Ensuring the cross-group comparability of latent variable models for ranking data with the EPC-interest

–BLINDED VERSION–

ABSTRACT

Ranking data are often encountered in political science, for example when modeling voters' choices in "ranked choice" voting systems or their stated priorities for political goals in a questionnaire. Group comparisons on such measures must first establish whether conclusions of interest are insensitive any differences in measurement–that is, whether the measures are "equivalent". To investigate this, the "EPC-interest", a measure of the sensitivity of conclusions to measurement equivalence, was recently introduced in this journal. However, the measure has not been applied to ranking data.

This paper therefore extends the EPC-interest to the case of latent variable models with ranking data. We demonstrate the approach on a multilevel latent class analysis of three "(post)materialism" measures using ranking data from 67,568 respondents in 48 different countries. In this application, violations of measurement equivalence assumptions reverse substantive conclusions of interest; we show that the EPC-interest can detect this problem.

1

The newly developed methods discussed in this paper have been implemented in user-friendly software for latent variable modeling. Program inputs and data for the examples discussed in this article are provided in the electronic appendix (`http://`[blinded]).

1. INTRODUCTION

Ranking data are central to political science, occurring for example when voters rank candidates in "ranked choice" ("instant-runoff") voting systems such as the Australian federal or London mayoral elections (see e.g Karvonen, 2004; Toplak, 2010), or when respondents rank objects in a questionnaire, for example post-materialist value orientations in the World Values Survey (Inglehart, 1997; Inglehart and Welzel, 2010).

Such procedures yield categorical and multivariate data, and, since they preclude certain joint outcomes, lead to complex dependencies that must be accounted for. Probabilistic models with rankings as outcomes that account for these characteristics were introduced by Thurstone (1927), and extended by Luce (1959) and McFadden (1974).

These approaches treat the choices as perfectly measured; however, rankings may contain measurement error, so that latent variable models become necessary. An example can be found in Inglehart's (1997) value scale measuring (post)materialism (see Table 1). Respondents are asked to rank their first and second priorities out of four, thus leading to a partial ranking of values priorities. These priorities are divided into "materialist" (M in Table 1) and "postmaterialist" (P) priorities. However, there is error in the measurement of "postmaterialism" when using a respondent's ranked priorities. For instance, while valuing a beautiful countryside might be a measure of "postmaterialism" for, say, a college professor, a farmer depending on ecotourism might prioritize

the countryside for more "materialist" reasons. To account for such measurement errors, Croon (1989) introduced a latent class model for ranking data. Kamakura, Wedel and Agrawal (1994) extended this model to include covariates relating to class membership. Other approaches to modeling ranking outcomes with latent variables were discussed by Jackson and Alwin (1980), Francis et al. (2002) and Maydeu-Olivares and Böckenholt (2005).

Whenever imperfectly measured variables are to be compared over groups, a concern is whether such comparisons are "fair"–that is, whether conclusions about the comparisons of interest are uncontaminated by any differences in measurement across the groups (Oberski, 2012). For example, Costa Rica's economy depends strongly upon ecotourism, whereas Russia's does not, conceivably causing measurement errors in the item mentioned to differ over these countries. Such differences may lead a researcher to erroneously conclude that Costa Ricans are less materialist than others. Similarly, when comparing countries with different percentages of women in parliament, the country with the higher percentage[1], Costa Rica (39%), may erroneously appear to have fewer "materialists" than Russia, the country with the lower percentage (14%). Thus, measurement differences may contaminate conclusions about both mean differences and relationships between variables.

The assumption that groups are comparable in measurement is often called "measurement equivalence" or "measurement invariance", and evaluating measurement invariance is therefore a concern in any comparative study (see Steenkamp and Baumgartner, 1998; Vandenberg and Lance, 2000; Schmitt and Kuljanin, 2008, for reviews).Recently, Oberski (2014) introduced a measure to evaluate measurement invariance in terms of the substantive parameters of interest, the "EPC-interest". The researcher initially fits a model that assumes full measurement invariance; EPC-interest then estimates

---

[1]See Section 2 for data sources.

the expected change in the parameters of interest that would occur if assumptions of measurement invariance were (partially) relaxed. To continue the previous example, EPC-interest might indicate how much of the substantive mean difference in "materialism" between Costa Rica and Russia is contaminated by possible measurement differences, and whether this contamination is capable of reversing the conclusions or not. While useful, currently EPC-interest is only available for linear structural equation models for continuous data. Categorical observed and latent variables are, however, commonly found in the social sciences. Ranking data, moreover, have never been examined for cross-group equivalence in this sense.

This paper therefore develops the EPC-interest in more generality. This allows for application of the EPC-interest approach to measurement equivalence evaluation in, for example, IRT models (see also Oberski and Vermunt, 2013), ordinal factor analysis, latent class analysis, and to models for ranking data. We demonstrate the application of the EPC-interest to ranking data using comparison of the Inglehart (1997) (post)materialism ranking measures over 48 countries. Each country's GDP per capita and percentage of women in parliament is related to the country proportion in categories of "(post)materialism" as measured by three separate ranking tasks. Since the measurement of these (post)materialism categories may be flawed, (post)materialism becomes a latent variable that is discrete: a "latent class" variable (Böckenholt, 2002; Moors and Vermunt, 2007). The measurement model for this variable accounts for the intrinsic dependencies between respondents' ranking choices, while the structural model accounts for dependencies between respondents within a country using a multilevel specification with a nonparametric random effect distribution (Vermunt, 2003). The EPC-interest is then applied to investigate the impact of measurement invariance assumptions on the latent relationship estimates of interest.

Table 1: Value priorities to be ranked for the three WVS 2010–2012 ranking sets. "Materialist" concerns are marked "M" while "postmaterialist" concerns are marked "P". The wording is "People sometimes talk about what the aims of this country should be for the next ten years. On this card are listed some of the goals which different people would give top priority. Would you please say which one of these you, yourself, consider the most important?"; "And which would be the next most important?".

Option #	M/P	Value wording
Set A		
1.	M	A high level of economic growth
2.	M	Making sure this country has strong defense forces
3.	P	Seeing that people have more say about how things are done at their jobs and in their communities
4.	P	Trying to make our cities and countryside more beautiful
Set B		
1.	M	Maintaining order in the nation
2.	P	Giving people more say in important government decisions
3.	M	Fighting rising prices
4.	P	Protecting freedom of speech
Set C		
1.	M	A stable economy
2.	P	Progress toward a less impersonal and more humane society
3.	P	Progress toward a society in which ideas count more than money
4.	M	The fight against crime

Section 2 presents the data and the postmaterialism theory used in the example application model. Since the obtained measures are rankings, Section 3 then presents the multilevel latent class model for partial rankings. We then present the EPC-interest for such and other latent variable models in Section 4, after which the results of applying the EPC-interest to the postmaterialism model and data is discussed in Section 5. Finally, Section 6 provides some conclusions and discusses limitations and further connections with other work.

2. BACKGROUND AND DATA FOR EXAMPLE APPLICATION

Our example application employs the 2010–2012 World Values Survey[2] (WVS) comprising $n = 67,568$ respondents in 48 different countries (Appendix A provides a full list of countries). The WVS questionnaire includes Inglehart (1977)'s extended (post)materialism questions, developed to measure political values priorities. This extended version includes three sets of four priorities (Table 1). Of these, set B in Table 1 is known as the "short scale" that is commonly used in research on values priorities.

(Post)materialism theory is based on the "dual-hypotheses model" (Inglehart, 1981, p. 881). The first hypothesis holds that an individual's priority set reflects the socio-economic environment because one attaches relatively more importance to scarce objects. In line with this scarcity hypothesis, Inglehart (1997) and Inglehart and Welzel (2010) argued that socio-economic development indicators per country such as GDP per capita will correlate strongly with aggregated values scores.

The second, "socialization", hypothesis stresses the importance of experiences in the "formative" years. By a process of social metabolism, society gradually changes from materialist to postmaterialist. In line with the socialization hypothesis, associations of aggregated values priorities with socio-cultural variations across countries such as religious heritage and gender equality have been predicted and found by Inglehart, Norris and Welzel (2002).

In short, substantive interest in this literature focuses on the associations between (post)materialism and both socio-economic and socio-cultural country-level aggregates. We will follow these authors and examine the aggregate relationship of values priorities with log-GDP per capita and the percentage of women in parliament. These country-

---

[2]http://www.worldvaluessurvey.org/

6

level variables were obtained from the World Bank database[3] using the `WDI` package (Arel-Bundock, 2013) in `R 3.0.2` (R Core Team, 2012).

3. MODEL

To examine differences in (post)materialism with respect to GDP per capita and the proportion of women in parliament, we develop a latent variable model for ranking data. The advantages of this approach are that there is no need to establish a priori rules for combining the three ranking tasks into a measure of "(post)materialism"; that it does not presuppose a fixed number of "(post)materialism" classes; that it allows for measurement error in "(post)materialism"; and that it allows for an examination of the impact of violations of measurement invariance on the conclusions of interest.

In each of the C countries we observe $\{n_1, \ldots, n_C\}$ respondents. Each respondent i has ranked their first and second choice on three ranking tasks A, B, and C (see Table 1); that is, only first choice and second choice are known. We assume that each object (e.g. value priority) has a particular "utility" and that the probability of choosing an object for each rank is proportional to that utility (Luce, 1959; Böckenholt, 2002, p. 171). Following Croon (1989) and Böckenholt (2002, p. 172), we then model the probability of these rankings conditionally on unit i's membership of latent class variable X, so that object utilities can differ over respondents in different classes. This latent class variable X in our case will represent "(post)materialism". For example, for ranking task A in country c, respondent i's choices for first and second place are modeled as

$$P(A_{1ic} = a_1, A_{2ic} = a_2 | X_{ic} = x) = \frac{\omega_{a_1 x}}{\sum_k \omega_{kx}} \frac{\omega_{a_2 x}}{\sum_{k \neq a_1} \omega_{kx}}, \tag{1}$$

---

[3] `http://data.worldbank.org/`

where ω_{kx} is the "utility" of object k for respondents in class x. Crucially, the choice for second place, A_{2ic}, is modeled while excluding the alternative already chosen for first place, A_{1ic}. In other words, we model a sequential ranking process in which first place is chosen from all options, then second place is chosen from the remaining options. This dependency intrinsic to ranking tasks makes the model used here different from standard models for qualitative dependent variables such as multinomial regression.

To facilitate model interpretation, it is convenient to separate the class-specific utilities into log-linear intercept τ_k and slope λ_{kx},

$$\ln \omega_{kx} = \tau_k + \lambda_{kx}. \tag{2}$$

For identification, we use the "effects coding" (Vermunt and Magidson, 2013) sum-to-zero restrictions $\sum_k \tau_k = \sum_k \lambda_{kx} = \sum_x \lambda_{kx} = 0$. The attribute parameters λ_{kx} can then be interpreted as the deviation from option k's average log-utility in latent class x.

Primary interest focuses on how the probabilities of belonging to these latent (post)materialism classes differ as a function of the country-level covariates log-GDP per capita (Z_1) and percentage of women in parliament (Z_2). At the same time, it has become common practice to control for dependencies between units in the same country using a multilevel approach. We therefore model the probability that unit i in country c belongs to latent class x using the multilevel multinomial logistic regression

$$P(X_{ic} = x | Z_{1ic} = z_{1ic}, Z_{2ic} = z_2, G_c = g) = \frac{\exp(\alpha_x + \gamma_{1x} z_1 + \gamma_{2x} z_2 + \beta_{gx})}{\sum_t \exp(\alpha_t + \gamma_{1t} z_1 + \gamma_{2t} z_2 + + \beta_{tg})}, \tag{3}$$

where the country-level random effect variable G has been introduced. The random effects β_{tx} in Equation 3 could be assumed to have a Gaussian (normal) distribution,

leading to a standard multilevel model. However, in the past very differing classes of countries have been observed (Vermunt, 2003), so that the normality assumption is rather implausible. Since an incorrect parametric specification of the distribution of G can bias the results (e.g. Heckman and Singer, 1984, p. 275), we instead take G to be a country-level latent class variable with S classes and a freely estimated (nonparametric) distribution. Overall, then, our model can be seen as a multilevel multinomial regression of (post)materialism on country-level covariates, in which the nominal dependent variable is latent (corrected for misclassification), and the random effects distribution is nonparametric rather than Gaussian. The main parameters of substantive interest in Equation 3 are therefore the multinomial logistic regression coefficients γ_{mx}.

Combining the three ranking tasks, each modeled according to Equation 1, with the multilevel latent class regression model in Equation 3 leads to the full data likelihood

$$L(\boldsymbol{\theta}) = P(A_1, A_2, B_1, B_2, C_1, C_2 | Z_1, Z_2) =$$
$$\prod_{c=1}^{C} \sum_{G} P(G_c) \prod_{i=1}^{n_c} \sum_{X} P(X_{ic} | Z_{1ic}, Z_{2ic}, G_c) \times$$
$$P(A_{1ic}, A_{2ic} | X_{ic}) P(B_{1ic}, B_{2ic} | X_{ic}) P(C_{1ic}, C_{2ic} | X_{ic}), \quad (4)$$

where the realizations of variables have been dropped for clarity, and $\boldsymbol{\theta}$ is a vector collecting the free parameters of the model[4]. To obtain the likelihood in Equation 4, G is assumed to fully account for the dependency between units within countries, and X fully accounts for the dependency between units within a country, i.e. units are i.i.d. given X, and country likelihoods are independent given G.

Figure 1 shows a graphical representation of the model in Equation 4. The param-

---

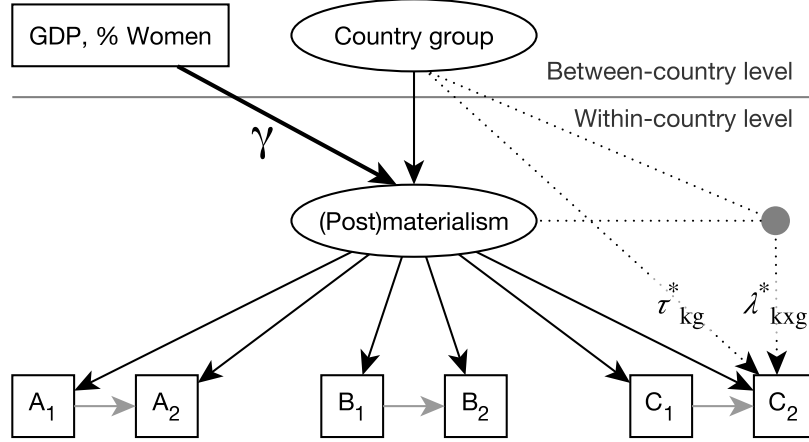[4]There are $T(11 + S) - 3$ free parameters overall.

Figure 1: Graphical representation of the multilevel latent class regression model for (post)materialism measured by three partial ranking tasks. Observed variables are shown in rectangles while unobserved ("latent") variables are shown in ellipses.

eters of primary substantive interest are the effect-coded logistic regression coefficients for the effect of GDP per capita and percentage of women in parliament on the latent postmaterialism class (the stronger line labeled γ in Figure 1). The top part of the model in Figure 1 is a multilevel latent class regression. The gray arrows between indicators within the same set are meant to show that a fixed conditional relationship is specified between the first and second choice within each set (Equation 1).

Equation 4 is, in fact, a "full measurement invariance" model: the measurement of (post)materialism by the three ranking tasks is not allowed to differ over countries. This excludes potential violations of "scalar" and "metric" measurement invariance, which can be conceptualized as, respectively, a direct main effect and a direct interaction effect of the grouping variable (Mellenbergh, 1989; Kankaraš, Moors and Vermunt, 2010; Kankaraš, Vermunt and Moors, 2011). Since we do not employ multiple groups (fixed effects) to deal with country effects, but a latent class multilevel (random effects) approach, instead of considering main and interaction effects from the countries directly, we consider these for the country random effects (G) (see De Jong, Steenkamp and Fox,

2007; Fox and Verhagen, 2011; Verhagen and Fox, 2013; Jak, Oort and Dolan, 2013, 2014). Thus, measurement invariance violations could be parameterized by extending Equation 2 to include country group class (g) effects:

$$\ln \omega_{kxg} = \tau_k + \lambda_{kx} + \tau_{kg}^* + \lambda_{kxg}^*. \tag{5}$$

The base invariance model above can then be seen as fixing the intercept deviations τ_{kg}^* and slope deviations λ_{kxg}^* to 0. An example of this parameterization of measurement non-invariance for the second ranking of Set C (C_2) is shown in Figure 1 as the dotted main effect and interaction effects for scalar and metric invariance, respectively.

With three sets of three non-redundant priorities each, there are $9(S-1)T$ possible additional τ_{kg}^* and λ_{kxg}^* parameters representing misspecification in the full invariance model specified above. For example, with $S = 3, T = 3$, there will be 54 possible violations of invariance. This is clearly a very large number of potential violations of measurement invariance, with an astronomical ($\approx 1.8 \times 10^{16}$) number of possible subsets of non-invariant models, making the fitting of each of these possible submodels infeasible. The EPC-interest, which considers the possible effect of freeing each of the restrictions separately, is therefore an attractive alternative.

However, there are strong dependencies between the additional parameter estimates. First, estimates of parameters corresponding to different categories of the same variable will necessarily be strongly dependent. Second, parameters corresponding to the same ranking set (A, B, or C) will be highly dependent due to the necessary dependence between the first and second choice in a ranking task. Therefore, instead of considering only one possible misspecification at a time, we consider freeing sets of restrictions corresponding to all main effects and all interaction effects of each set. The advantages are that the estimated change in the parameters of interest will be

closer to the observed change when freeing these restrictions, and that a model space to be explored of order 10^{16} is reduced to the examination of the effect of six sets of misspecifications on $2(T - 1)$ parameters of interest, γ_{jx}.

In summary, we have attacked the problem of examining the differences in (post)-materialism over 48 countries with different levels of GDP per capita and percentage of women in parliament by formulating a multilevel latent class model for ranking data. Measurement invariance is important here because direct main and interaction effects from country groups on the ranking tasks could threaten the comparison of countries with different levels of the covariates. Therefore it becomes relevant to examine the impact of these possible violations of measurement invariance on the parameters of substantive interest. The following section will introduce a method for doing so without the need to fit the alternative models.

4. EPC-INTEREST

In the application discussed, as in any application of invariance testing, interest focuses on a particular (function of) free parameters of the model. As argued by Oberski (2012, 2014), the question of invariance testing is then relevant insofar as it affects these parameters of interest. Such parameters may differ in models that allow for non-invariance, which in the model introduced in the previous section was parameterized as direct main and interaction effects of a group random effect. Instead of investigating whether such effects are statistically significantly different from zero, we suggest to investigate the impact of such effects on the parameters of interest. This is done using the "expected parameter change" in the parameter of interest, "EPC-interest".

The "EPC-interest" estimates the change in a free parameter of the model that

one can expect to observe if a particular restriction were freed. It is therefore a method of sensitivity analysis. However, the researcher is not forced to estimate all possible alternative models, but can evaluate the sensitivity of the results after fitting the restrictive full invariance model. The EPC-interest is based on the work of Saris, Satorra and Sörbom (1987), who introduced the expected parameter change in a fixed parameter for linear structural equation models (SEM), and Bentler and Chou (1993), who introduced the expected parameter change in a free parameter after freeing a fixed parameter for SEM. Oberski (2014) suggested using the EPC-interest for invariance testing using linear structural equation models with continuous data.

In this section we develop the EPC-interest in some generality so as to be useful for a larger set of models than just that applied in the example. In principle EPC-interest developed below applies to any situation in which parameters of interest are estimated by maximum likelihood and the researcher wishes to consider the likely effect on the parameters of interest that freeing some nuisance parameter might have. This is particularly useful for the case of latent variable modeling, in which the parameters of interest are typically the structural parameters (e.g. regression coefficients between latent variables) and the nuisance parameters are often measurement parameters. In the special case of invariance testing, the nuisance parameters can be conceived of as direct main and interaction effects of group dummy variables or random effects, or, equivalently, as differences between groups in measurement parameters.

We also extend the EPC-interest to evaluate the probable impact of several misspecifications at the same time. This last extension is particularly relevant in categorical data models, since parameters related to different categories of the same variable are often highly dependent. In such cases, the impact of misspecifications is better evaluated in sets of strongly related parameters. In the application to ranking data discussed above, for instance, parameters relating to the different ranking objects in the

same set will be highly dependent, since the second choice in each ranking is modeled conditionally upon the first choice (Equation 1). The impact of freeing the parameter λ_{111} can therefore not be seen separately from that of $\lambda_{211}, \lambda_{121}$, etc. We therefore consider the impact of all possible interactions of country grouping G with the latent (post)materialism variable X jointly for each of the three ranking tasks.

The key concept is considering the likelihood not only as a function of the free parameters of the model, but also as a function of the parameters that are fixed to obtain the full invariance model. Collecting the free parameters in a vector $\boldsymbol{\theta}$ and the fixed parameters in a vector $\boldsymbol{\psi}$, we assume the likelihood can be written as an explicit function of both sets of parameters, $L(\boldsymbol{\theta}, \boldsymbol{\psi})$. The maximum-likelihood estimates $\hat{\boldsymbol{\theta}}$ of the free parameters can then be seen as obtained under the full invariance model that sets $\boldsymbol{\psi} = 0$, i.e. $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \boldsymbol{\psi} = 0)$. Further, define the parameters of substantive interest as $\boldsymbol{\pi} := \mathbf{P}\boldsymbol{\theta}$, where $\mathbf{P}$ is typically a logical (0/1) selection matrix, although any linear function of the free parameters $\boldsymbol{\theta}$ may be taken. Interest then focuses on the likely value these free parameters $\boldsymbol{\pi}$ would take if the fixed $\boldsymbol{\psi}$ parameters were freed in an alternative model, $\hat{\boldsymbol{\pi}}_a = \mathbf{P} \arg\max_{\boldsymbol{\theta}, \boldsymbol{\psi}} L(\boldsymbol{\theta}, \boldsymbol{\psi})$.

We now show how these changes in the parameters of interest as a consequence of freeing the fixed parameters $\boldsymbol{\psi}$ can be estimated without fitting the alternative model. Let the Hessian $\hat{\mathbf{H}}_{\mathbf{ab}}$ be the matrix of second derivatives of the likelihood with respect to vectors $\mathbf{a}$ and $\mathbf{b}$, evaluated at the maximum likelihood solution of the full invariance model, $\hat{\mathbf{H}}_{\mathbf{ab}} := (\partial^2 L / \partial\mathbf{a}\partial\mathbf{b}')|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$. The expected change in the parameters of interest is then measured by the EPC-interest,

$$\text{EPC-interest} = \hat{\boldsymbol{\pi}}_a - \hat{\boldsymbol{\pi}} = \mathbf{P}\hat{\mathbf{H}}_{\boldsymbol{\theta\theta}}^{-1}\hat{\mathbf{H}}_{\boldsymbol{\theta\psi}}\mathbf{D}^{-1}\left[\frac{\partial L(\boldsymbol{\theta}, \boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right] + O(\boldsymbol{\delta}'\boldsymbol{\delta}), \qquad (6)$$

where $\mathbf{D} := \hat{\mathbf{H}}_{\boldsymbol{\psi\psi}} - \hat{\mathbf{H}}_{\boldsymbol{\theta\psi}}'\hat{\mathbf{H}}_{\boldsymbol{\theta\theta}}^{-1}\hat{\mathbf{H}}_{\boldsymbol{\theta\psi}}$ and the deviation from the true values is $\boldsymbol{\delta} := \boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}}$,

with $\boldsymbol{\vartheta}$ collecting the free and fixed parameters in a vector, $\boldsymbol{\vartheta} := (\boldsymbol{\theta}', \boldsymbol{\psi}')'$. Note that, apart from the order of approximation term $O(\boldsymbol{\delta}'\boldsymbol{\delta})$, Equation 6 contains only terms that can be calculated after fitting the invariance model. Thus, it is not necessary to fit the alternative model to obtain the EPC-interest.

In the structural equation modeling literature, the expected change in the fixed parameters $\boldsymbol{\psi}$ is commonly found and implemented in standard SEM software. This measure is commonly know as the "EPC", but to avoid confusion we term it "EPC-self" here. The EPC-self and EPC-interest both consider the impact of freeing restrictions, but differ in the target of this impact: the EPC-self evaluates the impact on the restriction itself, whereas the EPC-interest evaluates the impact on the parameters of interest. In spite of these differences, the two measures are related: this can be seen by recognizing that $-\mathbf{D}^{-1}\left[\left.\frac{\partial L(\boldsymbol{\theta},\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right] = \text{EPC-self} \approx \boldsymbol{\psi} - \hat{\boldsymbol{\psi}}$ so that, from Equation 6,

$$\text{EPC-interest} = -\mathbf{P}\hat{\mathbf{H}}_{\boldsymbol{\theta\theta}}^{-1}\hat{\mathbf{H}}_{\boldsymbol{\theta\psi}}\,\text{EPC-self} \approx -\mathbf{P}\hat{\mathbf{H}}_{\boldsymbol{\theta\theta}}^{-1}\hat{\mathbf{H}}_{\boldsymbol{\theta\psi}}\left(\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}\right) \tag{7}$$

Furthermore, since $\hat{\boldsymbol{\psi}}$ and $\hat{\boldsymbol{\theta}}$ are implicitly related by the fact that they are both solutions to the equation $\partial L/\partial\boldsymbol{\vartheta} = \mathbf{0}$, invoking the implicit function theorem yields $-\hat{\mathbf{H}}_{\boldsymbol{\theta\theta}}^{-1}\hat{\mathbf{H}}_{\boldsymbol{\theta\psi}} = \partial\boldsymbol{\theta}/\partial\boldsymbol{\psi}'$, so that

$$\text{EPC-interest} = \mathbf{P}\left(\frac{\partial\boldsymbol{\theta}}{\partial\boldsymbol{\psi}'}\right)\left(\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}\right), \tag{8}$$

that is, the EPC-interest can be seen simply as the coefficient of a linear approximation to the relationship between the free and fixed parameters, multiplied by the change in the fixed parameters. This demonstrates the difference with the sensitivity analysis approach common in econometrics (Magnus and Vasnev, 2007, p. 168) and applied to SEM by Yuan, Marshall and Bentler (2003), in which only $\partial\boldsymbol{\theta}/\partial\boldsymbol{\psi}'$ is considered: the

EPC-interest combines both the direction $(\partial\boldsymbol{\theta}/\partial\boldsymbol{\psi}')$ and the magnitude $(\boldsymbol{\psi} - \hat{\boldsymbol{\psi}})$ of the misspecification.

The derivation of the EPC-interest given in Equation 6 starts from the full invariance solution. We then find a hypothetical new maximum of the likelihood by setting the gradient of a Taylor expansion of the likelihood around the full invariance solution to zero:

$$\frac{\partial L(\boldsymbol{\theta}, \boldsymbol{\psi})}{\partial\boldsymbol{\vartheta}} = \mathbf{0} = \begin{pmatrix} \frac{\partial L(\boldsymbol{\theta},\boldsymbol{\psi})}{\partial\boldsymbol{\theta}}\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \\ \frac{\partial L(\boldsymbol{\theta},\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \end{pmatrix} + \begin{pmatrix} \hat{\mathbf{H}}_{\boldsymbol{\theta\theta}} & \hat{\mathbf{H}}_{\boldsymbol{\psi\theta}} \\ \hat{\mathbf{H}}_{\boldsymbol{\psi\theta}} & \hat{\mathbf{H}}_{\boldsymbol{\psi\psi}} \end{pmatrix} \begin{pmatrix} \boldsymbol{\theta} - \hat{\boldsymbol{\theta}} \\ \boldsymbol{\psi} - \hat{\boldsymbol{\psi}} \end{pmatrix} + O(\boldsymbol{\delta}'\boldsymbol{\delta}). \quad (9)$$

A similar device was used to derive the so-called "modification index" or "score test" for the significance of the hypothesis $\boldsymbol{\psi} = \mathbf{0}$ by Sörbom (1989, p. 373). Equation 6 follows directly by noting that $(\partial L(\boldsymbol{\theta}, \boldsymbol{\psi})/\partial\boldsymbol{\theta})|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \mathbf{0}$ and applying the standard linear algebra result on the inverse of a partitioned matrix $\left(\hat{\mathbf{H}}^{-1}\right)_{\boldsymbol{\theta\psi}} = -\hat{\mathbf{H}}_{\boldsymbol{\theta\theta}}^{-1}\hat{\mathbf{H}}_{\boldsymbol{\theta\psi}}\mathbf{D}^{-1}$ (e.g. Magnus and Neudecker, 2007, p. 12).

The accuracy of the approximation of the EPC-interest as a measure of the change in the parameters of interest is reflected in the order of approximation term, $O(\boldsymbol{\delta}'\boldsymbol{\delta})$. It can be seen that this accuracy is quadratic in the overall change in parameters, so that the approximation can be expected to work best when the misspecifications are not "too large". This result corresponds to results on the score test ("modification index") and "EPC-self" in the literature on structural equation modeling, which can be shown to be exact under a "sequence of local alternatives", i.e. when $\boldsymbol{\vartheta} = \lim_{n\to\infty}\hat{\boldsymbol{\vartheta}} + n^{-\frac{1}{2}}\boldsymbol{\delta}$ (Satorra, 1989, p. 135). It is important to note here that $\boldsymbol{\delta}$ is the deviation from the "true" value of $\boldsymbol{\vartheta}$, rather than the deviation from the limit of the parameter estimates under the alternative model. Therefore another view on the accuracy is that it will be better when the alternative model is not strongly misspecified. For this reason it is also

important to consider freeing sets of very strongly related parameters simultaneously, since a change in one of them will then imply a change in the others, and, consequently, a misspecified alternative model.

5. RESULTS

We now estimate the multilevel latent class model using 48 countries from the World Values Survey[5] for 2010–2012. Since our goal is to compare countries with different levels of GDP per capita and proportion of women in parliament, in this study we evaluate the sensitivity of these effect estimates of interest to measurement invariance assumptions using the EPC-interest.

The model as well as EPC-interest measures were implemented by one of the authors in the software Latent Gold Choice version 5.0.0.14157 and higher (Vermunt and Magidson, 2005, 2013, pp. 135–136). Program input for this analysis can be found in Appendix B, while output and data are included in the online Appendix at `http://[BLINDED]`.

We select three classes for both the latent "(post)materialism" variable and the country group class variable. For BIC values and the rationale behind these choices, please see Appendix C. Since class selection is not the focus of this example analysis, we will not discuss it here further.

The full invariance model imposes $\tau^*_{jkg} = 0$ and $\lambda^*_{jkxg} = 0$, i.e. there is no difference in measurement over the group random effect variable. Such violations of measurement invariance are relevant in the present analysis because they could potentially threaten the conclusions of substantive interest, namely the relationships γ between (post)materialism on the one hand and GDP per capita and percentage of women in

---

[5]`http://www.worldvaluessurvey.org`

parliament on the other. In this particular analysis, it was thought to be pertinent whether, after freeing some of the measurement invariance restrictions, one of these regression coefficients of interest could be estimated with a reversed sign.

We therefore calculated the EPC-interest for these parameters, of which our three-class model has four: two for each for each of the two independent variables. Measurement invariance violations can potentially take the form of 6 direct main effects (τ_{jkg}^*) –$(4-1)$ category effects in $(3-1)$ classes– and 12 direct interaction effects (λ_{jkxg}^*) for each of the three ranking tasks, totaling 54 possible misspecifications in the full invariance model. However, these misspecifications are strongly correlated and should not be considered separately. Rather, we consider the probable impact on the γ parameters of interest of freeing the direct main effects for each ranking task separately and of freeing the direct interactions for each ranking task separately. In short, rather than consider the direct and interaction effects for each of the 48 countries on each of the 3 unique categories of each of the three ranking tasks, making for $4 \times 3 \times (4-1) \times (48-1) \times (1+(3-1)) = 5076$ potential EPC-interest values, we evaluate direct effects of the country group random effect and consider their impact jointly for strongly correlated misspecifications, reducing the problem to 24 EPC-interest values of interest.

Table 2 shows these 34 EPC-interest values calculated from the full invariance model. These EPC-interest values estimate the likely change from the maximum likelihood estimates (column 3 in Table 2) in the four parameters of substantive interest after freeing the direct main or interaction effects for each of the three ranking tasks (columns 5–10 in Table 2). In the full invariance model, Class 1 corresponds to a "postmaterialism" class. The estimate -0.035 (s.e. 0.007) shown in Table 2) would therefore suggest that more prosperous nations tend to be less postmaterialist. This directly contradicts the theory of Inglehart (1997), which suggests that this coefficient should

Table 2: Full invariance multilevel latent class model: parameter estimates of interest with standard errors (columns 3 and 4), as well as expected change in these parameters measured by the EPC-interest when freeing each of six sets of possible misspecifications (columns 5–10).

				EPC-interest for...					
				τ^*_{jkg}			λ^*_{jkxg}		
		Estimates		Ranking task			Ranking task		
		Est.	s.e.	1	2	3	1	2	3
Class 1	GDP	-0.035	(0.007)	-0.013	0.021	-0.002	**0.073**	**0.252**	0.005
Class 2	GDP	-0.198	(0.012)	-0.018	-0.035	0.015	-0.163	-0.058	0.002
Class 1	Women	0.013	(0.001)	-0.006	0.002	0.000	-0.003	0.029	0.002
Class 2	Women	-0.037	(0.001)	0.007	-0.003	0.002	-0.006	-0.013	0.002

be positive.

Since the theory specifies only that certain coefficients should be positive or negative, the key focus of substantive interest should be whether misspecifications in the invariance model can potentially change the sign of a parameter of interest. In Table 2, we therefore look for EPC-interest values that, when added to the estimates in column 3, would change the sign of those estimates. It can be seen in the Table that two such EPC-interest are indeed present, namely the direct interaction effect of the country group class with the postmaterialism class on ranking tasks 1 and 2. The analogous finding (Kankaraš, Vermunt and Moors, 2011) if CFA had been used would be that of "metric equivalence" for ranking task 3 but not ranking tasks 1 and 2. This means that the attribute parameters that define the classes for these two tasks differ over country groups, and that after accounting for these differences the effect of GDP on postmaterialism is estimated to be positive rather than negative. This set of misspecifications is thus of substantive interest and should be amended in the model.

We therefore free these two sets of measurement invariance violations, allowing for differences in the parameters of ranking tasks 1 and 2 across country groups. Table 3 shows the substantive parameter estimates and EPC-interest interest values for

Table 3: Partially invariant multilevel latent class model: parameter estimates of interest with standard errors (columns 3 and 4), as well as expected change in these parameters measured by the EPC-interest when freeing each of four sets of remaining possible misspecifications (columns 5–7 and 10).

				EPC-interest for non-invariance of...					
				τ^*_{kg}			λ^*_{kxg}		
				Ranking task			Ranking task		
		Est.	s.e.	1	2	3	1	2	3
Class 1	GDP	-0.127	(0.008)	-0.015	-0.003	0.002			0.097
Class 2	GDP	0.057	(0.011)	-0.043	-0.013	0.002			0.161
Class 1	Women	0.008	(0.001)	-0.002	0.000	0.002			0.001
Class 2	Women	0.020	(0.001)	-0.007	-0.001	0.002			0.007

the resulting partial invariance model, which has 63 parameters and a log-likelihood of -418610.0 (BIC $= 837920.5$). The likelihood ratio test of improvement in model fit is highly significant ($\chi^2_{\text{df}=28} = 24607$). Moreover, the substantive logistic regression coefficient for the effect of GDP on the postmaterialism class (Class 2 in Table 3), is indeed positive after freeing the detected misspecifications. Re-calculating EPC-interest values for the remaining possible misspecifications reveals that none of the possible misspecifications in this partial invariance model has the potential to change the substantive conclusions. In this sense, we therefore conclude that the partial invariance model fits "approximately", since none of the substantive conclusions based on it are threatened by measurement invariance violations.

Table 4 shows the sizes of the three (post)materialism classes (third row) as well as the "attribute parameters", i.e. each class's average log-utility. Thus, when reading each row horizontally, the class with the highest log-utility represents respondents who value that object highest. For example, priorities A.1, A.2, B.1, B.3, and C1 have the highest log-utilities in Class 1. Since all of these priorities are "materialist" (labeled "M" in Table 4), we also labeled Class 1 "materialist". A caveat with this label is that the materialist priorities that are most strongly related to this class also happen

Table 4: Estimated log-utilities under the final model. In each row, the highest log-utility has been printed in **bold face** to facilitate interpretation of the classes.

		Class 1	Class 2	Class 3
	Class label	"Materialist"	"Postmater."	"Mixed"
	Class size	0.569	0.213	0.218
	(s.e.)	(0.0114)	(0.0179)	(0.0280)
Set A				
M	1. Economic growth	**2.1102**	0.4837	0.4156
M	2. Strong defense	**-0.5285**	-1.4984	-0.9249
P	3. More say	-0.5519	**1.4683**	0.4643
P	4. More beauty	-1.0298	-0.4536	**0.0449**
Set B				
M	1. Order in the nation	**1.0016**	-0.5898	0.0435
P	2. More say	-0.4592	**0.6902**	-0.2763
M	3. Rising prices	**0.4281**	-0.2269	0.3719
P	4. Freedom of speech	-0.9705	**0.1266**	-0.1390
Set C				
M	1. Stable economy	**2.0086**	0.0789	0.1715
P	2. Humane society	-0.7919	**0.4450**	-0.0943
P	3. Ideas	-1.1402	**-0.0593**	-0.4550
M	4. Fight crime	-0.0765	-0.4646	**0.3778**

to be the first item in each set, so that a primacy effect may play a role here as well. Class 2 is labeled "postmaterialist" because it has the highest log-utilities for all of the postmaterialist priorities (labeled "P" in Table 4), with the exception of A.4. Preferences in the third class appear to be for the most part in-between those of Classes 1 and 2. At the same time, however, this class has the highest log-utilities for A.4 (a "postmaterialist" object) and C.4 (a "materialist" object). For this reason we apply the label "mixed" to Class 3.

This combination of classes corresponds to the labels and groups created by Inglehart (1997). However, a key difference with the predefined groups of Inglehart (1997) and the latent classes in the current model is that the latent classes here arise purely from the model and the response patterns observed in the data, rather than being imposed a priori. Crucially also, our model allows for measurement error in the ranking
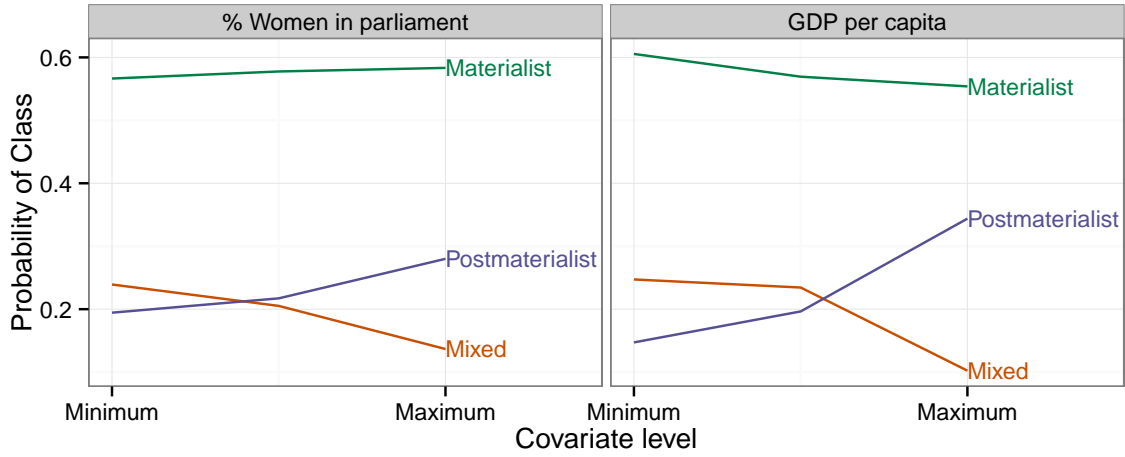
Figure 2: Estimated probability of choosing each class as a function of the covariates of interest under the final model.

tasks.

Finally, Figure 2 summarizes the loglinear parameter estimates of interest shown column 3 of Table 3. These graphs plot the covariate level on the horizontal axes and the probability of belonging to each of the classes defined in Table 4 on the vertical axes. The left-hand plot in Figure 2 shows the estimated marginal relationship between the percentage of women in parliament and the (post)materialism classes. It can be seen that the probability of the "post materialist" class is much larger for countries with a large percentage of women in parliament than it is for countries for fewer female parliamentarians: the predicted probability of belonging to the postmaterialist class for countries with the observed maximum percentage of women in parliament is 0.30, which is 7 percentage points above the average "postmaterialist" class membership of 0.23 (shown in Table 4). The probability of belonging to the purely economical "materialist" class also increases very slightly, to the overall detriment of membership of the "mixed" class in which "order in the nation" and "fighting crime" is sometimes endorsed.

22

The right-hand side of Figure 2 plots the estimated effect of log-GDP per capita on the three class membership probabilities. A similar pattern to that for percentage of women in parliament is observed for the "postmaterialist" and "mixed" classes, while the probability of belonging to the "materialist" class decreases somewhat with GDP per capita. The differences in predicted class membership are even larger over the observed GDP range than for the percentage of women in parliament: the predicted postmaterialist class membership at the maximum observed GDP is 0.40, which is 17 percentage points above the average. Greater prosperity is thus clearly strongly associated with postmaterialism. Crucially, the application of EPC-interest leads to greater confidence that this conclusion is not an artefact of differences in measurement.

6. DISCUSSION AND CONCLUSION

Whenever groups are compared, measurement equivalence is a concern. Particularly, it should be verified that substantive conclusions of interest are uncontaminated by possible cross-group differences in measurement. The EPC-interest, a measure introduced by Oberski (2014) for this purpose in the context of linear structural equation models, was extended in this paper to apply to categorical observed and latent variables as well as rankings and other types of data often encountered in the social sciences. An example application looked at the relationship between latent (post)materialism on the one hand and, on the other, log-GDP per capita and the percentage of women in parliament using data from 67,568 respondents in 48 countries. The EPC-interest is a useful guide for this purpose, reducing a potentially 5076-dimensional problem to the examination of Tables 2 and 3. Finally, we provided user-friendly software implementing the EPC-interest; program input and data for the example application is available in the online Appendix.

The EPC-interest fits within the recent literature on "approximate measurement invariance". That is, the goal of incorporating important measurement differences in the model while ignoring unimportant ones. Invariance testing using "approximate" fit measures such as the CFI and RMSEA (e.g. Cheung and Rensvold, 2002; Chen, 2007) can be thought of as such a method, since the fit of the model is thought to be "approximate", and relative to model complexity. Model-based approaches include random effects modeling (De Jong, Steenkamp and Fox, 2007; Verhagen and Fox, 2013; Jak, Oort and Dolan, 2013, 2014), Bayesian priors on measurement differences (Muthén and Asparouhov, 2012), and factor rotation towards an invariant solution (Muthén and Asparouhov, 2014). An advantage of these last three approaches is that a middle ground may be found between exact equality of measurement processes over groups and complete independence (Van De Schoot et al., 2013). However, a current problem is that there is no agreed-upon definition of "approximate". At the same time, the EPC-interest approach does not offer a middle ground, but it does solve the problem of defining what "approximate" fit might be (Van De Schoot et al., 2013, p. 12): for example, in the current application it was taken to be any measurement difference that cannot reverse the sign of the regression parameters of substantive interest. An interesting direction for future work is therefore to combine the sensitivity analysis approach of the EPC-interest with the approaches to approximate measurement invariance of Muthén and Asparouhov (2012, 2014), for example by weighting the cross-group measurement differences to be aligned by the possible change in parameters of interest.

The scope of the paper was limited to an application of the EPC-interest to a model with categorical observed and latent variables. While we consider this a useful contribution, we did not further investigate the performance of this measure. We do note that the actually observed change in the parameters of interest after freeing the suggested invariance restrictions was very close to that predicted by the EPC-interest

in this particular application. The Monte Carlo simulation using linear structural equation models in Oberski (2014) demonstrates that this is the case for linear structural equation models with relatively simple setups. However, an investigation of whether this is the case for categorical latent class models under a wider range of hypothetical conditions in a simulation study remains work for the future.

REFERENCES

Arel-Bundock, Vincent. 2013. *WDI: World Development Indicators (World Bank).* R package version 2.4.
   URL: *http://CRAN.R-project.org/package=WDI*

Bentler, P.M. and C.P. Chou. 1993. "Some new covariance structure model improvement statistics." In *Testing structural equation models*, ed. K.G. Jöreskog and J. Scott Long. Sage focus editions Sage Publications pp. 235–235.

Böckenholt, U. 2002. "Comparison and choice: Analyzing discrete preference data by latent class scaling models." In *Applied latent class analysis*, ed. Jacques A.P. Hagenaars and Allan L. McCutcheon. cambridge, UK: Cambridge University Press pp. 163–182.

Chen, F.F. 2007. "Sensitivity of goodness of fit indexes to lack of measurement invariance." *Structural Equation Modeling* 14:464–504.

Cheung, G.W. and R.B. Rensvold. 2002. "Evaluating goodness-of-fit indexes for testing measurement invariance." *Structural Equation Modeling* 9:233–255.

Croon, Marcel. 1989. "Latent Class Models for the Analysis of Rankings." In *New*

Developments in Psychological Choice Modelling, ed. G. De Soete, H. Feger and K. C. Klauer. Elsevier Science Publishers pp. 99–121.

De Jong, M.G., J.B.E.M. Steenkamp and J.P. Fox. 2007. "Relaxing Measurement Invariance in Cross-National Consumer Research Using a Hierarchical IRT Model." *Journal of consumer research* 34:260–278.

Fox, J.P. and A.J. Verhagen. 2011. "Random item effects modeling for cross-national survey data." In *Cross-cultural analysis: Methods and applications*, ed. Eldad Davidov, Peter Schmidt and Jaak Billiet. New York: Routledge pp. 467–488.

Francis, Brian, Regina Dittrich, Reinhold Hatzinger and Roger Penn. 2002. "Analysing partial ranks by using smoothed paired comparison methods: an investigation of value orientation in Europe." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 51:319–336.

Hagenaars, Jacques A P. 1990. *Categorical longitudinal data: Log-linear panel, trend, and cohort analysis.* Newbury Park: Sage.

Heckman, J. and B. Singer. 1984. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica: Journal of the Econometric Society* 52:271–320.

Inglehart, Ronald. 1977. *The silent revolution.* Princeton: Princeton University Press.

Inglehart, Ronald. 1981. "Post-materialism in an environment of insecurity." *The American Political Science Review* 75:880–900.

Inglehart, Ronald. 1997. *Modernization and postmodernization: Cultural, economic, and political change in 43 societies.* Cambridge, UK: Cambridge University Press.

Inglehart, Ronald and Christian Welzel. 2010. "Changing mass priorities: The link between modernization and democracy." *Perspectives on Politics* 8:551–567.

Inglehart, Ronald, Pippa Norris and Christian Welzel. 2002. "Gender equality and democracy." *Comparative Sociology* 1:321–345.

Jackson, David J and Duane F Alwin. 1980. "The factor analysis of ipsative measures." *Sociological Methods & Research* 9:218–238.

Jak, Suzanne, Frans J Oort and Conor V Dolan. 2013. "A test for cluster bias: Detecting violations of measurement invariance across clusters in multilevel data." *Structural Equation Modeling: A Multidisciplinary Journal* 20:265–282.

Jak, Suzanne, Frans J Oort and Conor V Dolan. 2014. "Measurement bias in multilevel data." *Structural Equation Modeling: A Multidisciplinary Journal* 21:31–39.

Kamakura, Wagner A, Michel Wedel and Jagadish Agrawal. 1994. "Concomitant variable latent class models for conjoint analysis." *International Journal of Research in Marketing* 11:451–464.

Kankaraš, Miloš, Guy Moors and Jeroen K Vermunt. 2010. "Testing for measurement invariance with latent class analysis." In *Cross-cultural analysis: Methods and applications*, ed. Eldad Davidov, Peter Schmidt and Jaak Billiet. New York: Taylor & Francis pp. 359–384.

Kankaraš, Miloš, Jeroen K Vermunt and Guy Moors. 2011. "Measurement equivalence of ordinal items: A comparison of factor analytic, item response theory, and latent class approaches." *Sociological Methods & Research* DOI: 10.1177/0049124111405301.

Karvonen, Lauri. 2004. "Preferential voting: incidence and effects." *International Political Science Review* 25:203–226.

Luce, R. Duncan. 1959. *Individual Choice Behavior: a Theoretical Analysis*. New York: John Wiley and Sons.

Lukočienė, Olga, Roberta Varriale and Jeroen K Vermunt. 2010. "The simultaneous decision(s) about the number of lower-and higher-level classes in multilevel latent class analysis." *Sociological Methodology* 40:247–283.

Magnus, J.R. and A.L. Vasnev. 2007. "Local sensitivity and diagnostic tests." *The Econometrics Journal* 10:166–192.

Magnus, J.R. and H. Neudecker. 2007. *Matrix Differential Calculus with Applications in Statistics and Econometrics, Third Edition*. New York: John Wiley & Sons.

Maydeu-Olivares, Albert and Ulf Böckenholt. 2005. "Structural equation modeling of paired-comparison and ranking data." *Psychological methods* 10:285.

McFadden, Daniel. 1974. "Conditional logit analysis of qualitative choice behavior." In *Frontiers in econometrics*, ed. Paul Zarembka. New York: Academic Press pp. 105–142.

Mellenbergh, G.J. 1989. "Item bias and item response theory." *International Journal of Educational Research* 13:127–143.

Moors, Guy and Jeroen Vermunt. 2007. "Heterogeneity in post-materialist value priorities. Evidence from a latent class discrete choice approach." *European Sociological Review* 23:631–648.

Muthén, Bengt and Tihomir Asparouhov. 2012. "Bayesian structural equation modeling: a more flexible representation of substantive theory." *Psychological methods* 17:313.

Muthén, Bengt and Tihomir Asparouhov. 2014. "IRT Studies of Many Groups: The Alignment Method." *Mplus Webnote* Version 2014-04-29.
URL: *http://www.statmodel.com/download/IRTAlignment.pdf*

Oberski, D.L. 2012. "Comparability of Survey Measurements." In *Handbook of Survey Methodology for the Social Sciences*, ed. Lior Gideon. New York: Springer-Verlag pp. 477–498.

Oberski, D.L. 2014. "Evaluating Sensitivity of Parameters of Interest to Measurement Invariance in Latent Variable Models." *Political Analysis* 22:45–60.

Oberski, D.L. and J.K. Vermunt. 2013. "A Model-Based Approach to Goodness-of-Fit Evaluation in Item Response Theory." *Measurement: Interdisciplinary Research & Perspectives* 11:117–122.
URL: *http://daob.nl/wp-content/uploads/2013/07/Oberski-Vermunt-A-model-based-approach-to-goodness-of-fit-evaluation-in-item-response-theory-2013-07-28.pdf*

R Core Team. 2012. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.
URL: *http://www.R-project.org/*

Saris, W.E., A. Satorra and D. Sörbom. 1987. "The Detection and Correction of Specification Errors in Structural Equation Models." *Sociological Methodology* 17:105–129.

Satorra, A. 1989. "Alternative Test Criteria in Covariance Structure Analysis: A Unified Approach." *Psychometrika* 54:131–151.

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

Sörbom, D. 1989. "Model modification." *Psychometrika* 54:371–384.

Steenkamp, JBEM and H. Baumgartner. 1998. "Assessing measurement invariance in cross-national consumer research." *Journal of Consumer Research* 25:78–107.

Thurstone, Louis L. 1927. "A law of comparative judgment." *Psychological review* 34:273.

Toplak, Jurij. 2010. Preferential voting: definition and classification. In *Annual Meeting of the Midwest Political Science Assication 67th Annual National Conference.*

Van De Schoot, Rens, Anouck Kluytmans, Lars Tummers, Peter Lugtig, Joop Hox and Bengt Muthén. 2013. "Facing off with Scylla and Charybdis: a comparison of scalar, partial, and the novel possibility of approximate measurement invariance." *Frontiers in Psychology* 4:770.

Vandenberg, R.J. and C.E. Lance. 2000. "A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research." *Organizational Research Methods* 3:4–70.

Verhagen, AJ and JP Fox. 2013. "Bayesian tests of measurement invariance." *British Journal of Mathematical and Statistical Psychology* 66:383–401.

Vermunt, J. K and J. Magidson. 2013. *Technical guide for Latent GOLD 5.0: Basic and advanced.* Belmont, MA: Statistical Innovations Inc.

Vermunt, Jeroen K. 2003. "Multilevel latent class models." *Sociological methodology* 33:213–239.

Vermunt, Jeroen K. and Jay Magidson. 2005. *Technical Guide for Latent GOLD Choice 4.0: Basic and Advanced.* Belmont, MA: Statistical Innovations.

Yuan, K.H., L.L. Marshall and P.M. Bentler. 2003. "Assessing the effect of model misspecifications on parameter estimates in structural equation models." *Sociological Methodology* 33:241–265.

A. COUNTRIES INCLUDED IN THE STUDY

ISO3 code	Country name	ISO3 code	Country name
DZA	Algeria	MAR	Morocco
ARM	Armenia	NLD	Netherlands, The
AUS	Australia	NZL	New Zealand
AZE	Azerbaijan	NGA	Nigeria
BLR	Belarus	PAK	Pakistan
CHL	Chile	PER	Peru
CHN	China	PHL	Philippines
COL	Colombia	POL	Poland
CYP	Cyprus	QAT	Qatar
ECU	Ecuador	RUS	Russian Federation
EGY	Egypt	RWA	Rwanda
EST	Estonia	SGP	Singapore
DEU	Germany	SVN	Slovenia
GHA	Ghana	ESP	Spain
IRQ	Iraq	SWE	Sweden
JPN	Japan	TTO	Trinidad and Tobago
JOR	Jordan	TUN	Tunisia
KAZ	Kazakhstan	TUR	Turkey
KOR	Korea, Republic of	UKR	Ukraine
KGZ	Kyrgyzstan	USA	United States
LBN	Lebanon	URY	Uruguay
LBY	Libya	UZB	Uzbekista
MYS	Malaysia	YEM	Yemen
MEX	Mexico	ZWE	Zimbabwe

B. LATENT GOLD CHOICE INPUT FOR THE FULL INVARIANCE MODEL

The input below fits the full invariance model described in the paper, setting the possible violations of invariance to zero (0). The option "score test" in the output section (only available in Latent Gold or Latent Gold Choice ≥ 5) is then used to obtain the EPC-interest values. Output and data for this example can be obtained from the online appendix at `http://`.

```
options

   maxthreads=all;
   algorithm
      tolerance=1e-008 emtolerance=0.01
      emiterations=450 nriterations=70 ;
   startvalues
      seed=0 sets=30 tolerance=1e-005 iterations=50;
   bayes
      categorical=0 variances=0 latent=0 poisson=0;
   missing  excludeall;

   // NOTE: The option "scoretest" for output is used to obtain
   //        the EPC-interest. This will also produce the score test ("MI")
   //        and EPC-self for the measurement invariance restriction

   output
      parameters=effect  betaopts=wl standarderrors profile
      probmeans=posterior
      frequencies bivariateresiduals estimatedvalues=regression
      predictionstatistics setprofile setprobmeans
      iterationdetails scoretest ;

    // There are several ways of modeling ranking data using LG or LGChoice.
    //    The most computationally efficient is to use the so-called "3-file"
    //    setup in LGChoice employed here (see LGChoice manual).

choice = 3
   alternatives 'inglehart_wvs6_long.alt' quote = single
   id=alt
   choicesets 'inglehart_wvs6_long.set' quote = single
```

```
    id=set;

variables
    groupid country;
    caseid id;
    choicesetid set ;
    dependent value ranking;
    independent NY_GDP_PCAP_CD, SG_GEN_PARL_ZS;
    attribute int1 nominal, int2 nominal, int3 nominal;
    latent
        GClass  group nominal 3,
        Class nominal 3;

equations
    // Group class intercept
    GClass <- 1 ;

    // Parameters of interest are logistic regression coefficients of
    //        NY_GDP_PCAP_CD and SG_GEN_PARL_ZS.
    Class <- 1 + GClass + NY_GDP_PCAP_CD + SG_GEN_PARL_ZS;

    // Below, sets of possible violations of measurement invariance have been
    //        explicitly restricted to equal zero using "(0)". This will
    //        produce EPC-interest, EPC-self, and Score test output.
    value <- int1 + int2 + int3 +
        int1 Class + int2 Class + int3 Class +
        (0) int1 GClass + (0) int2 GClass + (0) int3 GClass +
        (0) int1 Class  GClass +
        (0) int2 Class GClass +
        (0) int3 Class GClass ;
```

Table 5: Log-likelihood, number of parameters, and Bayesian Information Criterion (BIC) for models with different numbers of classes for the (post)materialism (within-country) and country group (between-country) latent class variables.

(Post)materialism (X) classes, $|\{G\}| = 1$				Country group (G) classes, $|\{X\}| = 3$			
#Classes	Log-lik	#Par	BIC(L^2)	#Classes	Log-lik	#Par	BIC
1	-460512.7	9	-10346.1	1	-447646.2	29	895616.3
2	-449836.9	19	-31586.2	2	-444754.8	32	889867.0
3	-447646.2	29	-35855.8	3	-443216.6	35	886824.1
4	-446211.1	39	-38614.4	4	-442734.2	38	885892.8
5	-445246.3	49	-40432.4	5	-442436.5	41	885330.9
6	-444776.4	59	-41260.3	6	-442110.5	44	884712.3
7	-444384.0	69	-41933.4	7	-441946.2	47	884417.3

C. MODEL SELECTION FOR THE EXAMPLE APPLICATION

In choosing the number of classes for the (post)materialism (within-country) and country group (between-country) latent class variables, we follow the advice of Lukočienė, Varriale and Vermunt (2010) to first fix the number of country-group classes to unity and choose a number of within-country classes, subsequently fixing the number of within-country classes to this chosen number and determining the number of country-group (between-country) classes. The left-hand side of Table 5 shows the log-likelihoods, number of parameters and Bayesian Information Criterion (BIC) values based on the L^2 for the model with one country-group class and an increasing number of (post)materialism classes. It can be seen that the BIC, which penalizes model complexity, decreases with each additional latent (post)materialism class. In fact, the BIC does not stop decreasing even when incrementing the number of classes to 14 (not shown in Table 5 for brevity).

In the literature on (post)materialism (e.g. Inglehart, 1997; Inglehart, Norris and Welzel, 2002; Inglehart and Welzel, 2010), the number of (post)materialism classes is typically fixed to three: "postmaterialist", "materialist", and "mixed". Clearly, using

the WVS ranking tasks and imposing full invariance, many more qualitative (post)materialism classes can be distinguished than the traditional three classes. This corresponds to findings by Moors and Vermunt (2007); however, these authors (following Hagenaars, 1990) also argued that "one can safely interpret the results (...) if adding another class does not result in important changes of the latent class weights for the other classes" (p. 637). While this is a somewhat subjective criterion, the three-class solution found in the data does correspond to the theoretical "postmaterialist", "materialist", and "mixed" classes, whose parameters appear to change little in the models with a greater number of classes. Moreover, the greatest reduction in BIC seen in Table 5 takes place when moving from a one-class to a two-class model, with relatively small improvements after three or more classes. We therefore follow the theoretical literature in selecting the three-class model.

While selecting the number of country-group classes, we find that the BIC improves little after three classes (right-hand side of Table 5), so that our initial full invariance model has three (post)materialism (within-country) and three country group (between-country) classes.