Studia Psychologica
UKSW
15(2) 2015 s. 83–99

JAN CIECIUCH[1]
University Research Priority Program „Social Networks”,
University of Zürich, Switzerland
Cardinal Stefan Wyszyński University in Warsaw, Poland

ELDAD DAVIDOV[2]
Institute of Sociology, University of Zürich, Switzerland

ESTABLISHING MEASUREMENT INVARIANCE ACROSS ONLINE AND OFFLINE SAMPLES. A TUTORIAL WITH THE SOFTWARE PACKAGES AMOS AND MPLUS

Abstract

Establishing measurement invariance for indicators measured using online and offline modes of data collection is a precondition for the comparison of such data. Furthermore, it may allow accumulating knowledge using data from different sources or pooling data collected using different methods. The main goal of the current paper is to present a tutorial outlining the procedure of testing for measurement invariance using the Amos and Mplus software packages. We focus on the following steps for performing a measurement invariance test: 1) model specification, 2) model identification, 3) model estimation and evaluation, and 4) model modification.

Keywords: measurement invariance; Amos; Mplus

TESTOWANIE RÓWNOWAŻNOŚCI POMIARU KWESTIONARIUSZY W WERSJACH ONLINE I OFFLINE: ZASTOSOWANIE PAKIETÓW OPROGRAMOWANIA AMOS I MPLUS

Streszczenie

Równoważność pomiaru konstruktów mierzonych online i offline jest warunkiem porównywalności zebranych danych. Umożliwia akumulację wiedzy uzyskanej na podstawie badań przeprowadzonych tymi dwiema różnymi metodami oraz łączenie danych zebranych tymi sposobami do dalszych analiz. Głównym celem artykułu jest prezentacja

[1] Correspondence address: jancieciuch@gmail.com
[2] Correspondence address: davidov@soziologie.uzh.ch

procedury testowania równoważności pomiaru w programach Amos i Mplus. Procedurę prezentujemy w kolejnych krokach testu równoważności: 1) specyfikacja modelu, 2) identyfikacja modelu, 3) estymacja i ocena modelu i 4) modyfikacja modelu.

Słowa kluczowe: równoważność pomiaru, Amos, Mplus

When is measurement invariance needed?

Similar data or measurements are often gathered under different conditions, among different groups, or using different modes of data collection such as online and offline. Researchers may be interested in pooling or comparing such data even though they may have been collected using different methods. For instance, if data may be cheaply produced using online methods of data collection, researchers would like to know if the measurement quality of such data is comparable to that of data collected using other modes of data collection. Scholars may also be interested in knowing if such data are compatible to be pooled together with similar data collected using offline modes of data collection. Thus, in these situations it is necessary to guarantee that the data are comparable and that pooling the data does not mix up constructs which are differently understood by respondents, that respondents do not behave differently while responding to the different questionnaires, or that constructs do not possess different measurement properties. Testing for measurement invariance in these situations allows researchers to establish whether data are comparable rather than to simply assume this.

What is measurement invariance?

Measurement invariance is a property of an instrument (usually a questionnaire) intended to measure a given psychological construct. Measurement invariance affirms that a questionnaire does indeed measure the same construct in the same way across various modes of data collection, but also across different groups, at various time points or under different conditions (Chen, 2008; Marsh et al., 2010; Meredith, 1993; Millsap, 2011; Steenkamp & Baumgartner, 1998; Van de Vijver & Poortinga, 1997; Vandenberg, 2002; Vandenberg & Lance, 2000).

One can differentiate between several levels of measurement invariance. Each level is defined by the parameters constrained to be equal across samples. The first and lowest level of measurement invariance is called configural invariance (Horn & McArdle, 1992; Meredith, 1993; Vandenberg & Lance, 2000). Configural invariance requires that each construct is measured by the same items. This level of invariance does not guarantee that the measurement properties are the same and, therefore, higher levels of invariance are necessary before meaningful comparisons can be made.

The second level is called metric invariance (Horn & McArdle, 1992; Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000). Metric invariance is tested by constraining the factor loadings between the observed items and the latent variable to be equal across the compared groups (Vandenberg & Lance, 2000). If metric invariance is established, one may assume that people in the different samples interpret the items in the same way, although it is still uncertain if the construct is measured in the same way. If metric invariance is established, covariances or unstandardized regression coefficients may be meaningfully compared across samples.

A third and higher level of measurement invariance is called scalar invariance (Vandenberg & Lance, 2000). Scalar invariance is tested by constraining not only the factor loadings but also the indicator intercepts to be equal across groups (Vandenberg & Lance, 2000). If scalar invariance is established, one may assume that respondents use the scale in the same way in each group; thus, it implies that the same construct (metric invariance) is measured in the same way (scalar invariance). If scalar invariance is established, one may compare also latent or observed means across samples, and pooling data from the different samples may be conducted more confidently.

Partial invariance is supported when the parameters of at least two indicators per construct (i.e., loadings for partial metric invariance and loadings plus intercepts for partial scalar invariance) are equal across groups. According to Byrne, Shavelson, and Muthén (1989) and Steenkamp and Baumgartner (1998), partial invariance is sufficient for meaningful cross-group comparisons.

Testing for measurement invariance using the Amos and Mplus software packages

There are several procedures for conducting tests of measurement invariance and many software packages which can do it (for a review see Davidov, Schmidt, & Billiet, 2011; Millsap, 2011). The most widely used method is multigroup confirmatory factor analysis (MGCFA; Bollen, 1989; Jöreskog, 1971). This method involves setting cross-group constraints on parameters and comparing more restricted models with less restricted models (Byrne et al., 1989; Meredith, 1993; Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000). Two popular structural equation modeling (SEM) software packages which are frequently used to test for measurement invariance are Amos (Arbuckle, 2012) and Mplus (Muthén & Muthén, 2012).

In both Amos and Mplus it is possible to write syntax and to draw the model using a graphical input. Drawing models in Mplus is a relatively new feature that was introduced in its 7th version (Muthén & Muthén, 2012), whereas this has been the main feature of Amos since its inception. However, Mplus includes many advanced features relevant for measurement invariance testing (e.g., handling of categorical data or new optimization procedures) which are missing in Amos. Below we demonstrate how such an analysis is run in both of these software packages using the path diagram in Amos and the syntax in Mplus. For further reading we refer to the manuals of both software packages.

Empirical testing of measurement invariance across various samples may be conducted using a structural equation modeling (SEM) approach in four steps of analysis: 1) model specification, 2) model identification, 3) model estimation and evaluation, and if necessary 4) model modification and looking for partial measurement invariance. These steps will be described below, and guidance on performing these steps in Amos 21 and Mplus 7.1 will be provided (see, e.g., also Byrne, 2004).

Model specification

Specification of the confirmatory factor analysis (CFA) model implies determining which items measure which constructs and how the constructs relate to each other. It is recommended that the specification of CFA in the multigroup analysis is preceded by a CFA in each group separately (Byrne, 2010).

Model specification in Amos

The specification in Amos requires drawing a path diagram. In the path diagram rectangles represent observed variables (items), while the ellipses represent latent variables or measurement errors (which like the latent variables are also unobserved). A model for three latent variables each loading on three items (with nine items in total) is presented in Figure 1.

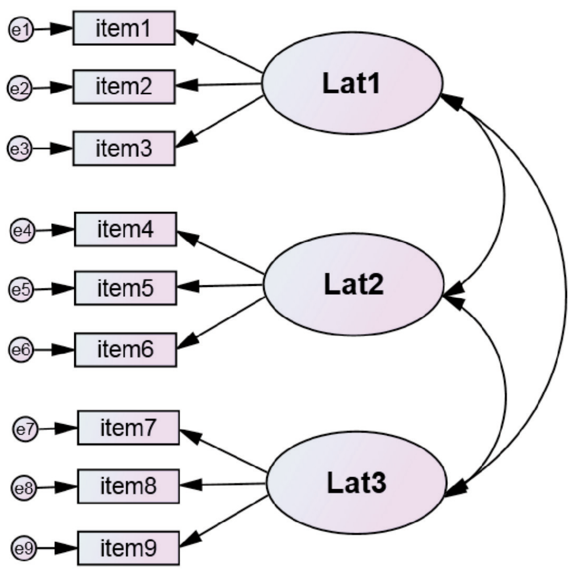

Figure 1. Specification of a CFA model in Amos.

The model consists of three latent variables: Lat1 (loading on item1, item2, and item3), Lat2 (loading on item4, item5, and item6) and Lat3 (loading on item7, item8, and item9). Nine measurement errors are represented by ellipses which load on their respective items. For example, e1 is the measurement error of item1. Thus, the observed scores of item1 depend on two unobserved components: the latent variable of interest (Lat1) and the measurement error (e1).

The specified model must be the same for the two online and offline samples. Therefore, in the box *Groups,* the online group and offline group should be introduced and the data should be imported for each group, respectively (using the menu option *File* → *Data files*), as illustrated in Figure 2.

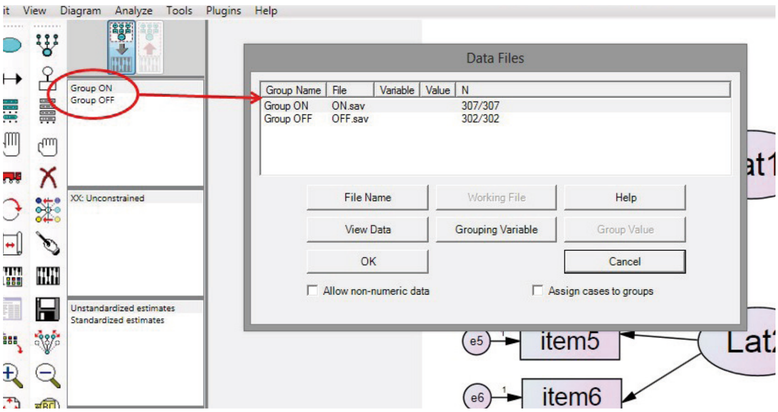

Figure 2. Importing data for groups in Amos.

After defining the groups, one may use the menu to conduct the invariance test (*Analyze* → *Multi-Group Analysis*), as presented in Figure 3.

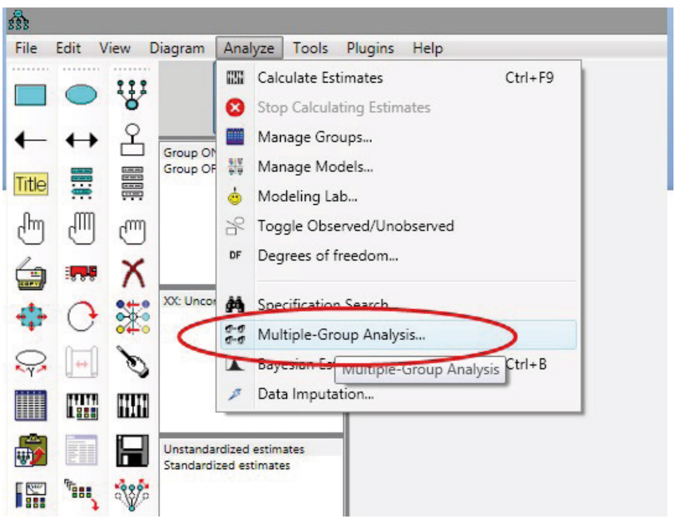

Figure 3. Multigroup analysis in Amos.

This option produces several models with increasingly strict equality constraints across groups. In the configural model there are no constraints. This model is labeled by Amos *unconstrained*. In the metric invariance model, the software constrains all factor loadings to be equal across the groups. This model is labeled by Amos *measurement weights*. In the scalar invariance model the software additionally constrains the intercepts to be equal across groups. This third model is labeled by the program *measurement intercepts*. Each set of constraints refers to a separate model. Model fit coefficients for these models are listed in the output. The program produces a few additional models with additional equality constraints on the measurement errors or on structural parameters such as covariances or latent means, which we do not discuss here.

Model specification in Mplus

Table 1 presents the syntax in Mplus of the model presented in Figure 1.

Table 1

Mplus Syntax and Explanations of Commands Used in Testing for Measurement Invariance

Syntax	Explanations
data: file is aaaa.dat;	Indicates the data source (the name of the file is aaaa) with a .dat format.
VARIABLE: Names are mode item1 item2 item3 item4 item5 item6 item7 item8 item9;	Indicates the grouping variable (in this case mode of data collection, but it could also be country or a cultural group) and the items in the data file.
grouping is mode (1 = off, 2 = on);	The variable mode contains two values, each indicating a different mode of data collection: the offline group (indicated by the value of 1) and the online group (indicated by the value of 2). This statement indicates that Mplus performs a group comparison across these two groups of the model described in the Model command below.
missing = all (999);	The value 999 in all variables indicates missing data.
ANALYSIS: model = configural metric scalar	This is a new convenience feature included in the 7.1 version of Mplus. It tests three models: configural, metric, and scalar invariance across the groups (online and offline) while adding the appropriate constraints of parameters for each model.

Syntax	Explanations
MODEL: LAT1 by item1 item2 item3; LAT2 by item4 item5 item6; LAT3 by item7 item8 item9;	Specification of the latent variables loading on their respective items. For example, latent variable LAT1 loads on three items: item1, item2, and item3.
OUTPUT: stand; modindices;	This command produces output about the estimated standardized parameters (in addition to the unstandardized ones) and the modification indices.

The model displayed in the *Model* command is tested for both groups. Mplus 7.1 and newer versions contain a convenience feature in the *Analysis* command as shown in the table. Three levels of measurement invariance are tested in separate models by adding the statement presented in the *Analysis* command in Table 1.

Model identification

Identification of the model implies adding constraints to the model to achieve a unique set of estimated parameters. Model parameters cannot be estimated if the model is not identified (Byrne, 2010). For example, the model presented in Figure 1 is not identified and cannot be estimated (for an extensive review about the topic of identification, see Bollen, 1989).

Several authors have proposed different ways to identify an SEM model (see, e.g., Little, Slegers, & Card, 2006). In the following illustration we use Little et al.'s (2006) second method that they described as the *marker-variable method*. This method chooses one indicator for each latent variable of interest, and constraints two parameters of the indicator in all the groups to be compared: the factor loading and the intercept. The factor loading of the marker (or reference) indicator is constrained to equal to 1, and the intercept of that indicator is constrained to equal to zero. The two constraints of the indicator parameters hold for all groups. The means of the latent variables are freely estimated in all groups.

Model identification in Amos

The researcher chooses one of the items for each latent variable to be the marker or reference indicator, and constrains its factor loading to 1 in all groups. It does not matter which indicator's factor loading is constrained to 1. Constraining the factor loadings to 1 is conducted by clicking the respective regression path, choosing the *object properties* window, introducing 1 as the regression weight, and marking in the box *for all groups*, as illustrated in Figure 4.

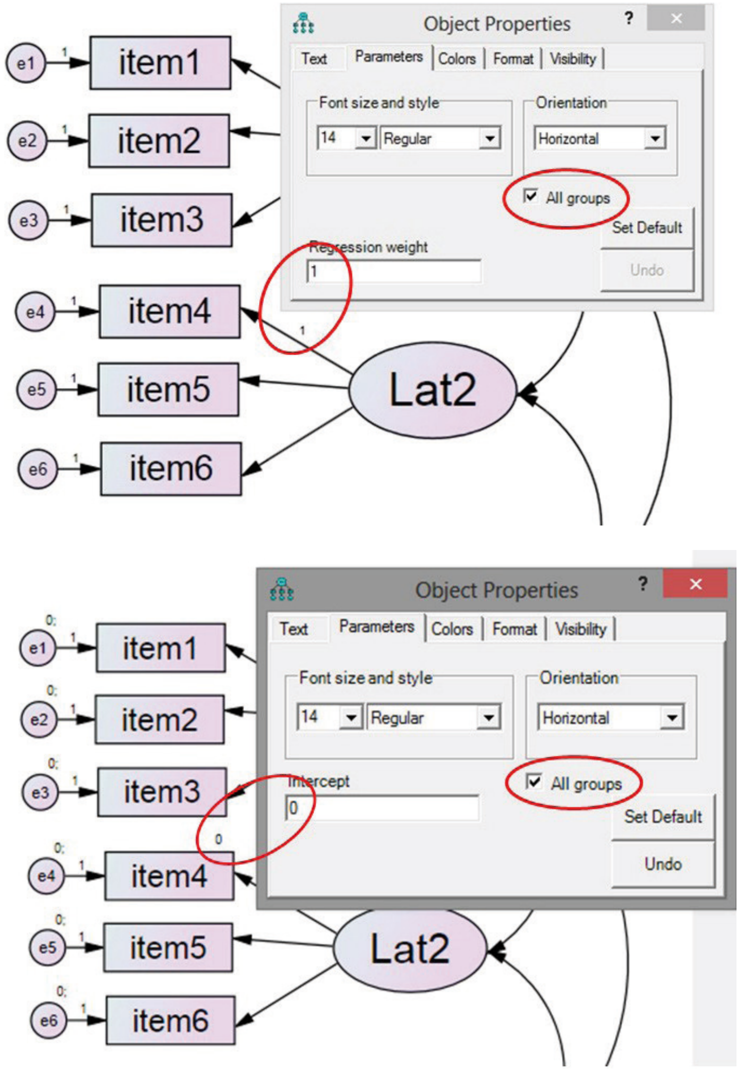

Figure 4. Model identification in Amos.

Before constraining the intercept of this item to zero, it is necessary to include mean and intercept parameters into the model because they are not estimated by default in Amos. This can be done by using the following path of commands in the menu: *view* → *analysis properties* → *estimation* → *estimate means and intercepts*, as presented in Figure 5.

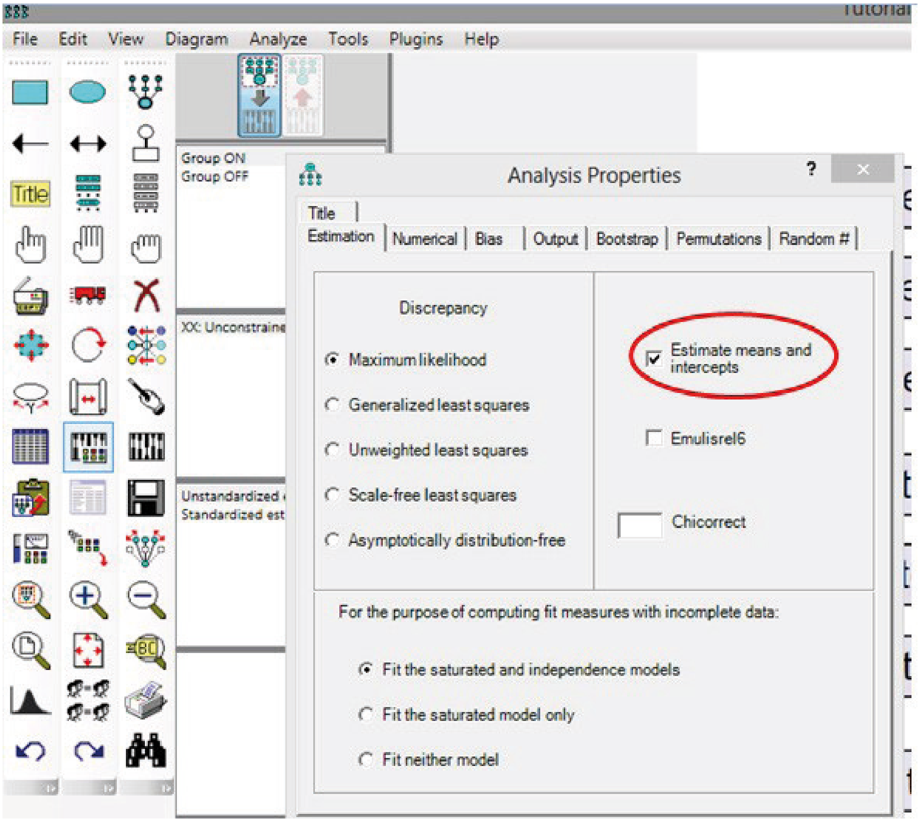

Figure 5. Estimating means and intercepts in Amos.

Constraining the intercept of the reference indicator to zero in all groups is conducted by clicking the respective indicator, choosing the *object properties* window, introducing the value '0' as the intercept parameter of that indicator, and marking in the box *for all groups* (see Figure 4), to guarantee that the constraint will apply for all the different samples (groups) in the analysis.

Model identification in Mplus

Mplus constrains by default the factor loading of the first indicator listed in the *Model* command for each latent variable to 1. Thus, the model described in the syntax presented in Table 1 for Mplus is identified. The means of the latent variables in one of groups are constrained to zero (Little, Slegers, & Card, 2006), and this group becomes the reference group for the mean comparison. Both the model evaluation and the decision regarding whether measurement invariance is established or not are based on the model fit comparison between the models.

Model estimation and evaluation

There are two main approaches to evaluate the quality of the model. The first approach relies on the global fit indices. The other approach, presented recently by Saris, Satorra, and van der Veld (2009), criticizes the use of global fit measures and focuses on testing local misspecifications. In this paper, we focus on the first approach, because it is the approach which is currently most often used by applied researchers.

The basic global fit measure is the chi-square (χ^2; see Jöreskog, 1969) which tests the hypothesis that the observed covariance matrix equals the hypothesized matrix. However, some problems with χ^2 have been recognized in the literature (e.g., Bentler & Bonett, 1980; Hu & Bentler, 1998; Hu, Bentler, & Kano, 1992; Kaplan, 1990). One of the problems of the χ^2 is that it is sensitive to sample size, and as a result it rejects good models with irrelevant or minor misspecifications in large samples (Bentler & Bonett, 1980). Instead, global fit indices and cutoff criteria were proposed in the literature, for example, by Hu and Bentler (1999) or by Marsh, Hau, and Wen (2004) for evaluating CFA models. Several other authors (Cheung & Rensvold, 2002; Chen 2007) proposed cutoff criteria for various fit measures to evaluate subsequent levels of measurement invariance in MGCFA. Table 2 presents a summary of global fit measures which are often used in the literature to evaluate measurement invariance and their recommended cutoff criteria.

A given level of measurement invariance is supported by the data when the changes of model fit indices are smaller than the values indicated in Table 2 on the last column when moving from a less restricted to a more restricted model. Thus, metric invariance is established when the change in model fit between the configural and the metric invariance models is smaller than the tolerable change indicated in Table 2. Scalar invariance is established when the change in model fit between the metric and the scalar invariance models is smaller than the tolerable change indicated in Table 2. Some scholars tolerate slightly larger changes to establish higher levels of measurement invariance (see, e.g., Byrne & Stewart, 2006).

Model modification and looking for partial measurement invariance

When the change of the fit indices is acceptable according to the cutoff criteria proposed by Cheung and Rensvold (2002) as well as by Chen (2007) which are indicated in Table 2, the researcher may conclude that measurement invariance is supported by the data. When this is the case, comparisons may be conducted across samples meaningfully, or samples may be pooled together. However, in reality, such a situation seldom occurs. Although the change of model fit indices usually exceeds the recommended cutoff criteria, one may still try to establish partial measurement invariance. Amos and Mplus provide, in the output, the

Table 2
Global Evaluation Criteria for Single CFA and Measurement Invariance in MGCFA

Coefficient	Definition	Cutoff criteria	
RMSEA Root mean square error of approximation	RMSEA reflects the degree to which a researcher's model reasonably fits the population covariance matrix and considers the degrees of freedom and sample size.	Single CFA (Hu & Bentler, 1999; Marsh, Hau, & Wen, 2004)	Measurement invariance in MGCFA, $N > 300$ (Chen, 2007)
		RMSEA < .008 – good fit RMSEA < .05 – very good fit	Δ RMSEA < .015
CFI Comparative fit index	CFI compares the fit of a researcher's model to a more restricted baseline model.	CFI > .90 – good fit CFI > .95 – very good fit	Δ CFI < .01
SRMR Standardized root mean square residual	SRMR compares the sample variances and covariances to the estimated ones.	SRMR < .08 – good fit SRMR < .05 – very good fit	Δ SRMR < .03 (when moving from configural to metric invariance model) Δ SRMR < .01 (when moving from metric to scalar invariance model)

Notes: Δ refers to the change in the fit measure when moving from a less restricted to a more restricted model. It indicates a tolerable change in the fit measure. If it is exceeded, it may imply violation of measurement invariance.

modification index (MI) and the expected parameter change (EPC) for misspecified parameters (Saris, Satorra, & Sörbom, 1987; Sörbom, 1989). The MI provides information on the minimal decrease in the χ^2 of a model when a given constraint is released. A decrease in χ^2 leads to an improvement of the model. The EPC provides a prediction of the minimal change of the given parameter when released (Saris et al., 1987). Thus, the EPC provides a direct estimate of the size of the parameter change in the modified model, whereas the MI provides a significance test for the estimated misspecification (Saris et al., 1987).

Researchers may look for the EPC of parameters which are constrained to be equal across groups that cause the largest misspecification and release them (Saris et al., 2009). This modification leads to a global model fit improvement. Researchers testing for (partial) metric measurement invariance may release the equality constraints of factor loadings which cause the largest misspecification in the model, whereas researchers testing for (partial) scalar invariance may release the cross-group equality constraints of intercepts which cause the largest misspecification in the model. If the parameters of two or more indicators of a given latent variable are still constrained to be equal across groups, and the change of global model fit indices is below the given cutoff criteria compared to a model with a lower level of invariance, one may conclude that partial invariance is supported by the data.

Testing for partial measurement invariance in Amos

Releasing certain constraints in Amos may be done in the *models* window. After opening a given model (*measurement weights* for the metric invariance model or *measurement intercepts* for the scalar invariance model), one can see all the equality constraints listed for that model, as presented in Figure 6.

To release some of these constraints, one simply deletes them from the list. After deleting the constraints, one should rerun the modified model and inspect the global fit indices. If the global fit indices of the modified model fulfill the cutoff criteria, the researcher may conclude that partial invariance is given if the parameters of at least two items per latent variable are equal across groups. Otherwise, one should repeat the procedure until the fit is satisfactory.

Testing for partial measurement invariance in Mplus

The model specified in Table 1 tests for scalar invariance after omitting the *Analysis* command. When the *Analysis* command is omitted, the default of Mplus sets all factor loadings and intercepts to be equal. In this model, factor loadings and intercepts are constrained to be equal across the online and offline samples. Releasing the equality constraints of single parameters (factor loadings or intercepts) in some of the groups is possible in the *Model* command. An additional statement introduced in the model for one of the groups will overwrite the default cross-groups equality constraints as described in Table 3 below. The statement indicates the name of the group whose equality constraints are released and lists the parameters that are released. Table 3 presents an example of how to specify partial metric and partial scalar invariance across the online and the offline samples, assuming that the loading of item5 and the intercept of item8 are noninvariant. All the other commands listed in Table 1 remain unchanged.

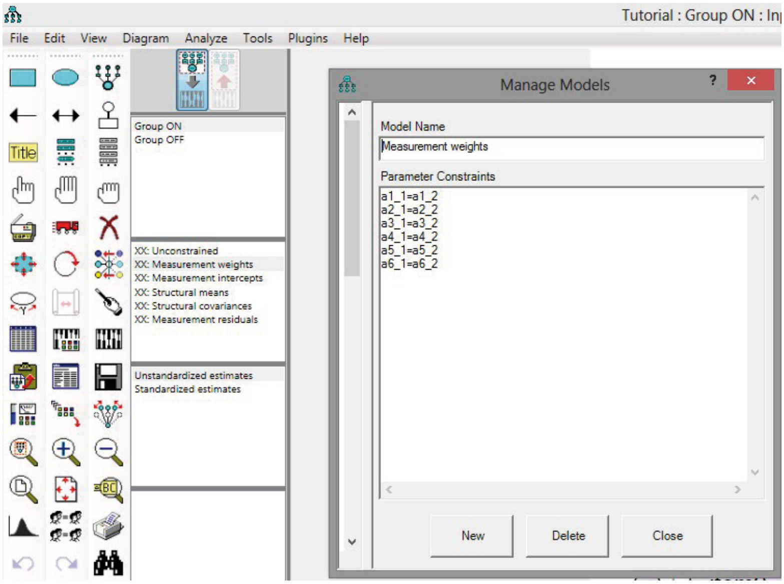

Figure 6. Equality constraints of parameters in Amos.

Table 3
Part of an Mplus Syntax Which Specifies Partial Metric and Partial Scalar Measurement Invariance (the Model *Command)*

Syntax	Explanations
MODEL: LAT1 by item1 item2 item3; LAT2 by item4 item5 item6; LAT3 by item7 item8 item9;	See explanations in Table 1 for this part of the syntax.
Model on LAT2 by item5; [item8];	The statement indicates that the factor loading of item5 on the latent variable LAT2 in the online group is released, that is, not constrained to be equal to the loading of this item in the offline group. It is thus freely estimated. An item name in brackets refers to a release of the equality constraint on the intercept of this item. The statement indicates that the intercept of item8 in the online group is not constrained anymore to be equal to the intercept of this item in the offline group and is thus freely estimated.

When more than two groups are compared, this part of the syntax should be extended and the parameters in the additional groups should be released as well.

What to report?

When reporting the results of a measurement invariance test a conclusive full account of the output would include a summary of the model specification, the factor loadings, and other parameter estimates for each group separately. Moreover, for the final models, a description of modifications made to the models should be provided along with the values of the global fit measures. However, due to restricted space or to word limits in many of the journals, it is often not possible to provide such an extensive report. Instead, much of this output may be included in an Appendix or on a website, or be provided by the author(s) upon request. If space allows, it would be very useful to include the following information in the body of the text at the very least:

1) Descriptive information on the items such as information about the scales used, and the distributions and the correlations between the variables.
2) Model specification and global fit measures of the single groups as well as the global fit for the different levels of measurement invariance. Although RMSEA, CFI, and SRMR are commonly used to discern between well-fitting and badly fitting models, it would be useful to complement this information by providing also the χ^2 values and the number of degrees of freedom for each model.
3) The estimator used. Maximum likelihood is the default estimator in Amos and Mplus, but it may be replaced by other more advanced procedures (such as robust weighted least squares, RWLS, to deal with categorical data; see Flora & Curran, 2004).
4) Information concerning all model modifications and global model fit of the accepted models.

Conclusions

Testing for measurement invariance is of great importance while conducting any kind of research involving more than one group or when combining data gathered using various modes of data collection. Testing for measurement invariance may be conducted to estimate whether data collected using online and offline modes of data collection are comparable, but is not restricted to this goal. The procedure described above can be extended and used for any study involving any group comparison, such as gender groups, age groups, language groups, or cultural groups. The logic and procedure of how to test for measurement invariance remains essentially the same in all cases and guarantees that conclusions from studies involving and combining various samples are not erroneous and that results are not biased.

Acknowledgment

The work of both authors on this paper was supported by the Scientific Exchange Program – SCIEX (Switzerland) and the University Research Priority Program (URPP) Social Network, University of Zurich. The work of the first author was partially supported by Grants DEC-2011/01/D/HS6/04077 from the Polish National Science Centre. The second author would like to thank GESIS Cologne, the EUROLAB, for their hospitality during work on this paper.

Both authors would like to thank Lisa Trierweiler for the English proof of the manuscript.

References

Arbuckle, J. L. (2012). *Amos 21 user's guide*. Chicago, IL: IBM SPSS.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Bentler, P. M. & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance-structures. *Psychological Bulletin*, *88*(3), 588–606. doi: 10.1037//0033-2909.88.3.588

Byrne, B. M. (2004). Testing for multigroup invariance using AMOS Graphics: A road less traveled. *Structural Equation Modeling*, *11*(2), 272–300. doi: 10.1207/s15328007sem1102_8

Byrne, B. M. (2010). *Structural equation modeling with Amos. Basic concepts, applications, and programming.* New York, London: Routledge Taylor & Francis Group.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures – the issue of partial measurement invariance. *Psychological Bulletin*, *105*(3), 456–466. doi: 10.1037/00332909.105.3.456

Byrne, B. M. & Stewart, S. M. (2006). The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process. *Structural Equation Modeling*, *13*(2), 287–321. doi: 10.1207/s15328007sem1302_7

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464–504. doi: 10.1177/0734282911406661

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, *95*(5), 1005–1018. doi: 10.1037/a0013193

Cheung, G. W. & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255. doi: 10.1207/S15328007SEM0902_5

Davidov, E., Schmidt, P., & Billiet, J. (2011). *Cross-cultural analysis: Methods and applications*. New York: Routledge.

Horn, J. L. & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, *18*(3–4), 117–144. doi: 10.1080/03610739208253916

Flora, D. B. & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*(4), 466–491. doi: 10.1037/1082-989X.9.4.466

Hu, L. T. & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424–453. doi: 10.1037/1082-989x.3.4.424

Hu, L. T. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *66*(1), 1–55. doi: 10.1080/10705519909540118

Hu, L. T., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure-analysis be trusted. *Psychological Bulletin*, *112*(2), 351–362. doi: 10.1037/0033-2909.112.2.351

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*(2), 183–202. doi: 10.1007/bf02289343

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36*(4), 409–426. doi: 10.1007/bf02291366

Kaplan, D. (1990). Evaluating and modifying covariance structure models – A review and recommendation. *Multivariate Behavioral Research*, *25*(2), 137–155. doi: 10.1207/s15327906mbr2502_1

Little, T. D., Slegers, D. W., & Card, N. A. (2006). A non-arbitrary method of identifying and scaling latent variable in SEM and MACS models. *Structural Equation Modeling*, *13*(1), 59–72. doi: 10.1207/s15328007sem1301_3

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cut-off values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, *11*(3), 320–341. doi: 10.1207/s15328007sem1103_2

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment*, *22*(3), 471-491. doi: 10.1037/a0019227

Meredith, W. (1993). Measurement invariance, factor-analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543. doi: 10.1007/bf02294825

Millsap, R. E. (2011). *Statistical approaches to measurement invariance.* New York, London: Routledge Taylor & Francis Group.

Muthén, L. K. & Muthén, B. O. (2012). *Mplus user's guide. Seventh edition*. Los Angeles, CA: Muthén & Muthén.

Saris, W. E., Satorra, A., & Sörbom, D. (1987). The detection and correction of specification errors in structural equation models. *Sociological Methodology*, *17*, 105–129.

Saris, W. E., Satorra, A., & van der Veld, W. M. (2009). Testing structural equation models or detection of misspecifications? *Structural Equation Modeling*, *16*(4), 561–582. doi: 10.1080/10705510903203433

Sörbom, D. (1989). Model modification. *Psychometrika*, *54*(3), 371–384. doi: 10.1007/bf02294623

Steenkamp, J.-B. E. M. & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*(1), 78–90. doi: 10.1086/209528

Van de Vijver, F. J. R. & Poortinga, Y. H. (1997). Towards an integrated analysis of bias in cross-cultural assessment. *European Journal of Psychological Assessment*, *13*(1), 29–37. doi: 10.1027/1015-5759.13.1.29

Vandenberg, R. J. (2002). Toward a further understanding of and improvement in measurement invariance methods and procedures. *Organizational Research Methods*, *5*(2), 139–158. doi: 10.1177/1094428102005002001

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*(1), 4–70. doi: 10.1177/109442810031002