally weighted least squares (DWLS) estimation. A robust a mean- and variance-adjusted test statistic can be requested in M*plus* using the command “ESTIMATOR = WLSMV” or from `lavaan` using the argument `estimator` = “`WLSMV`”, where the “MV” stands for the mean and variance adjustment to the chi-squared test statistic. MI testing can be conducted by comparing the global fit indices such as chi-squared statistic (χ^2) or alternative fit indices (AFI) between invariance models. Among different criteria developed for MI testing, researchers have found that the chi-squared difference ($\Delta\chi^2$) test substantially outperforms other fixed cutoffs based on change in AFI (e.g., change in CFI) by showing greater power and a better ability to control Type I error rate across different scenarios (Sass et al. 2014).

The $\Delta\chi^2$ tests of ordinal estimators in MG-CFA usually require researchers to apply robust corrections during the testing procedures to mitigate the influences of not using consistent estimators for the weight matrix in fit function (Savalei 2014). Software such as M*plus* (Muthén and Muthén 2015) and `lavaan` (Rosseel 2012) both provide robust $\Delta\chi^2$ tests for researchers to compare invariance models estimated by DWLS. Unfortunately, even though robust $\Delta\chi^2$ tests are considered best practice for testing MI with ordinal data in MG-CFA, there are some important issues that warrant further attention.

Most simulation research of the mean- and variance-adjusted $\Delta\chi^2$ test utilizes the implementation provided by M*plus* with the DIFFTEST command when using ESTIMATOR = WLSMV. Researchers have found contradictory conclusions during simulations about its ability to control Type I error rate (see the following sections for details). The mean- and variance-adjusted $\Delta\chi^2$ test is also implemented in `lavaan` via the `lavTestLRT` function, but it has not been examined in a published Monte Carlo simulation. Furthermore, the corrected χ^2 statistic obtained through WLSMV also has been shown to be inappropriate to test the configural invariance assumption (whether the item-factor configurations are identical across groups) when the model is only an approximation of the true population model (Jorgensen et al. 2017), but evidence of inflated Type I error rates under certain conditions (Bandalos 2014) suggests that a test of overall model fit could yield inflated Type I errors even when models fit perfectly.

To address these issues, in the current study, we propose a nonparametric method for testing MI based on the permutation test. We compare the robust $(\Delta)\chi^2$

tests provided by M*plus* and `lavaan` with two simulation studies. Through these simulations, we provide researchers (a) explanations about contradictory conclusions in previous studies about the robust $\Delta\chi^2$ test in M*plus*, (b) systematic evaluations of the robust $\Delta\chi^2$ test provided by `lavaan`, and (c) a new solution that can outperform robust $\Delta\chi^2$ test under conditions when it fails to yield nominal error rates. The rest of this article is organized as follows. We first briefly introduce the robust $(\Delta)\chi^2$ tests provided by M*plus* and `lavaan`, then explain their problems in MI testing. After that, we illustrate the rationale of the permutation test we propose and explain its theoretical advantages. Lastly, we investigate the relative performances between methods through our simulations and provide recommendations for researchers.

2 The Robust $\Delta\chi^2$ Test in M*plus* for Testing MI with Ordinal Data

The robust $\Delta\chi^2$ test provided by M*plus* is a widely used implementation for MI testing with ordinal data in MG-CFA recommended by popular structural equation modeling textbooks (e.g., Kline 2016; Little 2013). Muthén and Muthén (2015) suggested that researchers use the DIFFTEST command in M*plus* in order to correctly scale $\Delta\chi^2$. The DIFFTEST command in M*plus* applies the mean and variance adjustment to the $\Delta\chi^2$ statistic between nested models, as discussed by Asparouhov and Muthén (2006; see also Satorra 2000). The parent model (e.g., a configural model) is fitted to the data, and matrices containing information about the model are saved in a separate output file. When the nested model (e.g., a scalar invariance model) is fitted and the text file containing matrices from the parent model is provided, DIFFTEST uses information from both models to compute a “scaled and shifted” $\Delta\chi^2$ statistic that asymptotically yields nominal Type I error rates. A more detailed explanation of the computation involved with the DIFFTEST command can be found in Asparouhov and Muthén (2006).

3 The Robust $\Delta\chi^2$ Test in `lavaan` for Testing MI with Ordinal Data

Besides M*plus*, empirical researchers could also use the “`lavTestLRT`” function provided by `lavaan` for MI testing (Rosseel 2012). When two nested models are supplied to the `lavTestLRT` function, the correction outlined by (Satorra 2000) is applied to produce a mean- and variance-adjusted $\Delta\chi^2$ statistic. Within the `lavTestLRT` function in `lavaan`, there are two options for how to compute the

Jacobian of the constraint function. The first option (`method` = “`exact`”) is to calculate an exact solution from a constraint function applied to the full parameter vector, which requires that the two models are nested in the parameter sense, not the more flexible sense of nested covariance structures (Bentler and Satorra 2010). The second option (`method` = “`delta`”) provides an approximation to the Jacobian and only requires models to be nested in covariance sense, such that the set of predictions that could possibly be made by the parent model include all possible predictions made by the nested model. In the present research, we used the second option, which is `lavaan`’s default method beginning with version 0.6-1.1109.

4 Problems with Currently Available Methods

Asparouhov and Muthén (2006) conducted a small simulation to show that their robust $\Delta\chi^2$ test effectively controls the Type I error rate when the total sample sizes are asymptotically large: 1100 and 2200. A follow-up study conducted by Sass et al. (2014) found contradicting results when sample sizes were more realistically small or moderate. Specifically, Sass et al. found that the Type I error rate of the robust $\Delta\chi^2$ test provided by M*plus* was always substantially inflated in all of their conditions with symmetrically distributed thresholds (range from 7–9%), and 6–9% in asymmetric conditions. One explanation to these contradicting results could be that the sample sizes that Sass et al. examined are in general smaller than the sample sizes in Asparouhov and Muthén (2006), and small samples are inconsistent with the derivation of the robust test statistic, which relies on asymptotic theory. However, if the $\Delta\chi^2$ statistic obtained from WLSMV requires more than 1000 observations, then its applicability will be severely limited, considering most of MI studies in psychology won’t have this large of sample size (Putnick and Bornstein 2016).

After thoroughly examining the results in Sass et al. (2014), we found another possible explanation. That is, in their simulations the scalar invariance model was different from the ordinary settings by unnecessarily constraining two additional parameters. Specifically, to make sure the configural model was identified, Sass et al. fixed the mean and variance of latent factor to 0 and 1 in both groups. When estimating the scalar invariance model, Sass et al. did not release these two constraints in the second group as suggested in literature, which resulted in an overly stringent scalar invariance model (Kline 2016; Little 2013). We believe this could be another reason that caused their inflated Type I error rates.

According to our knowledge, there is still no study evaluating the performance of the `lavTestLRT` function in `lavaan`, despite its use by empirical researchers (e.g., Antoniadou et al. 2016). Note that Satorra (2000) originally proposed the adjustment for the $\Delta\chi^2$ statistic to correct for continuous non-normal data, not categorical data. The utility of this $\Delta\chi^2$ correction with ordinal estimators like

WLSMV seems to rest quite heavily on the asymptotic assumption. We therefore think it is worthwhile to conduct a simulation to compare different implementations of the correction that might not be equivalent in small to moderate samples, such as the DIFFTEST procedure in M*plus* and the `lavTestLRT` procedure in `lavaan`.

Finally, besides the unsolved issues we mentioned for the robust $\Delta\chi^2$ test in M*plus* and `lavaan`, we believed there is also a common limitation shared by the robust χ^2 obtained from the WLSMV estimator in both software packages. Specifically, we believe the χ^2 obtained from WLSMV estimator might not be a valid statistic for evaluating the configural invariance in small to moderate samples because it is derived from asymptotic theory. Bandalos (2014) found inflated Type I error rates for the robust χ^2 statistic when the sample size is small, especially when thresholds are asymmetrically distributed.

5 Permutation Tests of MI with Ordinal Data

To solve the problem of $(\Delta)\chi^2$ test statistics mentioned above, we proposed a permutation test of MI with ordinal data, which would be free from asymptotic theory and should be able to control the Type I error rate reasonably well regardless of the sample size and distribution of the thresholds. Specifically, we propose to apply the random permutation testing to $(\Delta)\chi^2$ with ordered-categorical data to overcome the issue of the difference statistic not following a central χ^2 distribution. The focus of the present research is demonstrating how this approach works and evaluating its performance. The proposed random permutation test is a nonparametric method based on the idea of building an empirical reference distribution reflecting the null hypothesis that groups have the same model configuration and measurement parameters. In other words, the reference distribution is built under the assumption of a true null hypothesis that there is no effect of group membership on measurement properties (e.g., configuration, parameter values). This reference distribution is used to calculate a *p* value when testing the null hypothesis of invariance. The benefit of permutation testing is that building a nonparametric reference distribution alleviates many of the assumptions of standard parametric hypothesis tests. When testing for the effect of group membership on a test statistic, a null distribution can be built by randomly shuffling the grouping variable and saving the resulting test statistic after each shuffle. If there is no difference in measurement-model configurations or parameters between groups, the observed test statistic (calculated from the original data) should be consistent with the values created by randomly shuffling the grouping variable; that is, the observed value would only exceed the upper 95th percentile of the permuted values 5% of the time. This should keep the Type I error rate of the test procedure nominal (i.e., at 5% when using $\alpha = 0.05$). Building a null distribution this way is especially useful when the distribution of the test statistic is unknown.

6 Method

To address the issues of the currently available two methods mentioned in the introduction, we conducted two Monte Carlo studies. Study 1 is designed to compare the relative performances between the robust $\Delta\chi^2$ test provided by M*plus*, the robust $\Delta\chi^2$ test provided by `lavaan`, and our new proposed permutation method on detecting DIF. In Study 1, based on the assumption that researchers have confirmed configural invariance hypothesis, we conducted the $\Delta\chi^2$ tests between scalar and configural invariance model with the three methods above. The relative performances between methods were evaluated in terms of Type I error rate and power across 1000 replications within each condition. In simulation Study 2 we focused on the performance of the Type I error rate the χ^2 obtained from the three methods. In Study 2 we examined whether the corrected χ^2 provided in M*plus* and `lavaan` would reject the configural invariance model too often in comparison to the permutation method we proposed. In both simulations, we follow Sass et al. (2014) and used (0.036, 0.064) as the acceptable range for observed Type I error rates, In both simulations, data were generated in R using the `simulateData` function in `lavaan`. A two-group, single-factor, model with eight indicator variables was used as the population model. The factor loadings were fixed at 0.6 except in conditions when loadings were not invariant (i.e., when the loadings of first two items in Group 2 were different from Group 1). Residual variances for indicator variables were always set at $1 - \lambda^2$ so that latent item responses would have unit variance. The number of shuffling with each permutation test was set to be 500. The design factors we manipulated in the two simulations (i.e., sample size, distribution of thresholds, the number of categories per item, and the presence of measurement non-invariance) are illustrated as follows.

Study 1 evaluated the random permutation $\Delta\chi^2$ against analytically derived robust $\Delta\chi^2$ test statistics. The simulation design was a fully crossed 2 (response categories) × 2 (threshold symmetry) × 2 (sample size) × 2 (factor loading invariance) design resulting in 16 between-replication conditions used to generate data, each having 1000 replications. In each replication, four different $\Delta\chi^2$ tests were conducted: robust $\Delta\chi^2$ tests in M*plus* and `lavaan`, our permutation test for $\Delta\chi^2$, and an unadjusted $\Delta\chi^2$ test as a reference.

In Study 1, we set the sample size as 300 (150 per group) or 600 (300 per group). These settings are similar to the small and medium sample sizes Sass et al. (2014) used. The number of categories per item was set to be 2 or 5 to represent the dichotomous and ordinal scales that researchers frequently used in practice. In addition, we also simulated either symmetrically or asymmetrically distributed thresholds, given that previous studies have found that he distribution of thresholds could affect the results of $\Delta\chi^2$ related tests (e.g., Sass et al. 2014). Specifically, in conditions with ordinal items, the symmetric and asymmetric thresholds are set to be (−1.30, −0.47, 0.47, 1.30) and (−0.25, 0.38, 0.84, 1.28) as used by Sass et al. (2014). Threshold values for symmetrically and asymmetrically dichotomous items are set to be 0 and 0.7 respectively, as the average of the conditions manipulated in previous research (Beauducel and Herzberg 2006; Rhemtulla et al. 2012). The non-invariance we manipulated in the current study is limited to factor loadings. Specifically, in Study 1 we created non-invariance by subtracting 0.25 (Sass et al. 2014) from the factor loadings for Items 1 and 2 in the population model in the focal group. Specifically, in non-invariant conditions, the factor loadings of Items 1 and 2 in the model will be 0.60 in the reference group but were 0.6 − 0.25 = 0.35 in the focal group. In contrast, Items 3–8 in both groups always had factor loadings of 0.60 in all conditions.

There were two models compared in each replication: a configural invariance model and a scalar invariance model. The configural model had the factor loadings and thresholds freely estimated for both groups, whereas the latent variable in each group had its estimated mean and variance fixed to be 0 and 1, respectively. Further, in the configural model, the variances of the latent response variables (i.e., scales of normally distributed responses assumed to underlie observed discrete item responses) were fixed to 1 in both groups (i.e., we used the so-called "delta" method of identification available in M*plus* and `lavaan`). The scalar invariance model had the factor loadings and thresholds constrained to equality across groups. Constraining the measurement parameters across groups allowed the latent variable mean and variance to be estimated in the focal group rather than fixed to 0 and 1.

The simulation conditions of Study 2 are almost identical to those of Study 1 except we removed the non-invariant conditions and the estimation of scalar invariance, given the exclusive focus on Type I error rates of the χ^2 statistic for the configural invariance model. Additionally, in order to increase the magnitude of asymmetry in our data to better match the work of Bandalos (2014), we changed the distribution of asymmetric thresholds to (1.198) and (0.85, 1.10, 1.45, and 2.00).

7 Results

Type I error rates for tests of scalar invariance are shown in Table 1. Results showed that random permutation testing and `lavTestLRT` had reasonable Type I error control. The random permutation test had Type I errors within the nominal range of 0.036–0.064 in all eight equal measurement parameter conditions, whereas the M*plus* DIFFTEST procedure had inflated error rates in the two conditions where there were two response options with asymmetric thresholds, even though the inflation is not as severe as Sass et al. (2014) found with ordinal data.

Power for scalar invariance tests are shown in Table 2. The M*plus* DIFFTEST procedure consistently showed the highest power, with `lavTestLRT` showing power equal to or greater than the random permutation test (see Table 2). All testing procedures showed higher power in conditions higher group sizes, more response categories, and symmetric thresholds.

The results of simulation Study 2 in Table 3 showed that the random permutation test of configural invariance had acceptable Type I error control in all eight study conditions. The mean- and variance-adjusted χ^2 tests provided by M*plus* and

Table 1 Type I error rates for $\Delta\chi^2$ tests

N	# Categories	Thresholds	Permutation	M*plus*	lavaan	Unadjusted
150	2	Symmetric	0.050	0.060	0.056	0.143
300			0.043	0.052	0.050	0.128
150	5		0.053	0.062	0.054	0.131
300			0.053	0.057	0.053	0.098
150	2	Asymmetric	0.053	0.065	0.054	0.135
300			0.056	0.078	0.065	0.139
150	5		0.050	0.053	0.047	0.131
300			0.054	0.062	0.056	0.128

Table 2 Power for $\Delta\chi^2$ tests

N	# Categories	Thresholds	Permutation	M*plus*	lavaan	Unadjusted
150	2	Symmetric	0.279	0.319	0.292	0.452
300			0.543	0.568	0.543	0.703
150	5		0.460	0.504	0.464	0.618
300			0.786	0.811	0.794	0.890
150	2	Asymmetric	0.214	0.258	0.225	0.361
300			0.406	0.457	0.427	0.588
150	5		0.342	0.370	0.335	0.519
300			0.707	0.733	0.712	0.831

Table 3 Type I error rates of χ^2 test in the configural invariance model

N	# Categories	Thresholds	Permutation	M*plus*	lavaan	Unadjusted
150	2	Symmetric	0.051	0.049	0.049	0.001
300			0.048	0.052	0.052	0.001
150	5		0.054	0.066	0.066	0.000
300			0.057	0.059	0.059	0.000
150	2	Asymmetric	0.047	0.049	0.049	0.006
300			0.039	0.051	0.052	0.004
150	5		0.049	0.202	0.199	0.014
300			0.035	0.100	0.101	0.002

lavaan performed nearly identically and showed inflated Type I errors in conditions with asymmetric thresholds with five response options. The error rates were especially inflated with five response options when the group sizes were 150 (20.2% and 19.9%), and improved but still inflated when the group sizes were 300 (10% and 10.1%). Lastly, the unadjusted χ^2 test provided by lavaan showed error rates well below the nominal value of 0.05 in all conditions.

8 Discussion

The purpose of the present research was to evaluate the use of random permutation testing applied to $\Delta\chi^2$ tests with ordered-categorical indicator variables. The research was focused on models estimated with the popular WLSMV estimator. When models with ordered-categorical data are estimated with WLSMV, the $\Delta\chi^2$ related tests require a mean and variance adjustment (Asparouhov and Muthén 2006; Satorra 2000). The random permutation test was introduced as an alternative that is easily implemented in any statistical software, and as a method that should control Type I errors as well or better than existing methods. Study 1 evaluated the random permutation $\Delta\chi^2$ test for measurement invariance in comparison to existing analytical robust solutions, and served as a follow-up to Sass et al. (2014). Study 2 expanded on the work of Jorgensen et al. (2017) and served as a follow-up to Bandalos (2014).

Overall, the random permutation test performed well in both simulations. In Study 1 the random permutation test was the only method that consistently showed Type I errors within the previously defined nominal range of 0.036 and 0.064. Further, the power of the random permutation test was increased in conditions with higher group sizes, more response categories, and symmetric response distributions. As would be expected based on the better error control, the random permutation test showed slightly less power than M*plus* DIFFTEST and `lavTestLRT`. The modification to the design of Sass et al. (2014) in simulation one did result in a better performance of the M*plus* DIFFTEST procedure. When the latent variable mean and variance were freely estimated in the focal group in the scalar invariance model, Type I error rates for the DIFFTEST procedure were closer to $\alpha = 0.05$ than what was reported by Sass and colleagues.

Study 2 replicated the poor Type I error control, previously reported by Bandalos (2014), of the mean- and variance-adjusted χ^2 when data were extremely asymmetric. The random permutation test showed no performance issues with Type I error control. These results show that random permutation testing should be considered an appropriate option for researchers to test DIF using item factor analysis models.

The present research suggests the random permutation testing procedure could be preferable over the parametric approaches in nonideal conditions (small to moderate samples with asymmetric thresholds) because permutation provides better control of the Type I error rate for both χ^2 and $\Delta\chi^2$ than the M*plus* DIFFTEST procedure or `lavaan`’s `lavTestLRT`.

References

Asparouhov, T., & Muthén, B. (2006). *Robust chi square difference testing with mean and variance adjusted test statistics*. M*plus* Web Notes No. 10. Retrieved from www.statmodel.com.

Antoniadou, N., Kokkinos, C. M., & Markos, A. (2016). Development construct validation and measurement invariance of the Greek cyber-bullying/victimization experiences questionnaire (CBVEQ-G). *Computers in Human Behavior, 65,* 380–390.

Bandalos, D. L. (2014). Relative performance of categorical diagonally weighted least squares and robust maximum likelihood estimation. *Structural Equation Modeling, 21*(1), 102–116.

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling, 13*(2), 186–203.

Bentler, P. M., & Satorra, A. (2010). Testing model nesting and equivalence. *Psychological Methods, 15*(2), 111–123.

Jorgensen, T. D., Kite B. A., Chen P.-Y., & Short S. D. (2017). Finally! A valid test of configural invariance using permutation in multigroup CFA. In L. A. van der Ark, M. Wiberg, S. A. Culpepper, J. A. Douglas, & W.-C. Wang (Eds.), *Quantitative psychology: The 81st annual meeting of the psychometric society, Asheville, North Carolina, 2016* (pp. 93–103). New York, NY: Springer. https://doi.org/10.1007/978-3-319-56294-0_9.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York, NY: Guilford.

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: Guilford Press.

Muthén, L. K., & Muthén, B. O. (1998–2015). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41,* 71–90.

Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*(3), 354–373.

Rosseel, Y. (2012). `lavaan`: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36.

Sass, D. A., Schmitt, T. A., & Marsh, H. W. (2014). Evaluating model fit with ordered categorical data within a measurement invariance framework: A comparison of estimators. *Structural Equation Modeling, 21*(2), 167–180.

Satorra, A. (2000). Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In R. D. H. Heijmans, D. S. G. Pollock, & A. Satorra (Eds.), *Innovations in multivariate statistical analysis: A Festschrift for Heinz Neudecker* (pp. 233–247). London, England: Kluwer Academic Publishers.

Savalei, V. (2014). Understanding robust corrections in structural equation modeling. *Structural Equation Modeling, 21*(1), 149–160.