



UvA-DARE (Digital Academic Repository)

Using product indicators in restricted factor analysis models to detect nonuniform measurement bias

Kolbe, L.; Jorgensen, T.D.

DOI

[10.1007/978-3-319-77249-3_20](https://doi.org/10.1007/978-3-319-77249-3_20)

Publication date

2018

Document Version

Final published version

Published in

Quantitative Psychology

[Link to publication](#)

Citation for published version (APA):

Kolbe, L., & Jorgensen, T. D. (2018). Using product indicators in restricted factor analysis models to detect nonuniform measurement bias. 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. 235-245). (Springer Proceedings in Mathematics & Statistics; Vol. 233). Springer. https://doi.org/10.1007/978-3-319-77249-3_20

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)

Using Product Indicators in Restricted Factor Analysis Models to Detect Nonuniform Measurement Bias



Laura Kolbe and Terrence D. Jorgensen

Abstract When sample sizes are too small to support multiple-group models, an alternative method to evaluate measurement invariance is restricted factor analysis (RFA), which is statistically equivalent to the more common multiple-indicator multiple-cause (MIMIC) model. Although these methods traditionally were capable of detecting only uniform measurement bias, RFA can be extended with latent moderated structural equations (LMS) to assess nonuniform measurement bias. As LMS is implemented in limited structural equation modeling (SEM) computer programs (e.g., *Mplus*), we propose the use of the product indicator (PI) method in RFA models, which is available in any SEM software. Using simulated data, we illustrate how to apply this method to test for measurement bias, and we compare the conclusions with those reached using LMS in *Mplus*. Both methods obtain comparable results, indicating that the PI method is a viable alternative to LMS for researchers without access to SEM software featuring LMS.

Keywords Factor analysis · Product indicators · Measurement invariance
Nonuniform measurement bias

1 Introduction

Measurement bias entails that scales function differently across groups, irrespective of true differences in the construct that the scale was designed to measure. Let T denote the construct of interest measured by a set of observed variables X . Moreover, let V be a set of variables other than T . The formal definition of measurement bias involves a violation of measurement invariance (Mellenbergh 1989):

L. Kolbe (✉) · T. D. Jorgensen

Research Institute of Child Development and Education, University of Amsterdam,
1001 NG, P.O. Box 15776 Amsterdam, The Netherlands
e-mail: L.Kolbe@uva.nl

T. D. Jorgensen

e-mail: T.D.Jorgensen@uva.nl

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

where f_1 is the conditional distribution of X given T and V , and f_2 the conditional distribution of X given T . If measurement invariance holds (i.e., $f_1 = f_2$), the measurement of T by X is invariant with respect to V . But if measurement invariance does not hold (i.e., $f_1 \neq f_2$), the measurement of T by X is biased with respect to V . A distinction can be made between uniform and nonuniform bias. Uniform bias implies that the extent of bias is constant for all levels of the construct T , whereas nonuniform bias implies that the extent of bias varies with T .

A common method to test for measurement bias with respect to a grouping variable is multiple-group confirmatory factor analysis (MGCFA; Vandenberg and Lance 2000), which requires sufficiently large samples for each group. An alternative for testing measurement bias is restricted factor analysis (RFA; Oort 1992, 1998). An advantage of this method over MGCFA is that the potential violator V may be categorical or continuous, observed or latent, and multiple violators can be investigated simultaneously. Moreover, RFA does not require the division of the sample into subsamples by V . The latter advantage comes at the cost of additional assumptions—namely, homogeneity of residual variances across groups.¹ If these additional assumptions hold, RFA should have more power than MGCFA to detect measurement bias.

When using RFA, the potential violator V is added to a common factor model as an exogenous variable that covaries with T . Uniform bias can be assessed by testing the significance of direct effects of V on X . To assess nonuniform bias, an extension for modeling latent interactions is required. RFA is commonly extended with latent moderated structural equations (LMS; Barendse et al. 2010). This allows for assessing nonuniform bias by testing the significance of interaction effects of $T \times V$ on X . Although this method generally has high power to detect measurement bias (Barendse et al. 2010, 2012; Woods and Grimm 2011), a disadvantage is that LMS is only implemented in the commercial structural equation modeling (SEM) software *Mplus* (Muthén and Muthén 2012).² Moreover, most traditional SEM fit indices to test for model fit are not available when using the LMS method in *Mplus*, except for Akaike's Information Criterion (AIC; Akaike 1973) and Bayesian Information Criterion (BIC; Schwartz 1978).

In this chapter, we introduce the product indicator (PI) method to model latent interactions in RFA models. The PI method has received a great deal of attention in the general context of modeling interactions among latent variables in SEM (Henseler and Chin 2010; Lin et al. 2010; Little et al. 2006; Marsh et al. 2004),

¹In traditional RFA models, common-factor variances are also assumed to be equal across groups. However, when extending RFA to include a latent interaction factor with product indicators (described immediately following), differences in common-factor variances can be captured by the covariance between the common factor and the latent interaction factor.

²LMS is also available in the open-source R package `nlsem` (Umbach et al. 2017), but the implementation is very limited. It is not possible to test measurement bias using RFA models in the `nlsem` package, so we do not consider it further.

but has never been studied in light of testing measurement bias. First, we discuss the detection of measurement bias using RFA models, then we introduce the PI method, and finally we demonstrate how to test for measurement bias using RFA with PI by means of an illustrative example. We compare the results of PI to LMS on the same simulated data set.

2 Restricted Factor Analysis

2.1 Detection of Measurement Bias with RFA Models

In RFA models, the construct T can be modeled as a latent factor with multiple measures X as observed indicators. The possible violator V is added to the measurement model as an exogenous single-indicator latent variable and is allowed to covary with the common factor T . The violator V may represent a grouping variable by using a dummy-coded indicator. The observed scores X are modeled as

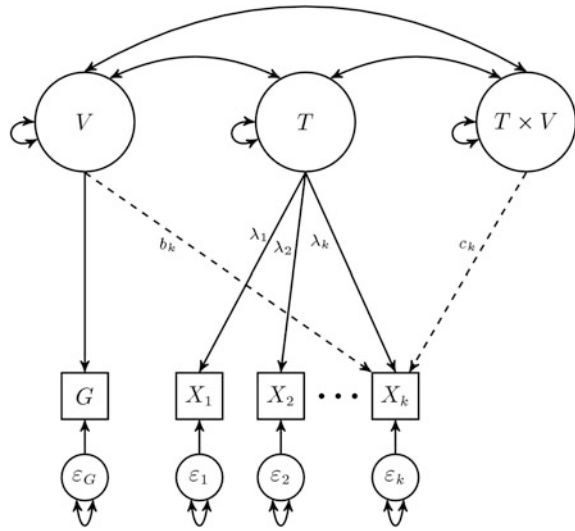
$$x_j = \boldsymbol{\tau} + \boldsymbol{\lambda}t_j + bg_j + ct_jg_j + \boldsymbol{\delta}\boldsymbol{\epsilon}_j \quad (2)$$

where x_j is a vector of observed scores, t_j is the common factor T score, g_j is a dummy code for group membership V , and $\boldsymbol{\epsilon}_j$ is a vector of the residual scores of subject j . Moreover, the vector $\boldsymbol{\tau}$ contains intercepts, $\boldsymbol{\lambda}$ is a vector of factor loadings on the common factor T , and $\boldsymbol{\delta}$ is a vector of residual factor loadings. The vectors b and c are of special interest and contain regression coefficients. A nonzero element in b or c indicates uniform or nonuniform bias, respectively.

Figure 1 illustrates an example of an RFA model to test for measurement bias using two anchor items. The violator V is modeled as a latent variable with a single indicator G representing group membership. For visual simplicity, the measurement model of $T \times V$ is excluded from Fig. 1, but those details are discussed in the following subsection. Measurement bias can be examined by comparing the fit of an unconstrained model with several constrained models. In the unconstrained model, all items are regressed on V and $T \times V$, except for the items in the anchor set. Each constrained model involves fixing the regression of the studied item onto V and $T \times V$ at zero.

The pair of constraints for each item can be tested simultaneously, where the null hypothesis of no measurement bias implies both b and c coefficients corresponding to the studied item are zero in the population. These constraints can be tested via model comparison of a constrained and unconstrained model, producing a likelihood ratio test statistic that is distributed as χ^2 random variable with 2 df . A significant test statistic indicates that the studied item is biased with respect to V , and 1- df follow-up tests of the individual b and c coefficients can reveal whether that indicator's bias is uniform or nonuniform. Our study focuses only on the 2- df omnibus test for each indicator.

Fig. 1 An example of testing measurement bias using an RFA model. Dashed arrows represent effects that may be estimated to test for uniform and nonuniform bias. The indicators X_1 and X_2 serve as anchor items



2.2 Product Indicators

The use of PI to model interactions among latent variables was originated by Kenny and Judd (1984). The PI method involves the specification of a measurement model for the latent interaction factor. Generally, product terms are built by multiplying the indicators of the associated latent variables, which serve as indicators for the latent interaction factor. All indicators, including the product indicators, are assumed to be multivariate normally distributed if the maximum likelihood estimation procedure is used. Because products of normal variables are not themselves normally distributed, this assumption is violated. Thus, a robust maximum likelihood estimator is used to relax this assumption (see Marsh et al. 2004).

Several variants of the PI method have been proposed, among which is the double-mean-centering strategy (Lin et al. 2010) that we implement herein. The double-mean-centering strategy is superior to other strategies because it eliminates the need for a mean structure and does not involve a cumbersome estimation procedure. Although the orthogonalizing and double-mean-centering strategy perform equally well when all indicators are normally distributed, the double-mean-centering strategy performs better when the assumption that all indicators are normally distributed is violated (Lin et al. 2010).

The Double-Mean-Centering Strategy. The first step of the double-mean-centering strategy involves mean-centering the indicators of the latent variables of interest. Each of the mean-centered indicators of one latent variable are multiplied by the mean-centered indicators of the other latent variable. Then, the resulting product indicators are centered at their means and are used as indicators of the latent interaction factor. If the common factor T has I indicators and the violator variable V has J indicators, then the latent interaction factor can have up to $I \times J$ product

indicators, although matching schemes have been proposed to reduce the number of product indicators (Marsh et al. 2004). In RFA, however, these matching schemes would be irrelevant when the common factor only interacts with a single-indicator violator construct (or with multiple single-indicator violators). Figure 2 shows an example of an RFA model with a latent interaction using the PI method. All possible cross-products are used in this example (i.e., each indicator of T is multiplied by the single indicator of V), and all indicators of T and V are centered at their means.³

3 Illustrative Example

We simulated a single data set to demonstrate how to apply the PI method in R (R Core Team 2016) to test for measurement bias, and to compare the conclusions with those reached using LMS. See Barendse et al. (2012) for *Mplus* syntax to apply LMS.

3.1 Data Generation

Data were generated for two groups, each with a group size of $n = 100$. We considered a scale of $k = 10$ items, 40% of which were biased: two uniformly biased items and two nonuniformly biased items. This way, we are able to investigate the performance of LMS and PI using a hypothetical scale with a substantial degree of measurement bias. Item scores of subject j in group g were generated using the following model:

$$x_j = \boldsymbol{\tau}_g + \boldsymbol{\lambda}_g t_j + \boldsymbol{\delta}_g \boldsymbol{\varepsilon}_j \quad (3)$$

where x_j is a vector of 10 item scores, t_j is the common factor score, and $\boldsymbol{\varepsilon}_j$ is a vector of 10 unique factor scores (residuals) for subject j . Moreover, $\boldsymbol{\tau}_g$ is a vector containing 10 intercepts, $\boldsymbol{\lambda}_g$ is a vector of 10 common factor loadings, and $\boldsymbol{\delta}_g$ is a vector of 10 residual factor loadings of group g . Following Barendse et al. (2010), differences in the common factor were simulated by drawing common factor scores from a standard normal distribution for the reference group $t^r \sim N(0, 1)$ and from a normal distribution with a lower mean for the focal group $t^f \sim N(-0.5, 1)$. Residual factor scores were drawn from a standard normal distribution $\boldsymbol{\varepsilon}_j \sim N(0, 1)$.

³In the case of a dummy-coded indicator, the mean is the proportion of the sample in Group 1. Mean-centering does not affect the variance, so a 1-unit increase in a mean-centered dummy code still represents a comparison of Group 1 to Group 0, just as the original dummy code does.

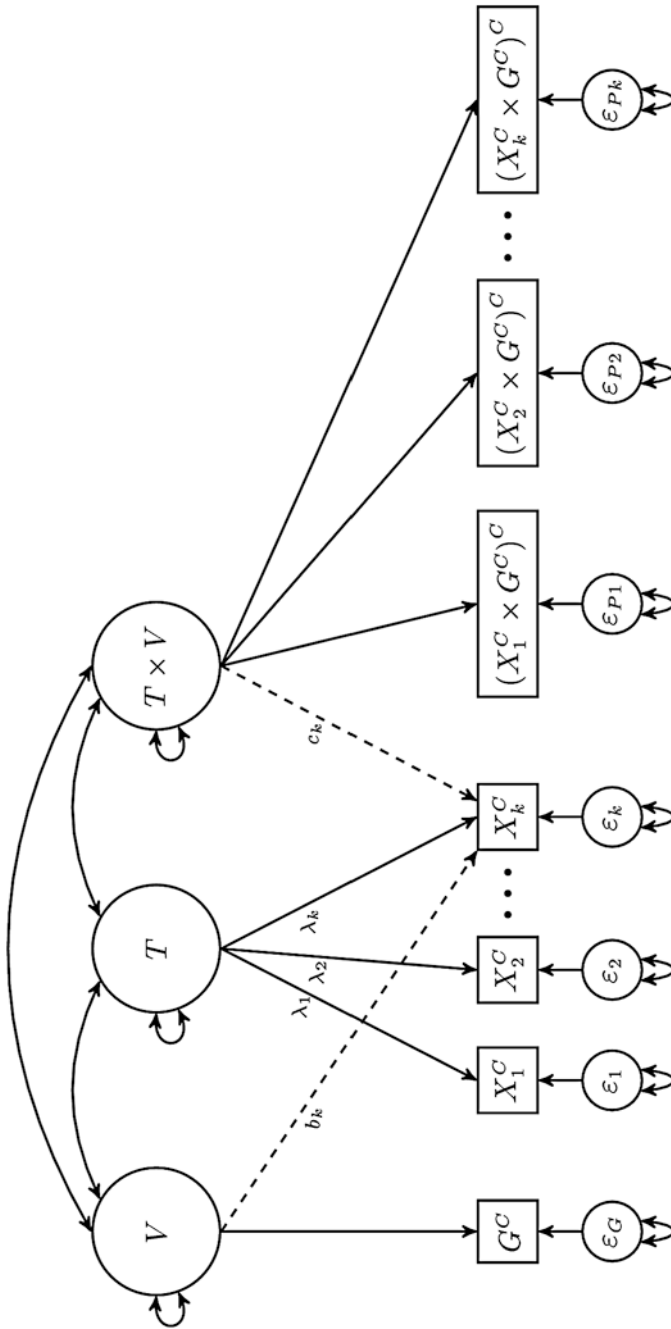


Fig. 2 An example of testing measurement bias using an RFA model with PI. Dashed arrows represent effects that may be estimated to test for uniform and nonuniform bias. The indicators X_1 and X_2 serve as anchor items

The same magnitude of uniform and nonuniform bias used by Barendse et al. (2010) was used. To introduce uniform bias, all intercepts τ were equal to 0, except for the intercept for the second and third item in the focal group, which were chosen equal to 0.5 (small uniform bias) and 0.8 (large uniform bias), respectively. Moreover, all common factor loadings were fixed at 0.8, except for the factor loadings of the fourth and fifth item in the focal group, which were chosen equal to 0.55 (small nonuniform bias) and 0.3 (large nonuniform bias), respectively. The residual factor loadings were set equal to the square root of $1 - \lambda_g^2$. Table 1 presents R syntax to generate this data set.

Table 1 R syntax for data generation for the illustrative example

```
## set seed
RNGkind("L'Ecuyer-CMRG")
.Random.seed <- as.integer(c(407, 1945764513, -1852313839, 178524778,
-983224279, -1572978333, -68534343))
## specify group size
Nn <- 100
## draw latent-trait values
thetal <- rnorm(Nn)
theta2 <- rnorm(Nn, -0.5, 1)
## draw scores on residual factor
residual <- matrix(NA, 2*Nn, 10)
for (j in 1:Nn) {
  for (i in 1:10) {
    residual[j, i] <- rnorm(1)
  }
}
## model parameters reference group
loading1 <- rep(0.8, 10)
delta1 <- sqrt(1 - loading1^2)
## model parameters focal group
tau2 <- c(0, -0.5, -0.8, 0, 0, 0, 0, 0, 0, 0)
loading2 <- c(0.8, 0.8, 0.8, 0.55, 0.3, 0.8, 0.8, 0.8, 0.8, 0.8)
delta2 <- sqrt(1 - loading2^2)
## simulate indicator scores reference group
x1 <- matrix(NA, Nn, 10)
for (j in 1:Nn) {
  for (i in 1:10) {
    x1[j,i] <- loading1[i] * thetal[j] + delta1[i] * residual[j, i]
  }
}
## simulate indicator scores focal group
x2 <- matrix(NA, Nn, 10)
for (j in 1:Nn) {
  for (i in 1:10) {
    x2[j,i] <- tau2[i] + loading2[i]*theta2[j] + delta2[i]*residual[j,i]
  }
}
## combine scores of both groups
dat <- as.data.frame(rbind(x1, x2))
dat$group <- rep(c(1, 2), each = Nn)
names(dat) <- paste0("x", 1:11)
```

3.2 Application

Table 2 shows R syntax for the application of PI in RFA models to detect measurement bias in the simulated data set. The RFA models with PI are fitted with the R package `lavaan` (version 0.5–23; Rosseel 2012). In our example, we apply the double-mean-centering strategy. First, the `indProd()` function in the `semTools` package (version 0.4–14; `semTools` Contributors 2016) with the argument `doubleMC = TRUE` is used to transform the data in order to be suitable for this strategy. This way, the indicators of the common factor T and violator V are mean-centered and indicators of the interaction factor $T \times V$ are built by multiplying the mean-centered indicator of V by each mean-centered indicator of T . The resulting product indicators are mean-centered again. After the data are prepared, one constrained model for each studied item must be specified. We use the ninth and tenth items, which are both bias-free, as anchor items, so they are not tested for measurement bias. Hence, the studied items are the first eight items, four of which are biased, which leads to eight constrained models in total. The unconstrained model is the same across items.

The first factor of the unconstrained model is the common factor T with 10 mean-centered observed variables X^C as indicators. The second factor is the violator V with a mean-centered single indicator G^C representing group membership. The residual variance of G^C is fixed at 0. The interaction factor $T \times V$ is the third factor

Table 2 R syntax for the application of PI in RFA in the illustrative example

```
## required package
library(semTools)
## prepare data
datDMC <- indProd(dat, 1:10, 11, match = FALSE, doubleMC = TRUE)
## additional parameters
paramc <- paste0("group + group.by.theta =~ x", 1:8)
## specify and fit unconstrained model
mod.un <- c('
  theta =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10
  group =~ 1*x11
  group.by.theta =~ x1.x11 + x2.x11 + x3.x11 + x4.x11 + x5.x11 +
                    x6.x11 + x7.x11 + x8.x11 + x9.x11 + x10.x11
  x11 ~~ 0*x11', paramc)
mod.un.fit <- cfa(mod.un, data = datDMC, estimator = "MLM")
## specify and fit constrained models
out <- matrix(NA, nrow = 8, ncol = 2,
              dimnames = list(paste0("x", 1:8), c("X2", "p")))
for (i in 1:length(paramc)) {
  mod.con <- mod.un[-(i+1)] # remove b and c for the i-th studied item
  mod.con.fit <- cfa(mod.con, data = datDMC, estimator = "MLM")
  outfit <- lavTestLRT(mod.con.fit, mod.un.fit,
                      method = "satorra.bentler.2001")
  out[i,1:2] <- c(outfit[2,5], outfit[2,7])
}
## print results
out
```

of the unconstrained model with double-mean-centered product indicators. For example, the first indicator of the interaction factor is obtained by mean-centering $G^C \times X_1^C$. For all factors in the unconstrained model, the factor loading λ of the first indicator is fixed at unity for identification. Covariances between all three factors are freely estimated. Finally, factor loadings of all items on V and $T \times V$ are added, except for the anchor items. The constrained models are built by removing factor loadings of the studied item on V and $T \times V$ from the unconstrained model. The estimator to be used for the unconstrained and constrained models is set to “MLM”, which involves maximum likelihood estimation with robust standard errors and a Satorra-Bentler scaled test statistic (Rosseel 2012).

To test each of the eight items for measurement bias, likelihood ratio test statistics are calculated using the `lavTestLRT()` function in the `lavaan` package (version 0.5–23; Rosseel, 2012). This involves comparing the fit of the unconstrained model with each constrained model. By setting the argument `method = “satorra.bentler.2001”`, a scaled $\Delta\chi^2$ test statistic with 2 *df* is computed as described by Satorra and Bentler (2001). An item is flagged as biased with respect to violator V when the $\Delta\chi^2$ statistic is significant using a criterion of $\alpha = 0.05$.

3.3 Results of Measurement Bias Detection

Table 3 presents the results of measurement bias detection using RFA with LMS and PI. When the PI method was applied, the $\Delta\chi^2$ statistics of three out of four truly biased items were significant. The item with small nonuniform bias, Item 4, was not flagged as biased, which is consistent with previous Monte Carlo studies showing that power to detect uniform bias is greater than to detect nonuniform bias (Barendse et al. 2010, 2012). Moreover, none of the $\Delta\chi^2$ statistics of the bias-free

Table 3 Results of testing measurement bias using RFA models with PI and LMS

Item	PI		LMS	
	$\chi^2_{df=2}$	<i>p</i>	$\chi^2_{df=2}$	<i>p</i>
1	0.425	0.809	0.674	0.714
2	19.396	0.000	17.696	0.000
3	38.755	0.000	28.000	0.000
4	5.217	0.074	6.283	0.043
5	10.105	0.006	10.656	0.005
6	0.145	0.930	0.201	0.904
7	0.948	0.622	0.772	0.680
8	0.246	0.884	0.196	0.907

Note Bold cells indicate significant measurement bias. Items 9 and 10 were used as anchor items, so they were not tested for measurement bias

items were significant. Thus, none of the items were incorrectly flagged as biased using PI. The LMS method obtained comparable results, but correctly flagged all truly biased items as biased with respect to violator V .

4 Discussion

In this chapter, we proposed the use of PI in RFA models as an alternative to LMS to test nonuniform measurement bias. The illustrative example showed that this method obtains results comparable to LMS. Because RFA with LMS can only be implemented in *Mplus* (Muthén and Muthén 2012), knowing that PI performs at least as well as LMS provides more researchers the opportunity to test for nonuniform bias using SEM software package. An additional advantage of PI is the availability of more traditional SEM fit indices to test for model fit that are not available when using LMS in *Mplus*, nor when using other available strategies for modeling interactions with latent variables (e.g., random effects models which treat item responses as cross-nested within items and subjects). However, several aspects of the use of PI in RFA models are yet unclear, for example, which items should serve as product indicators for the interaction factor (e.g., all items, only anchor items, or anchor items and studied items). In addition, RFA models assume strict invariance, that is, equal residual variances across groups. Future research could investigate how violations of strict invariance affect Type I error rates.

References

- Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267–281). Budapest, Hungary: Akademia Kiado. https://doi.org/10.1007/978-1-4612-1694-0_15.
- 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. <https://doi.org/10.1007/s10182-010-0126-1>.
- Barendse, M. T., Oort, F. J., Werner, C. S., Ligtvoet, R., & Schermelleh-Engel, K. (2012). Measurement bias detection through factor analysis. *Structural Equation Modeling*, *19*, 561–579. <https://doi.org/10.1080/10705511.2012.713261>.
- Henseler, J., & Chin, W. W. (2010). A comparison of approaches for the analysis of interaction effects between latent variables using partial least squares path modeling. *Structural Equation Modeling*, *17*, 82–109. <https://doi.org/10.1080/10705510903439003>.
- Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin*, *96*, 201–210. <https://doi.org/10.1037/0033-2909.96.1.201>.
- Lin, G.-C., Wen, Z., Marsh, H. W., & Lin, H.-S. (2010). Structural equation models of latent interactions: Clarification of orthogonalizing and double-mean-centering strategies. *Structural Equation Modeling*, *17*, 374–391. <https://doi.org/10.1080/10705511.2010.488999>.
- Little, T. D., Bovaird, J. A., & Widaman, K. F. (2006). On the merits of orthogonalizing powered and product terms: Implications for modeling interactions among latent variables. *Structural Equation Modeling*, *13*, 497–519. https://doi.org/10.1207/s15328007sem1304_1.

- Marsh, H. W., Wen, Z., & Hau, K.-T. (2004). Structural equation models of latent interactions: Evaluation of alternative estimation strategies and indicator construction. *Psychological Methods, 9*, 275–300. <https://doi.org/10.1037/1082-989X.9.3.275>.
- Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research, 13*, 127–143. [https://doi.org/10.1016/0883-0355\(89\)90002-5](https://doi.org/10.1016/0883-0355(89)90002-5).
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Oort, F. J. (1992). Using restricted factor analysis to detect item bias. *Methodika, 6*, 150–160. <https://doi.org/10.1007/s10182-010-0126-1>.
- Oort, F. J. (1998). Simulation study of item bias detection with restricted factor analysis. *Structural Equation Modeling, 5*, 107–124. <https://doi.org/10.1080/10705519809540095>.
- R Core Team. (2016). R: A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from the comprehensive R archive network (CRAN). <https://www.R-project.org/>.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>.
- Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*, 507–514. <https://doi.org/10.1007/BF02296192>.
- Schwartz, G. (1978). Estimating the dimensions of a model. *Annals of Statistics, 6*, 461–464. <https://doi.org/10.1214/aos/1176344136>.
- semTools Contributors. (2016). semTools: Useful tools for structural equation modeling [Compute software]. Retrieved from <https://CRAN.R-project.org/package=semTools>.
- Umbach, N., Naumann, K., Brandt, H., & Kelava, A. (2017). Fitting nonlinear structural equation models in R with package nlsem. *Journal of Statistical Software, 77*(7), 1–20. <https://doi.org/10.18637/jss.v077.i07>.
- 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. <https://doi.org/10.1177/109442810031002>.
- Woods, C. M., & Grimm, K. J. (2011). Testing for nonuniform differential item functioning with multiple indicator multiple cause models. *Applied Psychological Measurement, 35*, 339–361. <https://doi.org/10.1177/0146621611405984>.