The Impact of Controlling for Extreme Responding on Measurement Equivalence in Cross-Cultural Research

M. Morren*

J.P.T.M. Gelissen*

J.K. Vermunt*

* Department of Methodology and Statistics

Tilburg University

P.O. Box 90153

5000 LE Tilburg,

The Netherlands

Keywords:

measurement equivalence, extreme response style, latent class factor models, cross-cultural research

Word counts:

Manuscript = 5423; Abstract = 150; Notes = 446; Appendix: 559.

This research was supported by a grant from the Netherlands Organisation for Scientific Research, MaGW/NWO project number (no. 400-05-164).

Abstract

Prior research has shown that extreme response style can seriously bias responses to survey questions and that this response style may differ across culturally diverse groups. Consequently, cross-cultural differences in extreme responding may yield incomparable responses when not controlled for. To examine how extreme responding affects the cross-cultural comparability of survey responses, we propose and apply a multiple-group latent class approach where groups are compared on basis of the factor loadings, intercepts and factor means in a Latent Class Factor Model. In this approach a latent factor measuring the response style is explicitly included as an explanation for group differences found in the data. Findings from two empirical applications that examine the cross-cultural comparability of measurements show that group differences in responding import inequivalence in measurements among groups. Controlling for the response style yields more equivalent measurements. This finding emphasizes the importance of correcting for response style in cross-cultural research.

The Impact of Controlling for Extreme Responding on Measurement Equivalence in Cross-Cultural Research

Cross-cultural comparisons in which people from different nations or ethnic backgrounds are asked how they feel about social issues or how they behave constitute an important part of research in the social and behavioral sciences. More and more attention is being paid to the validity of such comparisons (Berry, Poortinga, Segall, & Dasen, 2002; Johnson, Kulesa, Cho, & Shavitt, 2005; Van de Vijver, 1998; Van de Vijver & Leung, 1997). In particular, this field of research questions whether it is possible to compare people with different cultural backgrounds on their attitudes and values. It is likely that people with a different frame of reference – rooted in their experiences, their social interactions, and the norms and values shared by their group – understand the topics raised in a survey differently (Triandis, 1990; Wallace & Wolf, 1998). Consequently, because describing and explaining differences in attitudes is the aim of most cross-cultural studies, one should empirically establish that respondents from different groups have the same topic in mind while answering a survey-item (Krosnick, 1999; Tourangeau, 2003). If this is not the case, comparing attitudes between groups is similar to comparing apples and oranges. The methodological literature refers to this situation as measurement inequivalence.

In this contribution, we argue that such a lack of measurement equivalence (ME)[1] can be related to group differences in response styles which cause respondents from

---

[1] In other traditions ME is referred to as measurement invariance (MI) (Cheung & Rensvold, 2000; Meredith, 1993; Millsap, 1995; Steenkamp & Baumgartner, 1998), or as Differential Item Functioning (DIF) (Sijtsma & Molenaar, 2002). Alternatively, Adcock and Collier (2001) address ME as the contextual specificity of measurement validity.

culturally diverse backgrounds to respond differently to the items than one would expect on basis of their attitudes (Hui & Triandis, 1989). We show that these group differences in responding lead to what appears to be measurement inequivalence and that correcting for the response style results in more equivalent measurements.

The investigation is narrowed down to extreme response style (ERS) because it has repeatedly been shown that this response style seriously distorts attitude measurement in social survey research (see for instance Chun, Campbell, & Yoo, 1974; De Jong, Steenkamp, Fox, & Baumgartner, 2008). The response pattern of an item that is affected by ERS shows a higher frequency of extreme responses – the endpoints of the item scale – than one would expect based on the respondent's attitude. This impedes a correct estimation of model parameters when modeling group differences in attitudes. Moreover, an extreme response pattern may represent a truly extreme attitude as well as ERS; that is, ERS may confound genuine and stylistic variance (Van de Vijver & Leung, 1997). Thus, ERS leads to biased attitude measurement if not controlled for. Additionally, research findings show that people with differing cultural backgrounds may be subject to extreme responding to a different degree (Bachman & O'Malley, 1984; Gibbons, Zellner, & Rudek, 1999; Hui & Triandis, 1989; Johnson, et al., 2005; Marin, Gamba, & Marin, 1992). In this paper we show how a difference in ERS between culturally diverse groups imports measurement inequivalence in the data and, if not controlled for, biases the attitude measurement.

To this end, we propose a latent variable model that simultaneously allows for examining measurement equivalence as well as the detection of and the correction for ERS. Importantly, this model enables us to assess the implications of the presence of ERS for measurement equivalence. We build on the contributions of Moors (2003,

2004) and apply logistic Latent Class Factor Analysis (LCFA) (Eid, Langeheine, & Diener, 2003; Heinen, 1996; Vermunt & Magidson, 2004) instead of linear Structural Equation Modeling (SEM) that is commonly used in multiple group analyses (Billiet & McClendon, 2000; Byrne & Stewart, 2006; Cheung & Rensvold, 2000). As we will show, SEM is an inappropriate method to deal with the non-monotone response pattern caused by ERS because of the assumption of linear relationships between latent and observed variables. In contrast, the less restrictive LCFA approach proposed here does not make such stringent assumptions, it allows for the detection of and the correction for ERS and measurement equivalence can be assessed.

In the remainder of this contribution, we illustrate how multiple-group analysis within the LCFA framework can be used to detect measurement inequivalence. We present a latent variable model that disentangles style and substance and we explain how this model can be adjusted to a LCFA model that can also detect and correct for ERS. We then show in an analysis of a generated data set in which we simultaneously detect measurement inequivalence and correct for ERS that specific forms of measurement inequivalence relate to the presence of extreme responding. Finally, we apply the multiple group LCFA approach to data obtained from four ethnic groups within the Netherlands using the Dutch survey The Social Position of Ethnic Minorities and Their Use of Services (SPVA)[2] and demonstrate the usefulness of the approach in an empirical application.

---

[2] In Dutch, the abbreviation SPVA stands for *Sociale Positie en Voorzieningengebruik van Allochtonen.* We thank Data Archiving and Networked Services (DANS) for providing the data files.

A Latent Class Factor Approach to Multiple-Group Analysis

Complex constructs such as people's attitudes cannot be observed directly. To obtain a valid and reliable measurement of such constructs, researchers usually ask respondents multiple questions which indicate several important aspects of the attitude (Bollen, 2002; Skondral & Rabe-Hesketh, 2004). These ideas about attitude measurement are applied by modeling the attitude as a latent – unobserved – variable (also called factor or trait) and the questions as observed variables (hereafter called items). Within this latent variable framework, an important goal of multiple-group analyses is to measure the extent to which the groups have different attitudes in terms of the group means of the latent variables. However, these group differences in latent means can only be compared validly and reliably when the same latent variable model can be applied within each group as well as across groups (Byrne & Watkins, 2003; Meade & Lautenschlager, 2004a, p. 60; Mullen, 1995; Van de Vijver & Leung, 1997).

In this paper, we investigate whether the items and the response scale of attitude measurements are used homogeneously – which indicates measurement equivalence – or rather heterogeneously – which indicates measurement inequivalence – by people who come from culturally diverse backgrounds but who actually have similar attitudes (Hui & Triandis, 1985; Van de Vijver & Leung, 1997). If measurement equivalence is absent, which is shown by the fact that the associations between the items and the attitudes differ across groups in strength and significance, then we hypothesize that this absence of measurement equivalence can be partly or even completely be explained by a confounding effect due to a group-specific presence of ERS. In Figure 1, we graphically illustrate various Latent Class Factor Models which allow the investigation of such issues in a multiple-group analysis on a pooled sample.

[Insert Figure 1 about here]

Here, Y_1 *to* Y_{10} represent the item responses which are directly related to the latent variables measuring the attitudes F_1 and F_2. Furthermore, the item responses may either be related directly, or indirectly, or in interaction with the effect of the latent variable measuring the attitudes to the observed group variable *G*. As we will explain below, which particular effects of the grouping variable G are included in the model depends on the type of measurement equivalence that the researcher seeks to investigate. The systematic variance among the item responses is captured by the factor loadings, i.e. the relations between the latent variables and the item responses; the random variation is represented by the error terms ε_j. In the models depicted in Figure 1, it is assumed that the five items in the first item subset do not relate directly to the second attitude which is modeled by fixing the item parameters β_{2j} to zero. In the same way, the parameters β_{1j} are fixed to zero for the five items in the second item subset.

An important advantage of the LCFA approach to multiple-group analysis in comparison to other well-known approaches that are based on the linear regression model is that the equivalence of item intercepts, factor loadings, factor means and (co)-variances – which is necessary for the evaluation of various forms of measurement equivalence – can be tested simultaneously without using restrictions. In particular, contrary to the linear regression model used in CFA analyses LCFA uses an ordinal logit regression model to measure the latent variables. The consequence of this difference in modeling is that whereas in the CFA approach the researcher needs to include certain restrictions in the model to fix the location and scale of the latent variables, this is unnecessary in the LCFA approach. The ordinal logit model for the latent variables in Figure 1 is described in Appendix A. In such multiple-group analyses, the group differences are introduced in the model by an explanatory covariate which measures group membership; in Figure 1 this is indicated by *G*. A direct effect of *G* on the latent variables denotes a group difference in the latent means and/or co-variances. These group differences in the attitudes can only be measured reliably and validly when the attitudes are measured equivalently across groups. Here, we are interested in two forms of measurement equivalence: scalar and metric equivalence[3].

Scalar equivalence is the most restrictive type of measurement equivalence and occurs when respondents from different backgrounds react similarly to the items given their attitudes. Establishing scalar equivalence is necessary to validly compare means of latent variables across groups. The situation of scalar equivalence is depicted in Figure 1a. Here, the group differences in the latent means are indicated by the dashed arrows between the observed variable *G* and the latent variables F_1 and F_2. The model in Figure 1a for the observed score of respondent *i* on item *j* is formally represented by:

$$E(Y_{ij} \mid F_{1i}, F_{2i}) = \beta_{0j} + \beta_{1j} F_{1i} + \beta_{2j} F_{2i} \qquad [1]$$

where the expected value of the response Y_{ij}, conditional on the attitudes F_{1i} and F_{2i}, depends on the item parameters β_{0j} representing the intercept, the parameters β_{1j} representing the influence of F_{1i}, and the parameters β_{2j} representing the influence of

---

[3] Prior to assessing scalar and metric equivalence, one also needs to establish configural equivalence, which holds that items measuring the attitudes exhibit the same configuration of loadings in all groups. Here we assume that configural equivalence has been established and the researcher now seeks to investigate more restrictive forms of equivalence of measurement instruments.

F_{2i}. The expected value of the errors ε_{ij} is zero because they are assumed to be unrelated and normally distributed.

A weaker form of measurement equivalence is metric equivalence, which is attained when the groups differ in their perception of the origin of the item scale but perceive the distances between the item categories and/or the order of the item categories similarly. Thus, metric equivalence is defined as groups having different item intercepts and error terms but equal factor loadings given their attitudes:

$$E(Y_{ij} \mid F_{1i}, F_{2i}, g) = \beta_{0jg} + \beta_{1j}F_{1i} + \beta_{2j}F_{2i} \qquad [2]$$

where the expectation of the response is conditional on the attitudes F_{1i} and F_{2i} and on group g to which individual i belongs. The subscript g of the parameter β_{0jg} denotes that the item intercept is group-specific; in other words, the intercepts are set free to vary across groups. In Figure 1b, the situation of metric equivalence is graphically represented for item 5 where a group-specific intercept is indicated by the dashed arrow representing a direct effect of the group variable G on Y_5.

Note that measurement equivalence is completely violated if the groups perceive the items completely different given their attitudes, resulting in group differences in the intercepts and the factor loadings:

$$E(Y_{ij} \mid F_{1i}, F_{2i,} g) = \beta_{0jg} + \beta_{1jg}F_{1i} + \beta_{2jg}F_{2i} \qquad [3]$$

where the subscript g of the parameters β_{1jg} and β_{2jg} denotes that the factor loadings are group-specific in addition to the intercepts. In Figure 1c, measurement inequivalence is represented for item 5 by the dashed arrows representing a direct effect of G on Y_5 β_{05g} and a group-specific factor loading β_{15g}.

In the LCFA approach to multiple-group analyses, we test for equivalence of certain parameters by constraining them to be equal across groups, yielding a more parsimonious model. If the groups respond equivalently then this more parsimonious model is preferred. However, if they respond inequivalently more complex models are required to avoid misspecification. By comparing model goodness-of-fit values, we assess which specific form of equivalence is attained. If fixing the group-specific parameters to equality does not deteriorate the model fit, the more parsimonious model is accepted and a particular type of measurement equivalence is attained. Specifically, the measurements are scalar equivalent if the fit of the model in equation [1] does not deteriorate compared to model described in [2]; the measurements are metric equivalent if the fit of the model in equation [2] does not deteriorate compared to the model in [3]. Note that – although in Figures 1b and 1c inequivalence is depicted for only one item – multiple or all items can of course be simultaneously inequivalent across groups.

Extreme Response Style

Apart from measurement inequivalence an additional problem surfaces in multiple-group analyses if the groups differ in their style of responding: the confounding of group differences in the attitudes and the response styles (Eid, et al., 2003; Poortinga & Van de Vijver, 1987; Van de Vijver & Leung, 1997). A straightforward manner to deal with this problem is to explicitly control for the response style by including one or more latent variables that accurately measure the response styles (Billiet & McClendon, 2000; Cheung & Rensvold, 2000; De Jong, et al., 2008). Figure 2 illustrates this approach which is a latent variable model that simultaneously detects and corrects for the response style.

[Insert Figure 2 about here]

In Figure 2, Y_1-Y_{10} indicate the item responses that relate to the latent variables representing the attitudes F_1 and F_2, and the extreme response style E. The response style and the attitude are disentangled by means of the model structure. Whereas the respondent's attitude only affects his or her answer to the items that reflect the same construct, the respondent's response style – by definition – affects the answers to all items regardless of their content (Hui & Triandis, 1989; Javeline, 1999; Johnson & Van de Vijver, 2003; Sudman, Bradburn, & Schwarz, 1996). The validity of the model is increased by including two weakly related attitudes: as ERS is unrelated to item content, it is expected that the response style is present across items treating diverse topics. Note that more substantive factors could be included to investigate whether the response style pertains to more items.

The model in Figure 2 was earlier applied within a Confirmatory Factor Analysis (CFA) framework to detect acquiescence (Billiet & McClendon, 2000); in this approach the latent variables as well as the observed variables are specified as continuous variables (Bollen, 1989; Joreskog, 1971). A consequence of the continuous specification in CFA is that the observed variables are required to relate linearly to the latent variables. However, in the case of ERS, the model in Figure 2 cannot be applied within the linear CFA framework because ERS violates the assumption of linearity (see below). In this contribution, we relax this assumption of linearity by using a Latent Class Factor Approach (LCFA) where the observed responses are specified as nominal variables and the latent constructs as ordinal variables[4]. By modeling each item category separately, assumptions concerning the

[4] The same model can be estimated with continuous latent variables without altering conclusions drawn in this paper. In that case, one should make additional restrictions to fix the location and scale of the latent variables. We chose for ordinal specification of the latent variables to facilitate the estimation procedure.

items as a whole are avoided and ERS influencing the item responses in a non-monotone manner can be detected by the model in Figure 2.

The non-monotonicity results from the particular response pattern that ERS causes among the item responses. The respondents subject to ERS are likely to select the extreme – positive and negative – categories more often than the other item categories, thereby leading to more observations in the extreme categories at both endpoints of the response scale (Moors, 2003). In contrast, the attitudes cause a monotone effect: the more positive the attitude of a respondent is, the more likely he or she is to select a positive answer and the more unlikely he or she is to select a negative answer. Therefore, an attitude induces a linear (and thus monotone) effect on the responses whereas ERS leads to a non-monotone relationship between the ERS factor and the responses. Figure 3 illustrates the non-monotone effect of ERS and the monotone effect of the attitude on the size of the category item parameters, which are in the case of the LCFA approach logit coefficients.

[Insert Figure 3 about here]

The x-axis in Figure 3 represent the item categories on the five-point response scale that runs from *totally agree* to *totally disagree* belonging to the item “In the Netherlands immigrants get many opportunities” which is part of the SPVA survey. Note that the same pattern appears for the other items. The y-axis describes the size of the parameters in the model that corrects for ERS illustrated in Figure 2. The dotted line shows the category-specific item parameters representing the influence of F_1, the crossed line illustrates the category-specific item parameters representing the influence of E.

Under the influence of the attitude, the size of the parameters increases along the item categories on the response scale. This illustrates the monotone manner in which the attitude relates to the observed item response. However, in the case of ERS, the size of the parameters decreases as well as increases along the response scale. This illustrates that ERS relates in a non-monotone manner to the item (see Figure 3). Because of this non-monotone pattern with respect to ERS, the item responses cannot be interpreted as responses of interval variables; that is, variables measured at an ordinal scale with equal distances between the item categories. Therefore, we model the item responses as nominal variables.

A nominal specification of the observed variables leads to a separate treatment of each response category: the response of individual *i* to item *j* is denoted by Y_{ij}, a response to a particular category by *c*, and the number of response categories by *C*. Note that for each attitude five items are included in the model, each having five categories and formulated as bipolar (the so-called Likert scales). The following multinomial logit model is used to model the relationship between the item responses and the attitudes:

$$P(Y_{ij} = c \mid F_{1i}, F_{2i}) = \frac{\exp(\beta_{0jc} + \beta_{1jc} F_{1i} + \beta_{2jc} F_{2i})}{\sum_{d=1}^{C} \exp(\beta_{0jd} + \beta_{1jd} F_{1i} + \beta_{2jd} F_{2i})} \qquad [4]$$

The probability of choosing category *c* of item *j* by individual *i*, conditional on F_{1i} and F_{2i}, is explained by the item parameters β_{0jc} representing the intercept and the parameters β_{1jc} and β_{2jc} representing the monotone relationship between the substantive F_{1i} and F_{2i} and the items. The error ε_{ij} is multinomially distributed.

As is typical of these multinomial logit models, each category *c* of item *j* has its own parameters, indicated by the index *jc* of the parameters β_{0jc}, β_{1jc}, β_{2jc} and β_{3jc} (Agresti, 2002). In the case of these category specific parameters, the identification of the category parameters can be accomplished by effect coding where the parameters are restricted to sum to zero across categories for each item. Another possibility would be dummy coding where the parameters are fixed to zero for one category.

To correct for the response style, we include a separate latent factor *E* measuring ERS in the model as is depicted in Figure 2. This leads to the following multinomial logit model:

$$P(Y_{ij} = c \mid F_{1i}, F_{2i}, E_i) = \frac{\exp(\beta_{0jc} + \beta_{1jc} F_{1i} + \beta_{2jc} F_{2i} + \beta_{3jc} E_i)}{\sum_{d=1}^{C} \exp(\beta_{0jd} + \beta_{1jd} F_{1i} + \beta_{2jd} F_{2i} + \beta_{3jd} E_i)} \qquad [5]$$

where response Y_{ijc} is conditional on the attitudes F_{1i} and F_{2i} and E_i. The influence of ERS on the response is explained by the parameters β_{3jc} representing the influence of E_i.

With respect to the attitudes, the parameters β_{1jc} and β_{2jc} are constrained to increase monotonically across the response scale by the restriction of the parameters as $\beta_{1j} \cdot c$ and $\beta_{2j} \cdot c$. A change from one category to the next (for example from 1 to 2) would denote an increase of β_{1j} by one since the difference between categories – denoted by c – equals one. In this way, a more parsimonious model can be estimated: only one parameter is needed for each item, assuming the distance between category 1 and 2 to be equal to the distances between the other adjacent categories. This adjacent-category ordinal logit specification can be implemented within the multinomial logit model so that the parameters reflecting ERS are specified nominally while the parameters[5] describing the substantive factors are constrained to

[5] In this paper, the effects of the latent variables on the item responses are referred to as factor loadings as is usual in CFA. Due to the discrete specification of the observed variables in LCFA, these effects actually are logit coefficients. Since the factor loadings and logit coefficients are conceptually equal and the only difference is the specification of the observed variables, we refer to these effects as factor loadings.

monotonicity. Note that the model for the latent means and (co)variances is also an adjacent-category logit model as the factors are specified as ordinal variables (see Appendix A).

How ERS leads to Measurement Inequivalence

Metric or scalar equivalence may be violated in only one, a few or all items. If all items in both item subsets are affected, this may be caused by a difference in the style of responding between groups, because the response style affects all items simultaneously. In other words, the presence of a response style that differs across groups is likely to import measurement inequivalence in the data. Unfortunately, most comparative studies on measurement equivalence among culturally diverse populations focus on the detection of measurement inequivalence without correcting for ERS (Mullen, 1995; Myers, Calantone, Page Jr., & Taylor, 2000; Raju, Laffitte, & Byrne, 2002; Reise, Widaman, & Pugh, 1993; Steenkamp & Baumgartner, 1998).

To show which model parameters appear as inequivalent as a consequence of ERS, we generated a data set where groups are simulated to differ in ERS. Previous studies within the latent variable framework simulated group differences in ERS by generating the item intercepts to differ across groups (Meade & Lautenschlager, 2004a, 2004b). A disadvantage of this approach is that it assumes that ERS violates scalar equivalence and invalidates the possibility that ERS violates other forms of equivalence. Cheung and Rensvold (2000) explicitly examined how ERS affects the model parameters and simulated the group differences in ERS by generating group differences in the response patterns, with a group that was severely subject to ERS having many extreme answers. By running SEM models on this data set, they showed that the difference in the response patterns leads to inequivalent intercepts and factor loadings. More specifically, the group subject to a high level of ERS has higher loadings and lower intercepts than the group subject to a low level of ERS (Cheung & Rensvold, 2000).

As we discussed before, these results based on the SEM approach should be viewed with caution because ERS violates the assumption of linearity. Therefore, we generated a data set based on a latent variable model that detects and corrects for ERS by specifying the observed responses nominally with respect to the ERS factor and ordinally with respect to the substantive factors. To detect measurement inequivalence, we extend the model in [5] by including an exploratory group variable g as follows:

$$P(Y_{ij} = c \mid F_{1i}, F_{2i}, E_i, g) = \frac{\exp(\beta_{0jgc} + \beta_{1jg} c\, F_{1i} + \beta_{2jg} c\, F_{2i} + \beta_{3jc}\, E_i)}{\sum_{d=1}^{C} \exp(\beta_{0jgd} + \beta_{1jg} d\, F_{1i} + \beta_{2jg} d\, F_{2i} + \beta_{3jd}\, E_i)} \quad [6]$$

The response Y_{ij} is explained by the item parameters β_{0jgc} representing the group specific category intercept, the ordinally-restricted parameters $\beta_{1jg}c$ and $\beta_{2jg}c$ representing the group specific influence of F_{1i} and respectively F_{2i}, and the unrestricted parameters β_{3jc} representing the influence of E_i on the item responses. Note that although the parameters β_{3jc} (representing the influence of the style factor) are restricted to equality across groups (no superscript g), this assumption could be relaxed to investigate how groups differ in ERS. We allowed for group differences in the latent group means of the factor that measures ERS. By comparing models that correct and do not correct for ERS, we examine which model parameters appear to be inequivalent as the result of these group differences in ERS.

We generate a data set by a latent factor model in which five 5-category variables are related to two continuous latent variables measuring one attitude and ERS. Three groups, each consisting of 1000 observations, are assumed to differ in their style of responding by specifying different latent means for each group ($\mu_g = 2$, 0 and -2). To ensure that the group differences in the latent means of the attitude and ERS are not confounded, the groups do not differ in the attitude. Note that although only one attitude is included in the model, the effects of the attitude and ERS on the items cannot be confounded because the items are restricted to relate to the attitude in a monotone manner and allowed to relate to ERS in a non-monotone manner. This is accomplished by assigning values to the parameters describing the relationship between the manifest and latent variables. For each item, the category parameters are restricted to relate to the attitude in a monotone way by sequentially assuming the values -2, -1, 0, 1, and 2. These values restrict the effect of the attitude to be the same for each pair of categories as the inter-category distance is always 1. Furthermore, a respondent who scores highly on the dimension is likely to choose the positive outer category (see the SPVA example depicted in Figure 3). For the category item parameters relating to ERS, the values 1.5, -1, -1, -1, and 1.5 are assumed, indicating a non-monotone pattern. These values signify that respondents with a high score of ERS are more likely to select the outer categories than the categories 2, 3 or 4 (see also Figure 3).

We estimated various models on this generated data set to find out how the group differences in ERS import measurement inequivalence in the results. For this purpose we used the syntax module of the Latent GOLD 4.5 program[6] (Vermunt & Magidson, 2008), a program for the Maximum Likelihood estimation of latent class

[6] See Appendix B for the details of model specification using the syntax module of the Latent GOLD 4.5 program.

models[7] and other types of latent variable models (see Appendix B). A comparison of the results between the models that control and do not control for the response style informs us about how ERS affects the model parameters. To compare the models, we report in Table 1 both the log-likelihood values and the Bayesian Information Criterion (BIC) values. The latter fit measure introduces a penalty for the sample size and the number of parameters (Burnham & Anderson, 2004; Raftery, 1999). The best model in terms of fit and parsimony has the lowest value of BIC. Note that although we choose to simulate the data using continuous factors[8], we estimate the models with ordinal factors to preserve continuity with the other models presented in the paper.

[Insert Table 1 about here]

Table 1 reports the fit statistics for the models estimated on the simulated data set. For the models that correct for ERS, the model of scalar equivalence has the best model fit (model C_{ERS}). This result is expected as the measurements are simulated to be scalar equivalent. More interesting is that among the models that do not correct for ERS the model of metric equivalence is preferred (Model B). Thus, group differences in ERS cause equivalent measurements to *appear* as group differences in parameters β_{0jc}; that

---

[7] A well-known problem with these models is the occurrence of local minima. Here, we deal with this problem by using 100 sets of starting values, 250 iterations using the Expectation-Maximization algorithm and a low minimum convergence criterion (1e-005).

[8] Theoretically, the style factor represents a continuous dimension; however, we approach the dimension as ordinal to avoid inappropriate assumptions about normal distribution of respondents on this dimension and to facilitate the estimation process. In the Latent Class Factor Approach the latent variables have three categories (the latent classes) that are restricted to be ordinally located with equal distances on an underlying continuous dimension (see Appendix A).

is, as groups having unequal intercepts. These results show that correcting for ERS is crucial in making valid conclusions with respect to metric and scalar equivalence.

An Empirical Application

We now illustrate the importance of this finding with a dataset collected[9] in 2002 among the four largest ethnic minorities in the Netherlands, namely Turks, Moroccans, Surinamese and Antilleans. Response rates lie between 44% for Surinamese and Antilleans and 52% for Turks. The SPVA survey treats the Social Position and Utility Use of Ethnic Minorities by focusing on the cultural, economic and social life of ethnic minorities in the Netherlands (Dagevos, Gijsberts, & Van Praag, 2003). In this application, we use two sets of five questions, each subset referring to an aspect of the cultural dimension; that is, family values and the attitude toward the Dutch society. One item subset contains three items that are negatively worded; the other subset contains one item negatively worded. The respondents were asked to report on a fully labeled 5-point Likert scale, ranging from totally agree (1) to totally disagree (5), with neither agree nor disagree as a neutral midpoint. For the statistical analyses, the category order was reversed in order to facilitate the interpretation of scale which now runs from a negative (1) toward a positive (5) response to the items. Descriptive statistics of the items are reported in Table 2.

[Insert Table 2 about here]

We estimated various models for our data set; as in the generated data example, the model selection is based on log-likelihood and BIC values. We use a LCFA model that corrects for ERS by including an ERS factor and simultaneously tests for

[9] Since the data is collected among households, we only include the answers given by the heads of the households to have independent observations.

measurement equivalence described in [6].

As in the generated example, we specified six models of which three models correct for ERS. The first model depicts measurement inequivalence where the latent means, the latent (co-) variances, the intercepts and the factor loadings are simultaneously allowed to differ across groups (see Figure 1c). By using the situation of measurement inequivalence as a baseline model, we avoid inappropriate assumptions about measurement equivalence. To test for metric equivalence, this baseline model is compared to a more restrictive model where the factor loadings are restricted to equality across groups. Scalar equivalence is tested by additionally restricting the intercepts to equality across groups[10] (Byrne, Shavelson, & Muthen, 1989; Vandenberg & Lance, 2000); if the model fit does not deteriorate significantly, the restrictions are confirmed to be appropriate. To investigate whether the conclusions with respect to metric and scalar equivalence are affected by ERS, these three models are re-estimated while controlling for ERS. In Table 3, the fit statistics are reported for all models.

[Insert Table 3 about here]

Comparing Models D, E and F in Table 3 illustrates that according to the BIC values Model D – the baseline model – is preferred: the model in which the factor loadings as well as the intercepts are allowed to differ between groups. The increase in BIC values of Models E and F compared to Model D confirm that the equality restrictions are inappropriate. However, this conclusion clearly alters when the same analyses are

[10] The model selection does not include partial equivalence models where only some of the factor loadings are restricted to equality because this paper focuses on how the presence of a response style affects measurement equivalence. Since a response style is assumed to affect all items simultaneously, it presumably violates equivalence of all items simultaneously.

controlled for ERS in Models D_{ERS}, E_{ERS} and F_{ERS}. First, the model fit improves substantially between the models with and without a style factor which illustrates the necessity of introducing an ERS factor into the model. Controlling for ERS causes the model with unequal intercepts and equal factor loadings (Model E_{ERS}) to fit best. Thus, accounting for ERS yields a substantive reduction in the group differences in the factor loadings. The magnitude of the reduction is evaluated by inspecting the WALD statistics which allow to test whether β_{1jg} is equal across groups g for each item j belonging to factor F_1 (Buse, 1982; Vermunt & Magidson, 2005, p. 69).

[Insert Table 4 about here]

Table 4 reports the WALD statistics for the group differences in the intercepts as well as the factor loadings of Model D and Model D_{ERS}. The large number of substantial reductions in the WALD statistics show that controlling for ERS decreases the group differences in both the intercepts and the factor loadings. However, the decrease in values of the WALD statistics is larger with respect to the intercepts than with respect to the factor loadings. These results indicate that the group differences are diminished substantially by controlling for the response style, and especially the group differences in the intercepts. The fact that the measurements are not scalar equivalent, even after controlling for ERS, is likely to be caused by unknown causes not taken into account in this model. These findings are in accordance with the results in the generated data example where controlling for ERS decreased the group differences in the intercepts. Therefore, we conclude that the ERS factor partly explains the group differences in the intercepts of the set of items that were taken from the SPVA data.

Conclusion

This paper demonstrates that ERS imports inequivalence in measurements among groups if this response style is not explicitly controlled for. This conclusion is drawn from separate findings. First, correcting for ERS reduces the measurement inequivalence in both the item intercepts and the factor loadings. This conclusion holds in the case of secondary data as well as in the generated data set. Using a LCFA multiple-group analysis, we find that the presence of ERS violates metric and scalar equivalence in the models that do not control for ERS. In the generated data set the presence of group differences in ERS violates scalar equivalence. In the Dutch data set of the four largest minorities, the group differences in ERS violate scalar as well as metric equivalence. The group differences in the intercepts that remain after controlling for ERS are ascribed to unknown group differences not considered here.

In this paper we focused primarily on the extent to which ERS leads to measurement inequivalence when comparing attitudes across culturally diverse groups. However, the model can be extended by including covariates to control for other possible socio-demographic or cultural group differences, for instance language proficiency, level of education or gender. Additionally, the assumption that ERS is measured equivalently across culturally diverse groups could be relaxed by allowing the factor loadings related to ERS to differ between groups. Finally, one could specify a more parsimonious model by assuming that ERS affects all items similarly.

To conclude, in this paper we show that equivalent measurements could appear as inequivalent measurements because of a group difference in ERS for which the researcher does not control. Thus, to investigate metric and scalar equivalence adequately, one should control for ERS. We have shown that Latent Class Factor Analysis is a straightforward method to appropriately investigate measurement equivalence because it enables multiple-group analyses while simultaneously correcting for ERS.

References

Adcock, R., & Collier, D. (2001). Measurement Validity: A Shared Standard for Qualitative and Quantitative Research. *American Political Science Review, 95*(3), 529–546.

Agresti, A. (2002). Categorical Data Analysis. New York: John Wiley & Sons.

Bachman, J. G., & O'Malley, P. M. (1984). Yea-Saying, Nay-Saying, and Going to Extremes: Black-White Differences in Response Styles. *Public Opinion Quarterly, 48*, 491-509.

Berry, J. W., Poortinga, Y. H., Segall, M. H., & Dasen, P. R. (2002). *Cross-Cultural Psychology. Research and Applications.* (Second ed.). Cambridge (UK): Cambridge University Press.

Billiet, J. B., & McClendon, M. J. (2000). Modeling Acquiescence in Measurement Models for Two Balanced Sets of Items. *Structural Equation Modeling, 7*(4), 608-628.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*: John Wiley & Sons, Inc.

Bollen, K. A. (2002). Latent Variables in Psychology and the Social Sciences. *Annual Review Psychology, 53*, 605-634.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel Inference. Understanding AIC and BIC in Model Selection. *Sociological Methods and Research, 33*(2), 261-304.

Buse, A. (1982). The Likelihood Ratio, Wlad, and Lagrange Multiplier Tests: An Expository Note. *The American Statistician, 36*(3), 153-157.

Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for the Equivalence of Factor Covariance and Mean Structures: The Issue of Partial Measurement Invariance. *Psychological Bulletin, 105*(3), 456-466.

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.

Byrne, B. M., & Watkins, D. (2003). The Issue of Measurement Invariance Revisited. *Journal of Cross-Cultural Psychology, 34*(2), 155-175.

Cheung, G. W., & Rensvold, R. B. (2000). Assessing Extreme and Acquiescence Response Sets in Cross-Cultural Research Using Structural Equation Modeling. *Journal of Cross-Cultural Psychology, 31*(2), 187-212.

Chun, K.-T., Campbell, J. B., & Yoo, J. H. (1974). Extreme Response Style in Corss-Cultural Research. A Reminder. *Journal of Cross-Cultural Psychology, 5*(4), 465-480.

Dagevos, J., Gijsberts, M., & Van Praag, C. (2003). *Rapportage Minderheden 2003. Onderwijs, Arbeid en Sociaal-Culturele Integratie*. Den Haag: Sociaal en Cultureel Planbureau.

De Jong, M. G., Steenkamp, J.-B. E. M., Fox, J.-P., & Baumgartner, H. (2008). Using Item Response Theory to Measure Extreme Response Style in Marketing Research: A Global Investigation. *Journal of Marketing Research, 45*(1), 104 - 115.

Eid, M., Langeheine, R., & Diener, E. (2003). Comparing Typological Structures Across Cultures by Multigroup Latent Class Analysis. A Primer. *Journal of Cross-Cultural Psychology, 34*(2), 195-210.

Gibbons, J. L., Zellner, J. A., & Rudek, D. J. (1999). Effects of Language and Meaningfulness on the Use of Extreme Response Style by Spanish-English Bilinguals. *Cross-Cultural Research, 33*(4), 369-381.

Heinen, T. (1996). *Latent Class and Discrete Latent Trait Models: Similarities and Differences*. Thousand Oaks: Sage Publications.

Hui, H. C., & Triandis, H. C. (1985). Measurement in Cross-Cultural Psychology. A Review and Comparison of Strategies. *Journal of Cross-Cultural Psychology, 16*(2), 131-152.

Hui, H. C., & Triandis, H. C. (1989). Effects of Culture and Response Format on Extreme Response Style. *Journal of Cross-Cultural Psychology, 20*(3), 296-309.

Javeline, D. (1999). Response effects in polite cultures: a test of acquiescence in Kazakhstan. *The Public Opinion Quarterly, 63*(1), 1-28.

Johnson, T. P., Kulesa, P., Cho, Y. I., & Shavitt, S. (2005). The Relation Between Culture and Response Styles: Evidence From 19 Countries. *Journal of Cross-Cultural Psychology, 36*(2), 264-277.

Johnson, T. P., & Van de Vijver, F. J. R. (2003). Social desirability in cross-cultural research. In J. Harkness, F. J. R. Van de Vijver & P. P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 195-204). Hoboken, new Jersey: John Wiley & Sons.

Joreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36*(4), 409-426.

Krosnick, J. A. (1999). Survey Research. *Annual Review Psychology, 50*, 537-567.

Marin, G., Gamba, R. J., & Marin, B. V. (1992). Extreme Response Style and Acquiescence Among Hispanics. *Journal of Cross-Cultural Psychology, 23*(4), 498-509.

Meade, A. W., & Lautenschlager, G. J. (2004a). A Comparison of Item Response Theory and Confirmatory Factor Analytical Methodologies for Establishing Measurement Equivalence/ Invariance. *Organizational Research Methods, 7*(4), 361-388.

Meade, A. W., & Lautenschlager, G. J. (2004b). A Monte-Carlo Study of Confirmatory Factor Analytic Tests of Measurement Equivalence/Invariance. *Structural Equation Modeling, 11*(1), 60-72.

Meredith, W. (1993). Measurement Invariance, Factor Analysis and Factorial Invariance. *Psychometrika, 58*(4), 525-543.

Millsap, R. E. (1995). Measurement invariance, predictive invariance, and the duality paradox. *Multivariate Behavioral Research, 30*(4), 577-605.

Moors, G. (2003). Diagnosing Response Style Behavior by means of A Latent-Class Factor Approach. Socio-Demographic Correlates of Gender Role Attitudes and Perceptions of Ethnic Discrimination Reexamined. *Quality and Quantity, 37*, 277-302.

Moors, G. (2004). Facts and Artefacts in the Comparison of Attitudes Among Ethnic Minorities. A Multigroup Latent Class Structure Model with Adjustment for Response Style Behavior. *European Sociological Review, 20*(4), 303-320.

Mullen, M. R. (1995). Diagnosing Measurment Equivalence in Cross-National Research. *Journal of International Business Studies, 26*(3), 573-596.

Myers, M. B., Calantone, R. J., Page Jr., T. J., & Taylor, C. R. (2000). Academic Insights: An Application of Multiple-group Causal Models in Assessing Cross-Cultural Measurement Equivalence. *Journal of International Marketing, 8*(4), 108-121.

Poortinga, Y. H., & Van de Vijver, F. J. R. (1987). Explaining Cross-Cultural Differences: Bias Analysis and Beyond. *Journal of Cross-Cultural Psychology, 18*(3), 259-282.

Raftery, A. E. (1999). Bayes Factors and BIC. Comment on "A Critique of the Bayesian Information Criterion for Model Selection". *Sociological Methods and Research, 27*(3), 411-427.

Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement Equivalence: A Comparison of Methods Based on Confirmatory Factor Analysis and Item Response Theory. *Journal of Applied Psychology, 87*(3), 517-529.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory Factor Analysis and Item Response Theory: Two Approaches for Exploring Measurement Invariance. *Psychological Bulletin, 114*(3), 552-566.

Sijtsma, K., & Molenaar, I. W. (2002). *Introduction to Nonparametric Item response Theory*. Thousands Oaks, California: Sage Publications.

Skondral, A., & Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton, Florida: Chapman & Hall/CRC.

Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing Measurement Invariance in Cross-National Consumer Research. *Journal of Consumer Research, 25*(1), 78-90.

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking About Answers. The Application of Cognitive Processes to Survey Methodology*. San Francisco: Jossey-Bass Publishers.

Tourangeau, R. (2003). Cognitive aspects of survey measurement and mismeasurement. *International Journal of Public Opinion Research 15*(1), 3-7.

Triandis, H. C. (1990). Theoretical concepts that are applicable to the analysis of ethnocentrism. In R. W. Brislin (Ed.), *Applied Cross-Cultural Psychology* (Vol. 14). Newbury Park, California: Sage Publications, Inc.

Van de Vijver, F. J. R. (1998). Towards a Theory of Bias and Equivalence. *ZUMA-Nachrichten Spezial*, 41-65.

Van de Vijver, F. J. R., & Leung, K. (1997). *Methods and Data Analysis for Cross-Cultural Research* (Vol. 1). Thousand Oaks, California: Sage Publications.

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.

Vermunt, J. K., & Magidson, J. (2004). Latent Class Analysis. In E. S. Lewis-Beck, A. Bryman & T. F. Liao (Eds.), *The Sage Encyclopedia of Social Sciences Research Methods* (pp. 549-553). Thousand Oaks: Sage Publications.

Vermunt, J. K., & Magidson, J. (2005). *Technical Guide for Latent GOLD 4.0: Basic and Advanced*. Belmont Massachusetts: Statistical Innovations Inc.

Vermunt, J. K., & Magidson, J. (2008). *LG-SyntaxTM User's Guide: Manual for Latent GOLD 4.5 Syntax Module*. Belmont, MA: Statistical Innovations, Inc.

Wallace, R. A., & Wolf, A. (1998). *Contemporary Sociological Theory. Expanding the Classical Tradition.* (5th ed.). Upper Saddle River, New Jersey: Prentice Hall.

Table 1.

Model selection estimated with generated data (N=3000).

Model	Fit Statistics		
	Log-Likelihood	BIC (based on LL)	Number of parameters
Without a correction for ERS (two factors)			
A) Measurement inequivalence	-15976,6	32585,8	79
B) Metric equivalence	-16011,5	32575,4	69
C) Scalar equivalence	-19569,0	39370,2	29
With a correction for ERS (three factors)			
A_{ERS}) Measurement inequivalence	-15055,7	30936,1	103
B_{ERS}) Metric equivalence	-15061,1	30866,8	93
C_{ERS}) Scalar equivalence	-15130,0	30684,4	53

Table 2.

Mean observed item response per ethnic group (N=3576)

		Turks		Moroccans		Surinamese		Antilleans	
Factor 1: Attitude toward the Dutch society									
Item 1	In the Netherlands immigrants get many opportunities	3.53	(1.056)	3.42	(1.073)	3.26	(1.105)	3.25	(1.146)
Item 2	The Netherlands is hostile to immigrants	2.80	(1.015)	2.46	(.878)	2.39	(.877)	2.52	(.905)
Item 3	In the Netherlands your civil rights as an immigrant are respected	3.40	(.905)	3.55	(.858)	3.52	(.861)	3.45	(.842)
Item 4	The Netherlands is a hospitable country for immigrants	3.03	(.972)	3.48	(.915)	3.70	(.884)	3.60	(.908)
Item 5	The Netherlands is tolerant towards foreign cultures	3.83	(.911)	3.58	(.874)	3.83	(.816)	3.69	(.827)
Factor 2: Family values									
Item 6	A man and woman are allowed to live together without being married	2.55	(1.249)	2.12	(1.104)	3.90	(1.046)	3.95	(1.041)
Item 7	Married people with children should not divorce	3.12	(1.188)	2.67	(1.138)	2.64	(1.090)	2.42	(1.051)
Item 8	The best family is: two married parents with children	3.54	(1.093)	4.05	(.932)	3.47	(1.219)	3.40	(1.212)
Item 9	A daughter aged 17 is allowed to live by herself	2.00	(.934)	1.94	(.953)	2.41	(1.039)	2.60	(1.139)
Item 10	The opinion of the parents should be important in the choice of a partner for their child [a]	3.45	(1.043)	3.48	(1.089)	2.64	(1.108)	2.47	(1.069)
N (unweighted)		914		858		1022		782	
Response Rate		52		52		44		51	

Note. Items 2, 7, 8 and 10 are formulated in reversed manner where a positive answer indicates a conservative attitude. For the other items a positive answer indicates a modern attitude. Standard deviations between parentheses.

Table 3.

Model selection estimated with SPVA data (N=3576).

Model	Fit Statistics		
	Log-Likelihood	BIC (based on LL)	Number of parameters
Without a correction for ERS (two factors)			
D) Measurement inequivalence	-44153.0	90057.0	214
E) Metric equivalence	-44283.8	90073.1	184
F) Scalar equivalence	-45088.6	90700.8	64
With a correction for ERS (three factors)			
D_{ERS}) Measurement inequivalence	-41819.7	85758.5	259
E_{ERS}) Metric equivalence	-41901.3	85676.2	229
F_{ERS}) Scalar equivalence	-42564.0	86019.8	109

Table 4.

WALD statistics for group differences in the intercepts and loadings for model D and D_{ERs}.

	Intercepts		Factor loadings	
	Model D [a]	Model D_{ERS} [a]	Model D	Model D_{ERS}
Item 1	55.21***	131.87***	8.73*	18.68***
Item 2	209.20***	63.38***	4.42	6.13
Item 3	97.38***	36.37***	8.31*	5.42
Item 4	220.07***	71.68***	23.37***	3.76
Item 5	150.91***	146.37***	48.46***	46.73***
Item 6	550.63***	243.18***	20.75***	16.33**
Item 7	134.07***	63.75***	29.67***	27.08***
Item 8	100.34***	90.10***	13.01**	15.15**
Item 9	65.05***	46.66***	43.50***	32.66***
Item 10	302.62***	124.84***	16.71**	4.36

Note. start values are used to ascertain that the four ethnic groups have positive factor loading parameters; Model D: BIC=90086; Model D_{ERS}: BIC= 85724.

* the groups differ at a significance level of $p<.05$;

** the groups differ at a level of $p<.01$;

*** the groups differ at a significance level of $p<.001$

Figure 1.

The pooled approach to multiple-group analyses of metric and scalar equivalence.

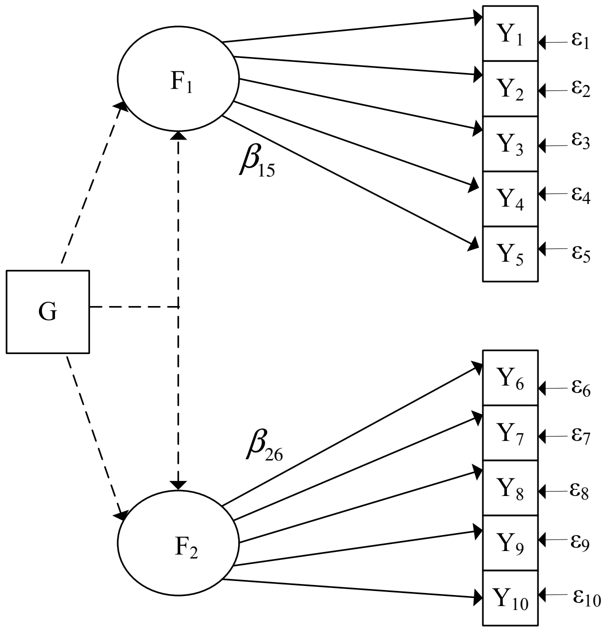


Figure 1a: scalar equivalence

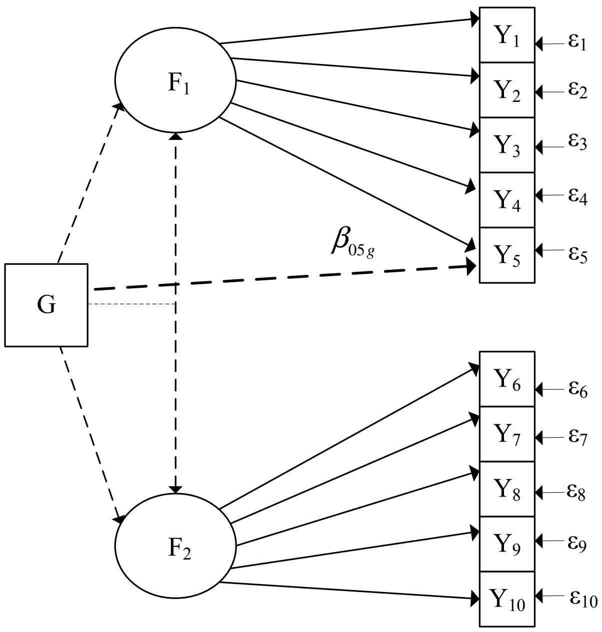


Figure 1b: metric equivalence

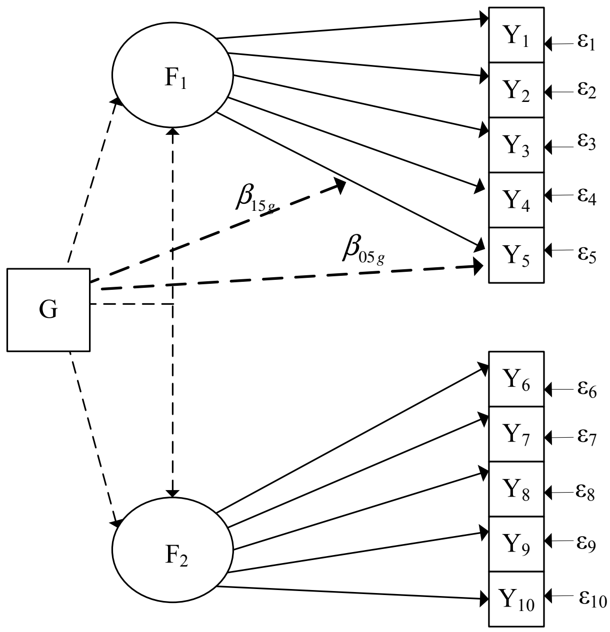


Figure 1c: measurement inequivalence

Note. The dashed arrows indicate that the groups differ with respect to these relationships. The parameters are not denoted as category-specific item parameters to simplify the graphical display. In Figure 1a, the item parameters β_{1j} and β_{2j} are described by β_{15} and β_{26} for item 5 and 6. In Figure 1b, the inequivalent intercepts of item 5 are indicated by β_{05g}, and in Figure 1c the inequivalent factor loadings of item 5 are indicated by β_{15g}.

Figure 2.

The latent variable model for the detection of a response style.

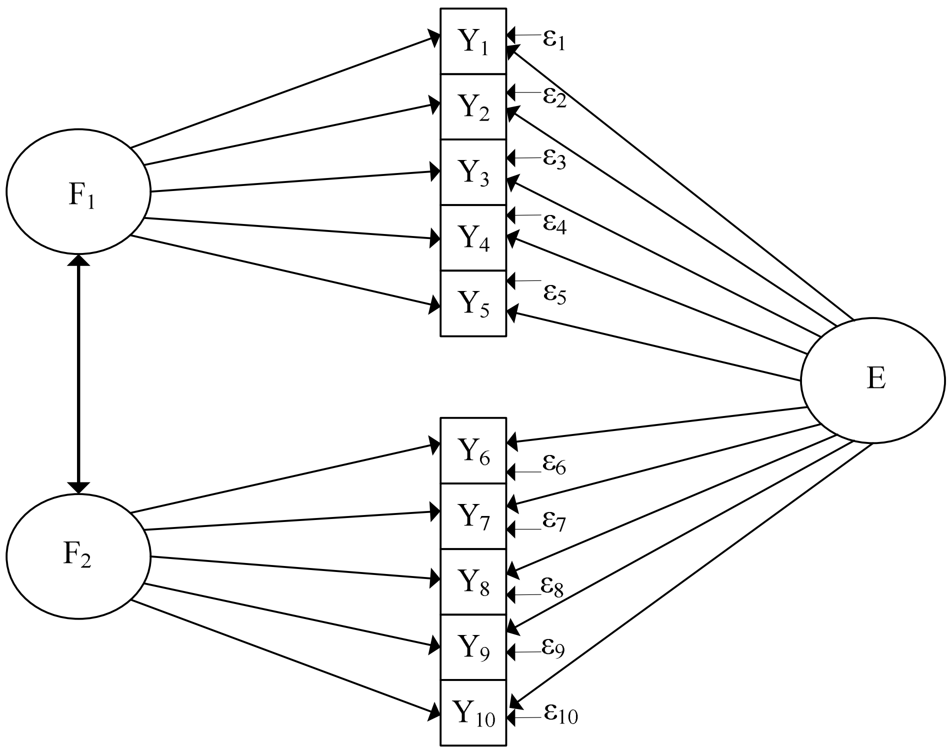
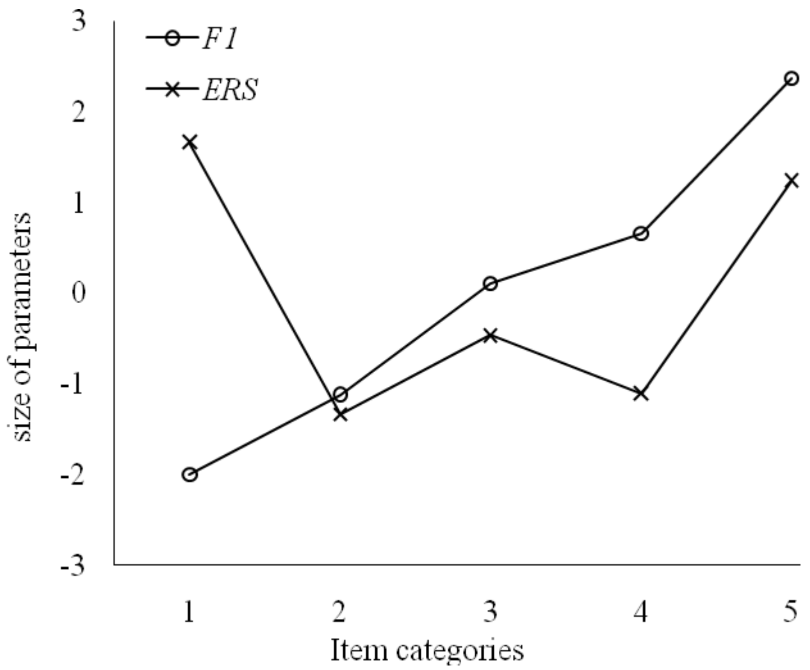


Figure 3.

The size of the category-specific item parameters relating the responses to the factors F1 and ERS estimated with SPVA data (N=3576).



Note. The graphs are based on the model parameters for the first item, estimated under the assumption of scalar equivalence and all observed variables are nominally specified with respect to all latent variables. Parameters are logit coefficients.

Appendix A

The Model for Latent Means and (Co) variances

The model for the means and the (co)variances of latent variable k for group g can be represented as:

$$\mathbf{F}_i \sim N(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g) \qquad [7]$$

where the multivariate vector $\mathbf{F}_i$ is normally distributed with a vector containing group specific means $\boldsymbol{\mu}_g$ and a group specific co-variance matrix $\boldsymbol{\Sigma}_g$. Using dummy coding, the factor means are restricted to zero in one group. In the empirical application of the model using the SPVA data set this reference group is the Turks.

In the case of the regression model used for the ordinal latent variables, an adjacent-category ordinal logit model as is described in [5] is used:

$$P(F_i = k \mid g) = \frac{\exp\left(\gamma_{0k} + \gamma_{1g} \cdot k\right)}{\sum_{k'=1}^{3} \exp\left(\gamma_{0k'} + \gamma_{1g} \cdot k'\right)}, \qquad [8]$$

where the probability that respondent *i* belongs to class *k* of variable *F* is estimated given the respondent's group membership *g*. As one can see, the model has a similar structure as the model in equation [6] where the observed responses are modeled as a function of latent variables. In equation [7] the latent variables are modeled as a function of ethnicity. Note that the group structure in the latent means and co-variances described in equations [7, 8] applies to every estimated model in this paper.

Appendix B

Latent GOLD 4.5 Syntax used for assessing Measurement Equivalence

We used the syntax module of Latent GOLD 4.5 to estimate models A to F and Model A_{ERS} to F_{ERS} from Table 1 and Table 3. The variables and equations sections of the syntax file for the most complex model D_{ERS} is as follows:

```
.
  variables
     dependent
        Y1 nominal, Y2 nominal, Y3 nominal, Y4 nominal,
        Y5 nominal, Y6 nominal, Y7 nominal, Y8 nominal,
        Y9 nominal, Y10 nominal;
      independent ethnicity nominal coding=first;
      latent
         F1      ordinal 3 scores=(-1 0 1),
         F2      ordinal 3 scores=(-1 0 1),
         ERS     ordinal 3 scores=(-1 0 1);
  equations
      F1        <- 1 + ethnicity;
      F2        <- 1 + ethnicity;
      ERS       <- 1 + ethnicity;
      F1        <-> F2 |ethnicity;
      Y1 - Y5   <- 1|ethnicity + (~ord) F1|ethnicity + ERS;
      Y6 - Y10  <- 1|ethnicity + (~ord) F2|ethnicity + ERS;
.
```

In the variables section we provide the relevant information on the dependent, independent, and latent variables to be used in the analysis. The first three equations define the regression models for the latent variables – which contain an intercept (indicated with "1") and an effect of ethnicity – and the fourth defines the association between F1 and F2 which is modelled as a conditional effect depending on the group. In other words, the association between F1 and F2 is group specific. The last two equations define the multinomial regression models for items Y1 to Y5 and Y6 to Y10, respectively. The term "(~ord)" before F1 and F2 indicates that the nominal dependent variable concerned should be treated as ordinal in this term. As an alternative, we could define the items to ordinal instead of nominal and put "(~nom)" before ERS to indicate that the ordinal items should be treated as nominal for these terms.

The other estimated models can easily be derived from this syntax example. For example, removing "|ethnicity" yields a model without ethnic group difference in the intercepts and the factor loadings representing respectively scalar and metric equivalence, removing "(~ord)" yields a model in which the term concerned remains a standard multinomial logit term, and removing ERS from the latent variable definition and the equations yields a model without ERS factor.