To cite this chapter:

Cloos, L., & Flake, J. (2022). Partial measurement invariance. In B. B. Frey (Ed.), *The SAGE Encyclopedia of Research Design* (pp. 1170-1171). Sage. https://doi.org/10.4135/9781071812082

Sage Reference

The SAGE Encyclopedia of Research Design

For the most optimal reading experience we recommend using our website. A free-to-view version of this content is available by clicking on this link, which includes an easy-to-navigate-and-search-entry, and may also include videos, embedded datasets, downloadable datasets, interactive questions, audio content, and downloadable tables and resources.

Author: Leonie J. R. Cloos, Jessica K. Flake
Pub. Date: 2022
Product: Sage Reference
DOI: https://doi.org/10.4135/9781071812082
Keywords: measurement, parameters, confirmatory factor analysis
Disciplines: Research Methods & Evaluation (general), Research Methods for the Social Sciences, Research Design, Sociology
Access Date: March 26, 2024
Publishing Company: SAGE Publications, Inc.
City: Thousand Oaks,
Online ISBN: 9781071812082

© 2022 SAGE Publications, Inc. All Rights Reserved.

© 2022 by SAGE Publications, Inc.

Measurement invariance, also referred to as *measurement equivalence* or *lack of differential item functioning*, is a term used to describe a property of an instrument (e.g., test, survey, or questionnaire) that measures a latent construct with observed indicators. Full measurement invariance signifies that *all* of the parameters that determine the empirical relations between the indicators and the construct are equal for any given measurement condition (e.g., different time points or groups) and that the relations are not condition-specific. If full measurement invariance is achieved, then the differences in subjects' observed scores are due to differences in the latent construct rather than a condition-specific variable that systematically influences the responses. It is required whenever researchers use a measure's observed composite scores to compare conditions (e.g., pre- vs. posttest, men vs. women) or use scores to rank or compare subjects from different populations (e.g., for educational purposes). However, measurement parameters can vary across conditions. Partial measurement invariance (PMI) refers to a situation in which only some, but not all, of the parameters are equal across conditions of use of an instrument. Measurement invariance can be considered a matter of degree in that a subset of parameters can be invariant while others are not, and an instrument can be partially invariant.

A lack of measurement invariance is a concern in social science research because some statements on questionnaires (i.e., items) may have different meanings across time and groups, complicating the interpretation of responses to those statements. For example, the question "Who is the president of the United States?" on a general measure of intelligence will have a different level of difficulty and a different meaning to a U.S. sample versus that of any other country. If researchers use a latent factor model (e.g., structural equation modeling [SEM] or item response theory), they can relax the assumption of full invariance by determining which items are noninvariant and then allowing those items to vary across conditions. These partially invariant models enable researchers to account for measurement differences across conditions and allow them to estimate latent scores on the construct rather than analyzing observed scores, which do not account for the measurement differences. If researchers correctly identified all noninvariant items and at least two items are invariant, they may retain all items and use estimates from these models in subsequent analyses. The remainder of this entry presents methods to evaluate PMI and provides options of statistical analyses that may be carried out in case of PMI.

Evaluating PMI

The two methods of evaluating measurement invariance that have received the most application and attention so far are multiple group confirmatory factor analyses and item response theory. The term *PMI* is typically used with multiple group confirmatory factor analyses methods.

In factor models, the observed item response (xi) is modeled as a linear function of the latent factor (i.e., the construct) η.

© 2022 by SAGE Publications, Inc.

$xi = \tau i + \lambda i\eta + \varepsilon i$

The parameters in this equation are the intercepts τi, denoting the expected mean of item i, controlling for the latent factor η, the loadings λi, determining the relationship between the item and the latent factor, and the items' error term εi. The measurement parameters τi, λi, and εi are the focus of PMI investigations. In data in which subjects' observations differ on a certain condition, invariance of these parameters is investigated by fitting a series of hierarchical, nested models in which these parameters are constrained to be equal across conditions. A bottom-up approach is commonly adopted during which the fit of the models to the data is evaluated in the following steps.

First, the basic configuration of the items on the factors is tested for equality across conditions (configural invariance). If this model has an acceptable fit, then the nested model constraining the loadings to be equal is tested (metric invariance). Two subsequent models add restrictions on item intercepts (strong invariance) and error terms (weak invariance) across conditions to each previous model, respectively. The models are compared against each other using an X^2 difference test, where a significant p value indicates that a more restrictive level of invariance may not be tenable. Noninvariant parameter estimates will result in a poorer-fitting model when constrained to be equal, and a more restricted model will fail to demonstrate acceptable fit if there are noninvariant items. Researchers can then work to identify a PMI model and consider which items are noninvariant with respect to their loadings, intercepts, or error terms by examining the modification indices or residual covariance matrices of the more restricted model. To improve the model, they can release the restriction on the parameter with the highest modification index or residual estimate. Often the source of misfit concerns several items. Statistical software can simultaneously estimate a partially constrained model for each item, freeing the restriction on the parameter of the level where invariance was rejected (i.e., loading, intercept, or error term), and compare the partially invariant model of each item to the fully restricted model using an X^2 - difference test. A significant p value indicates that releasing the restriction on the item results in a significant difference in the model. This test statistic is corrected for multiple testing in which the number of tests equals the number of items. If several comparisons are significant, the model that yields the greatest improvement of the chi-square test statistic is selected. In a subsequent step, the model with this noninvariant item is again compared to several partially constrained models in which the remaining items are freed, respectively. This procedure is repeated until there are no more significant differences between models.

Determining the Anchor Items

Since PMI allows for parameters to vary across conditions, the multigroup model requires more parameters than a regular factor model. This requires additional steps for model identification, beyond what is typically

© 2022 by SAGE Publications, Inc.

applied in CFA. To test for PMI, researchers need to constrain a set of parameters across conditions, such that at least one item is invariant. This so-called anchor item remains restricted across the measurement invariance models that are tested. The estimation of other parameters is dependent on the anchor item, so it is necessary to carefully select an item that is noninvariant. This can be done with different statistical procedures such as a factor ratio test and the stepwise partitioning procedure. This causes a need for multiple testing, so it is recommended to evaluate anchor items based on theoretical considerations as well.

How to Respond to PMI

Once the noninvariant items are identified, the researcher can determine the final PMI model that applies across conditions. For example, if full intercept invariance was rejected and the source of noninvariance was identified, a PMI model would specify equal factor configuration and loadings across conditions but allow the intercepts of the items that are noninvariant to vary across conditions, as well as all items' error terms. The next section describes the consequences of PMI on observed scores and how to proceed with observed and latent models to investigate substantive research questions.

Observed Scores

When there is PMI, observed scores can be biased. The amount and direction of the bias depend on the pattern and magnitude of the noninvariance. Though this is an active area of methodological research, findings suggest that partially invariant loadings will bias the observed factor variances and partially invariant intercepts will bias the observed factor means. When the residual terms are noninvariant, observed scores will also bias reliability estimates. If researchers are interested in comparing variances or investigating the relationship of the scores to other constructs (e.g., estimate correlation and regression coefficients), they should ensure full loading invariance. If one wants to compare observed score means or determine prevalence rates between conditions, full invariance of both loadings and intercepts has to be established. When these requirements are not fulfilled, researchers are advised to proceed with statistical analyses in a latent factor model framework, which can be done using SEM.

SEMs and Factor Scores

SEMs can be used to obtain unbiased estimates of the group means and variances when observed scores

© 2022 by SAGE Publications, Inc.

are biased due to PMI. How to use these models for subsequent analyses depends on the research questions and the statistical analyses researchers are interested in. If loadings are noninvariant, fitting partially invariant SEMs, where noninvariant items are freely estimated across groups, are recommended to get unbiased estimates of factor variances and covariances. To determine the relationship to other constructs, the model can be extended by adding paths to exogenous variables to estimate unbiased regression coefficients. In the case of intercept differences, groups should be compared on their latent means in a partially invariant multigroup factor model that allows for different intercepts across groups. In PMI models, the latent factor is typically scaled by choosing one condition as the reference group, and fixing the mean of this group to 0 and the variances to 1. The mean of the other group is an unbiased estimate of difference between conditions rather than measurement differences. This approach is a *simultaneous* one where the parameters are directly interpreted from the model. Another option is the *multistage* approach where one first obtains factor scores from the PMI model and uses them for downstream statistical analyses such as *t*-tests, analyses of variance, and regression models. Those scores are corrected for measurement error as well as measurement differences across conditions and can be interpreted as *z* scores if the latent factor has been scaled by setting the mean to 0 and variance to 1 in the reference group. Comparing the results from analyses conducted with factor scores to those obtained with the original observed scores gives an estimate of how sensitive the results are to measurement noninvariance.

See also Confirmatory Factor Analysis; Construct Validity; Differential Item Functioning; Item Response Theory; Measurement Invariance

Leonie J. R. Cloos Jessica K. Flake

- measurement
- parameters
- confirmatory factor analysis

Byrne, B. M., & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology*, 34(2), 155–175. .

Johnson, E. C., Meade, A. W., & DuVernet, A. M. (2009). The role of referent indicators in tests of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(4), 642–657. .

© 2022 by SAGE Publications, Inc.

Millsap, R. E., & Kwok, O. M. (2004). Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological Methods*, 9(1), 93–115. .

Rensvold, R. B., & Cheung, G. W. (1998). Testing measurement models for factorial invariance: A systematic approach. *Educational and Psychological Measurement*, 58(6), 1017–1034. .

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, 18(4), 210–222. .

Steinmetz, H. (2013). Analyzing observed composite differences across groups: Is partial measurement invariance enough? *Methodology*, 9(1), 1–12. .

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature. *Organizational Research Methods*, 3(1), 4–70. .

https://doi.org/10.4135/9781071812082