Survey Research Methods (2018)
Vol. 12, No. 1, pp. 29-41
doi:10.18148/srm/2018.v12i1.7210

ISSN 1864-3361
http://www.surveymethods.org

Approximate measurement invariance and longitudinal confirmatory factor analysis: concept and application with panel data

Daniel Seddig
Institute of Sociology and Social Psychology
University of Cologne, Germany
and
Department of Psychology
University of Zurich, Switzerland

Heinz Leitgöb
Institute of Sociology
University of Eichstätt-Ingolstadt
Germany

This article addresses the approximate approach to assess measurement invariance with (longitudinal) confirmatory factor analysis. Approximate measurement invariance uses zero-mean, small-variance Bayesian priors to allow minor differences in estimated parameters across time, while still maintaining comparability of the underlying constructs. The procedure is illustrated for the first time with panel data on young peoples' preferences to maximize pleasure and enjoy life. Results indicate whereas the traditional approach of exact measurement invariance failed to establish scalar invariance across time and precluded comparisons of latent means, it was possible to establish approximate scalar invariance. Based on a monitoring procedure for model fit and convergence, a rather small prior variance was deemed sufficient to account for minor deviations of cross-time intercept differences from zero.

Keywords: Confirmatory factor analysis, Bayesian structural equation modeling, approximate measurement invariance, panel data

1 Introduction

Comparative research is a core subject in different fields of social science. Research interests can be cross-national and cross-cultural or longitudinal and developmental. For example, it may be of interest to compare mean scores of variables, such as ethics, values, attitudes, and social or political trust across cultures, countries, or across time (Davidov, Schmidt, & Billiet, 2011). Regardless of whether the comparison is focused on different contexts or time points, the mere application of the same scale across groups or time does not ensure that the same concept is actually analyzed. Response characteristics to survey questions may be different across groups or over time which implies that people from different contexts may not understand the concept in the same way or that the understanding has changed. Thus, if the studied variables are latent variables measured by multiple indicators, a comparison across groups or time requires measurement invariance (hereafter: MI) to ensure that the concept remains comparable across groups or time. However, methodology reports of relevant large-scale panel studies, such as the Socioeconomic Panel (SOEP) in Germany, British Household Panel Survey (BHPS) in Great Britain, National Crime Victimization Survey (NCVS) in the United States, and European Union Statistics on Income and Living Conditions (EU-SILC) usually do not include systematic assessments of MI for key concepts across time. It seems that testing for MI usually remains within the area of responsibility of the individual data user. For this reason, the article aims at capacitating researchers to apply the innovative concept of Bayesian approximate MI for situations where conventional approaches turn out to be inappropriate.

Since traditional tests of MI build on very restrictive assumptions about the equivalence of measurement properties across groups or time, it cannot always be achieved to a satisfactory degree.[1] With the introduction of Bayesian analytic properties into confirmatory factor analysis (hereafter: CFA) and structural equation modeling (hereafter: SEM), researchers are able to relax exact equality constraints by assuming that parameters are only approximately equal, but comparability of the underlying constructs is maintained. For this purpose, cross-group or cross-time parameter differences are expressed as Bayesian priors which are assumed to be shaped as a normal distribution with a mean of zero and a small variance (B. Muthén & Asparouhov, 2013; Van de

Contact information: Daniel Seddig, University of Cologne, Institute of Sociology and Social Psychology, Albertus-Magnus-Platz, 50923 Köln, Germany (email: dseddig@uni-koeln.de)

[1]See Van de Schoot, Schmidt, De Beuckelaer, Lek, and Zondervan-Zwijnenburg (2015) for a recent overview of recent developments in the analysis of MI.

Schoot et al., 2013).[2]

In the following we will emphasize the properties of Bayesian approximate MI as an alternative to exact MI and demonstrate how it can be assessed within longitudinal CFA. Further, we will demonstrate the use of Bayesian approximate MI in an empirical application with panel data. So far, to the best of our knowledge, Bayesian approximate MI has not been applied to longitudinal data. The variables of interest are young peoples' preferences to maximize pleasure and enjoy life which are conceptualized as a latent variable called "hedonism". We will conclude with a summary of our findings and will discuss the implications for future comparative research.

2 Measurement Invariance

The indication of MI is "whether or not, under different conditions of observing and studying phenomena, measurement operations yield measures of the same attribute" (Horn & McArdle, 1992, p. 117). If MI does not hold, assuming differences in latent variable means or regression coefficients can be incorrect or actual differences may be concealed. For longitudinal modeling, MI guarantees that the same latent variable is measured at all times and avoids interpreting change in the meaning or understanding of a latent variable as change in terms of an underlying developmental process (Ferrer, Balluerka, & Widaman, 2008; Little, 2013; Stoel, van den Wittenboer, & Hox, 2004; Widaman, Ferrer, & Conger, 2010).

To illustrate the parameters that can be subject to MI constraints, recall a confirmatory factor model (Bollen, 1989; Brown, 2015; Jöreskog, 1979; Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000):

$$y^t = \tau^t + \Lambda^t \eta^t + \epsilon^t \tag{1}$$

where y^t is a $(M \times 1)$ vector of M observed indicators (for time t consecutively), η^t is a $(N \times 1)$ vector of N latent unobserved variables, τ^t is a $(M \times 1)$ vector of indicator intercepts, Λ^t is a $(M \times N)$ matrix of factor loadings linking observed indicators and unobserved latent variables, and ϵ^t is a $(M \times 1)$ vector of measurement errors with $E(\epsilon^t) = 0$; $\mathrm{COV}(\eta^t, \epsilon^t) = 0$.[3]

MI is typically assessed by imposing different equality restrictions on the parameters across time (Hertzog & Nesselroade, 2003; Horn & McArdle, 1992; Meredith, 1993; Meredith & Teresi, 2006; Reise, Widaman, & Pugh, 1993; Vandenberg & Lance, 2000). Configural MI assumes structural equivalence. This means that only the patterns of factors and factor loadings are constrained to be equal across time. Metric MI requires the factor loadings to be equal across time (factor loading invariance), indicating that the meaning of a latent variable is the same across time. Metric MI is considered necessary to compare factor covariances or unstandardized regression coefficients. Scalar MI additionally requires indicator intercepts to be equal across time (intercept invariance), indicating the similar use or origin of a scale across time. Scalar MI is considered necessary to compare latent means.[4]

3 Approximate Measurement Invariance

Factor loadings and intercepts are usually constrained to be exactly equal across time. When exact MI does not hold, one can reduce the of number time points that are compared. However, this maybe be rather unsatisfactory. Another alternative is partial MI (Byrne, Shavelson, & Muthén, 1989) where non-invariant parameters are freely estimated and the remaining are held exactly equal. However, it has been shown that approximate MI provides a better alternative when relatively small parameter differences occur (Van de Schoot et al., 2013).

Bayesian approximate MI is based on less strict assumptions about parameter differences, i.e. parameter differences are allowed some leeway or "wiggle room" (Van de Schoot et al., 2013, p. 1). Differences are assumed to be almost zero, but not exactly. However, differences are still kept at a minimum to ensure that concepts remain approximately comparable. The degree of leeway can be expressed in terms of a prior distribution, usually a normal distribution with a mean of zero and a small variance, i.e. $N(0, \upsilon)$. The prior variance υ can be changed to express different levels of confidence in the assumption that the differences are (close to) zero. The larger υ, the more a prior is uninformative and reflects uncertainty. Values of less than 1 for υ are considered small (Asparouhov & Muthén, 2017). Figure 1 shows cross-time differences for a measurement parameter τ. Panel a displays exact MI with no differences and panel b shows Bayesian approximate MI with differences in the shape of a normal distribution with a mean of zero and a small variance. The prior variance in panel b implies that 95% of the variation of the differences is between -0.2 and 0.2 (B. Muthén & Asparouhov, 2012, p. 316).

4 In Brief: Principles of Bayesian Analysis

The principles of Bayesian statistics are described in great detail by several authors (Gelman, Carlin, Stern, & Rubin,

[2]Alternative approaches exist that estimate the variance of parameter differences from the data. These are random effect (Fox & Verhagen, 2017) or multilevel models (Davidov, Dülmer, Schlüter, Schmidt, & Meuleman, 2012; Jak, Oort, & Dolan, 2014).

[3]Note that this terminology can be used compatibly for cross-group and cross-time analysis. From here on, we refer to the cross-time case.

[4]Additional restrictions can be tested for other parameters, for example, equal disturbances of the same items across time, equal factor variances and covariances across time, and equal latent factor means across time (Vandenberg & Lance, 2000). However, for our purposes scalar MI is sufficient.

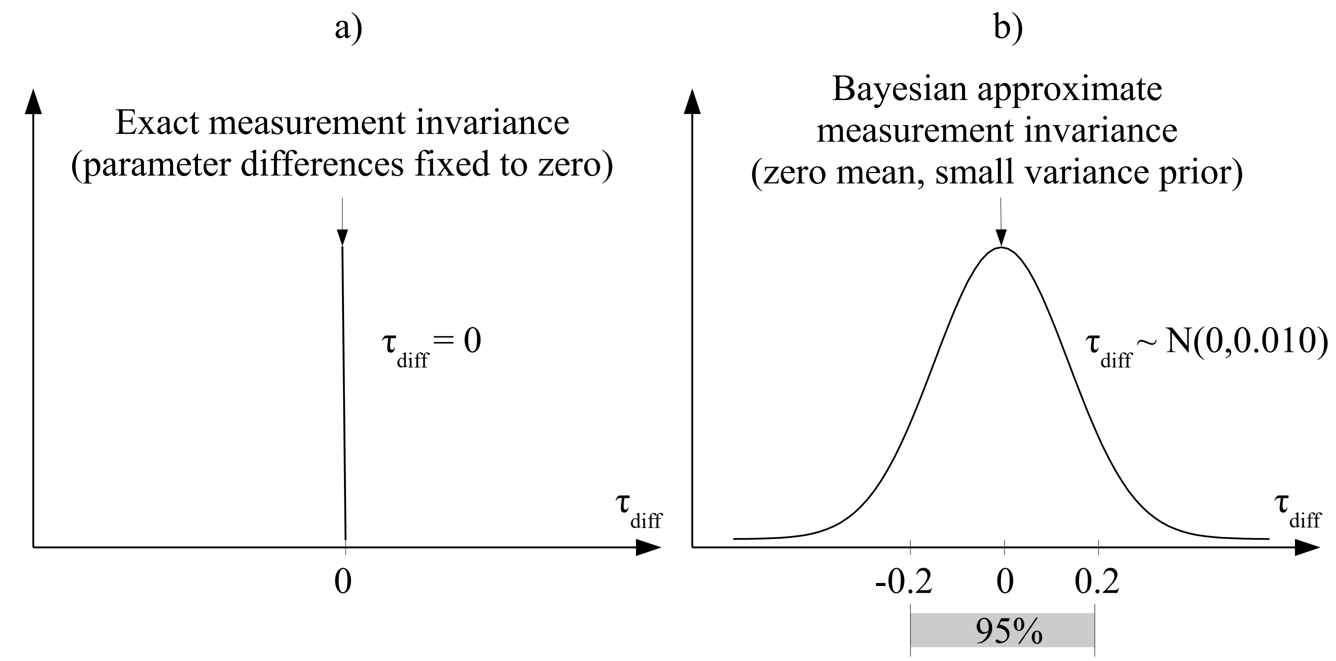

Figure 1. Exact (a) and Bayesian approximate (b) MI (see B. Muthén & Asparouhov, 2012)

2004; Kaplan, 2014; Lynch, 2007) and have also been introduced to structural equation modeling and related topics (Kaplan & Depaoli, 2012; Lee, 2007; B. Muthén & Asparouhov, 2012; Scheines, Hoijtink, & Boomsma, 1999; Song & Lee, 2012; Van de Schoot et al., 2014). The Bayesian perspective shares a specific understanding of probability that is related to the researcher's expectations or beliefs about a parameter of interest. When probability is generally seen as a quantification of uncertainty (Zyphur & Oswald, 2015), Bayesian probability refers to the researcher's uncertainty about the potential true value of a parameter. This implies parameters are treated as random and thus variable within a prespecified range of values that can be expressed in terms of a prior probability distribution. Useful and reasonable prior information as well as strong confidence in the prior belief can be translated into a sharp and narrow shaped prior distribution whereas uncertain information and poor confidence is reflected by a uniform and flat shaped distribution. A uniform and flat (uninformative) prior distribution does not favor any single parameter value to be more likely than another.[5]

Prior beliefs about the parameter of interest are updated with observed data (i.e., the likelihood) resulting in the posterior distribution, which is a trade-off between priors and data (i.e., between expectations and evidence). A strong prior will have more influence on the posterior and a weak prior will give way for the data to have a stronger influence on the posterior. The posterior distribution carries the usually desired information about a parameter. The posterior mean, mode, or median can serve as a point estimate and the posterior standard deviation is an indication of precision.

To determine the posterior parameter distribution, Bayesian estimation procedures use Markov chain Monte Carlo (hereafter: MCMC) sampling methods (Geyer, 2011; Lynch, 2007). MCMC algorithms (e.g., Gibbs sampler, Metropolis-Hastings) iteratively sample from multivariate posterior distributions or conditional distributions of each parameter. Previous parameter values are repeatedly updated each time the algorithm passes through the process until the joint posterior distribution can be approximated. Since MCMC sampling methods directly generate the posterior distribution, asymptotic arguments (e.g., normality of a parameter distribution) are not needed.

The iterative process is called a chain and it is possible to use more than one chain to obtain a posterior solution. The first half of iterations in a chain is usually discarded as "burn-in phase" and the second half is used to approximate the joint posterior distribution. Convergence of the iterative process can be assessed with the potential scale reduction factor (hereafter: PSR) (Gelman & Rubin, 1992). The PSR assesses within- and between-chain variation where between-chain variation should be small compared to within-chain variation (see also Asparouhov and Muthén, 2010b; Gelman and Shirley, 2011; Kruschke, Aguinis, and Joo, 2012).

Models can be compared by means of the deviance infor-

[5]Prior distributions are not necessarily uniform. Commonly used conjugate prior distributions are discussed in Lynch (2007, :68–70)

mation criterion (hereafter: DIC; Spiegelhalter, Best, Carlin, & van der Linde, 2002). The DIC contains a model complexity term that captures the effective number of parameters:

$$p_D = \overline{D} - D(\bar{\theta}) \tag{2}$$

where $\overline{D}$ is the average deviance across MCMC iterations and $D(\bar{\theta})$ is the average of parameters across the MCMC iterations. The DIC is then calculated as:

$$\text{DIC} = \overline{D} + p_D \tag{3}$$

Since parameters based on small variance priors are not recognized as actual parameters that add to model complexity, the DIC may be preferred over the Bayesian information criterion (BIC; Schwarz, 1978) in BSEM (Asparouhov, Muthén, & Morin, 2015). The model with the smallest DIC value is preferred. Further, the DIC is equivalent to the Akaike Information Criterion (AIC; Akaike, 1973) when prior information is absent (Ellison, 2004) and approximately equivalent given uninformative prior information (Ando, 2010, p. 215; see also Ward, 2008).

The predictive quality of the posterior estimates (i.e., model fit) can be assessed with posterior predictive checks (Lynch & Western, 2004; Scheines et al., 1999). Based on the posterior, a synthetic data set is generated and the discrepancy between generated and observed data is approximated as a distribution of differences of test statistics T (e.g., χ^2) over a number of iterations. The proportion of test statistics T from generated data (y^{gen}) exceeding T from observed data (y) is calculated as a posterior predictive p value (hereafter: PPP), indicating the "probability that a future observation would exceed the observed data, given the model" Lynch, 2007, p. 156:

$$\text{PPP} = p(T(y^{\text{gen}}) \geq T(y)|y) \tag{4}$$

Values around .50 imply equally distributed discrepancies and thus good fit. Extreme values, such as PPP less than .05, indicate poor fit.

It should be noted that the DIC and PPP are not suited to evaluate the adequacy of small variance priors (Hoijtink & Van de Schoot, 2017). However, we will use both criteria in conjunction with the procedure of Asparouhov et al. (2015, Appendix A) who propose to run a series of models with different prior variances until increasing the prior variance is considered unnecessary (see below).

The use of Bayesian methodology has steadily increased in recent decades (Van de Schoot, Winter, Ryan, Zondervan-Zwijnenburg, & Depaoli, 2016). Several studies have covered diverse substantive and methodological areas of research on MI and some successfully used approximate MI in situations where exact parameter constraints have been too strict and suggested incomparability of the variables under study (Bujacz, Vittersø, Huta, & Kaczmarek, 2014; Chan et al., 2015; Chiorri, Day, & Malmberg, 2014; Cieciuch, Davidov, Algesheimer, & Schmidt, 2017; Cieciuch, Davidov, Schmidt, Algesheimer, & Schwartz, 2014; Davidov et al., 2015; De Bondt & Van Petegem, 2015; Falkenström, Hatcher, Skjulsvik, Holmqvist Larsson, & Holmqvist, 2015; Jackson, Gucciardi, & Dimmock, 2014; Van de Schoot et al., 2013; Zercher, Schmidt, Cieciuch, & Davidov, 2015). However, to the best of our knowledge there exists no application with panel data.

5 Empirical Application

5.1 Data

Data for this analysis is taken from the German sociological and criminological panel study "Crime in the modern city (Crimoc)".[6] The study examines the onset of delinquency during early adolescence and the development of delinquency and explanatory dimensions during adolescence and young adulthood (Boers, Reinecke, Seddig, & Mariotti, 2010; Seddig, 2014, 2016). The study examines causes, consequences, and the development of delinquent behavior from adolescence to adulthood. In 2002 the study was first implemented in the city of Duisburg, Germany. A cohort of approximately 3,500 seventh-grade students (mean age 13) from different schools was interviewed annually in schools using paper-and-pencil questionnaires or via a postal questionnaire survey. Since 2009 data collection is conducted biennially. In 2017 the 12th survey was conducted.

The current analysis uses 4 waves of panel data covering ages 14, 16, 18, and 20. Although more data points are available, we chose this particular data set for two reasons. First, to observe substantial change in the latent means of hedonism we considered a longer time period spanning from early adolescence until young adulthood. Second, however, to keep the application simple and parsimonious, we decided 4 data points are adequate to illustrate the approximate MI procedure. Hence, we accepted two year gaps between measurement occasions.

The 4 waves of panel data contain information on N = 1,002 respondents (64% females; 36% males). Obviously, panel dropout is an issue in the Crimoc study (Reinecke, 2013). This can be explained by the code-based and anonymized panel construction procedure. A self-generated code based on six time-invariant personal characteristics has to be filled out by the respondents prior to each assessment. The code is used to identify data of the same respondent across time. However, many respondents do not reconstruct it correctly. Thus, compared to the separate cross-sectional

[6]The study is funded by the German Research Foundation (Deutsche Forschungsgemeinschaft). Project directors are Klaus Boers (University of Münster, Germany) and Jost Reinecke (University of Bielefeld, Germany). Additional information about the study can be found at http://www.crimoc.org.

Table 1
Means and standard deviations (n=1,002)

	Total		Female		Male	
Age	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
y_1 *(Desires)*						
14	3.00	1.19	3.00	1.17	3.00	1.23
16	2.80	1.14	2.81	1.13	2.80	1.17
18	2.49	1.12	2.45	1.13	2.57	1.11
20	2.29	1.03	2.21	1.01	2.44	1.05
y_2 *(Excitement)*						
14	2.45	1.14	2.40	1.13	2.53	1.14
16	2.42	1.06	2.30	1.00	2.63	1.14
18	2.13	0.97	1.95	0.92	2.47	0.98
20	2.00	0.93	1.81	0.87	2.36	0.94
y_3 *(Pleasure)*						
14	3.10	1.12	3.03	1.07	3.22	1.21
16	3.12	1.06	3.03	1.01	3.27	1.13
18	3.03	1.04	2.95	1.03	3.18	1.04
20	3.06	0.98	2.97	0.98	3.22	0.96

samples, the panel contains fewer males, fewer students from secondary schools, and fewer delinquents (Seddig, 2016). Consequently, the levels of hedonistic orientation may be underestimated.

5.2 Measures

To illustrate the application of approximate MI with Bayesian CFA we focus on adolescents' preferences to maximize pleasure and enjoy life. Such a hedonistic orientation is characterized by an affection for enjoyment, excitement, and consumption pronouncing the immediate gratification of short-term desires. As one of the ten universal values captured in Schwartz's theory of basic human values (Schwartz, 1992, 1994; Schwartz et al., 2012, 4), the concept of hedonism plays not only a prominent role in social psychology and cross-cultural research, but also in sociology (e.g. Hitlin & Piliavin, 2004). Hedonistic orientations are widely accepted among many adolescents and serve as a relevant factor in the explanation of various phenomena, such as youth delinquency (e.g. Boers et al., 2010; Seddig, 2014) or nonparticipation in higher education (e.g. Leitgöb, Paseka, Bacher, & Altrichter, 2012). Hence, we consider the development of hedonistic orientations as an appropriate example for a broad audience of researchers from various disciplines.

We will analyze hedonism with multiple indicators at each time point. Three indicator variables measure the concept of hedonism. Students were asked to rate their "understanding for people who do what they desire" (y_1) as well as their personal desires for "living a life of pleasure" (y_2) and "excitement" (y_3). These items were assessed on a 5-point scale, where 1 indicates the lowest and 5 indicates the highest level of agreement. Thus, we treated the scale as continuous (Finney & DiStefano, 2013; Rhemtulla, Brosseau-Liard, & Savalei, 2012). Table 1 shows item means and standard deviations. Figure 2 shows the longitudinal confirmatory factor model.

5.3 Analytical Procedure

We first test exact MI with the traditional (full information) maximum likelihood approach (Finkbeiner, 1979). Therefore, we follow a "bottom-up" strategy and begin with the least restrictive model (configural model). We then consecutively introduce cross-time equality restrictions on the factor loadings (metric model) and intercepts (scalar model). Finally, we replace exact equality constraints for parameters found to be non-invariant across time points with approximate equality constraints by using zero-mean, small-variance priors in the Bayesian approach.

We follow the outline of (Asparouhov et al., 2015, Appendix A) to assess which prior variance allows sufficient wiggle room for parameter differences, but is small enough so that the differences can be considered approximately zero. The procedure requires to begin with a very small prior variance (e.g., $\upsilon = 0.001$) and then gradually increase it in small steps to achieve identified solutions that allow to separate minor parameter differences from model misspecifications. The process is monitored with regard to model fit (DIC, PPP, and the 95% credibility intervals for the difference between observed and replicated chi-square values) and convergence. Increasing the prior variance is considered unnecessary when

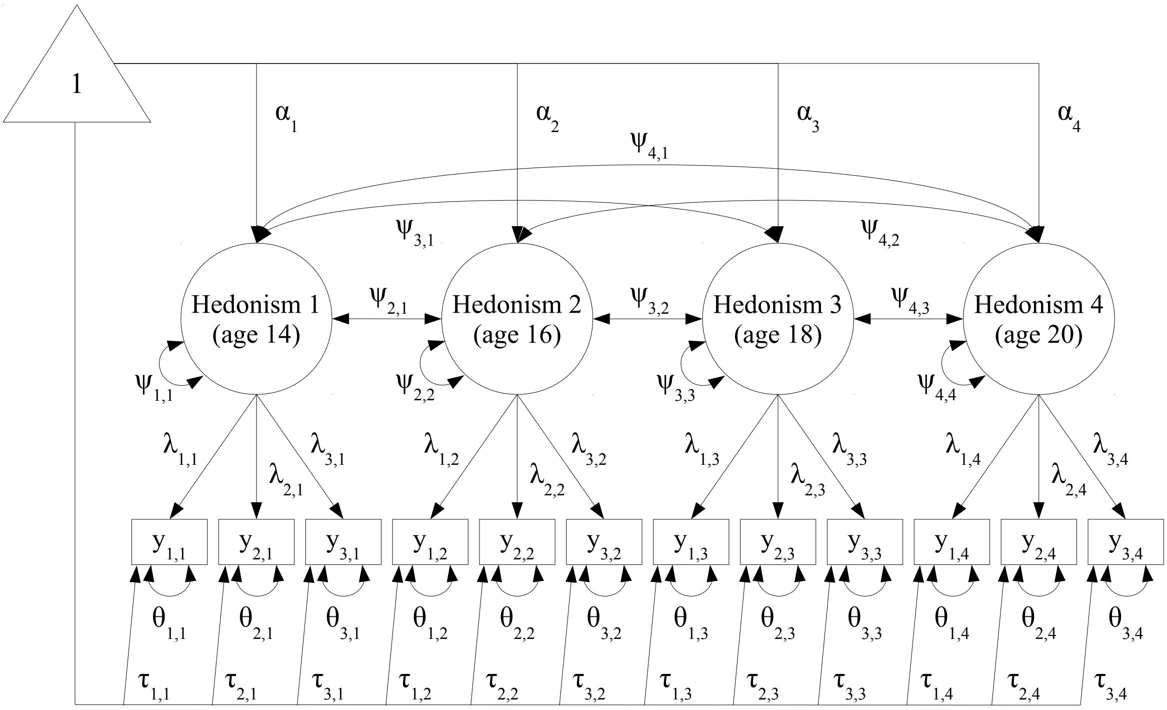

Figure 2. Longitudinal panel CFA model “hedonism”

Table 2
Maximum Likelihood CFA Model Fit (N = 1,002)

Model	χ^2 (df)	RMSEA	SRMR	CFI
Configural	35.630 (30)	0.014	0.017	0.998
Metric	42.982 (36)	0.014	0.021	0.997
Scalar	151.015 (42)	0.051	0.042	0.955

Note. df=degrees of freedom; RMSEA=root mean square error of approximation; SRMR=standardized root mean square residual; CFI=comparative fit index.

model fit differences between models become irrelevant.

The programm M*plus* Version 8 (L. Muthén & Muthén, 1998-2017) was used with the Gibbs sampler and two MCMC chains in this study. A process is assumed to be converged when the second half of the iterations has PSR values lower than defined by the default of the M*plus* command “*bconvergence*” (Asparouhov & Muthén, 2010b). The treatment of missing data in Bayesian estimation is similar to full information maximum likelihood, that is, the full information available from the data is used assuming missing at random (Asparouhov & Muthén, 2010a, 2010b).

5.4 Results

The results based on maximum likelihood CFA suggest that at least exact metric MI is supported (Table 2). Global fit statistics for scalar MI are also within the acceptable range (Hu & Bentler, 1999). However, the χ-difference to the metric model is very large. Further, Chen (2007) developed criteria for other fit statistics to indicate when MI is not given. According to these criteria, change in RMSEA should be less than .015, change in SRMR should be less than .010, and change in CFI should be less than -.010. The scalar model does not meet these criteria. This implies a misspecification in the scalar model that leads to a substantial decrease in model fit, but not to an overall rejection of the model. However, although the global fit of the scalar model is tolerable, the metric model may be the better choice. We admit that this is a very strict interpretation of model fit differences, which should be used with caution and may only serve illustrative purposes in this example. However, the Bayesian exact scalar MI model by far has the worst fit (upper section in Table 3).[7]

According to the maximum likelihood modification indices, the misspecification in the scalar model is located at the intercept level. The intercepts of items y_1 and y_3 are non-invariant at time points 1 and 4 (age 14 and 20). Therefore,

[7]In all models, autocorrelated disturbances were considered to fit the data structure appropriately (see the M*plus* input in Figure 3).

```
Analysis:
  Estimator=Bayes; Chains=2; Proc=2; Stvalues=ml;
  Biterations=1000000(100000); Bseed=3010;
Model:
  hedo1 by y11* y21* y31* (1-3); [y11 y21 y31] (i11-i13);
  hedo2 by y12* y22* y32* (1-3); [y12 y22 y32] (i21-i23);
  hedo3 by y13* y23* y33* (1-3); [y13 y23 y33] (i31-i33);
  hedo4 by y14* y24* y34* (1-3); [y14 y24 y34] (i41-i43);
 hedo1@1 hedo2-hedo4;
 [hedo1@0 hedo2-hedo4];
 y11 with y12 y13 y14;
 y12 with y13 y14;
 y13 with y14;
 y21 with y22 y23 y24;
 y22 with y23 y24;
 y23 with y24;
 y31 with y32 y33 y34;
 y32 with y33 y34;
 y33 with y34;
 Model priors:
 Do(1,3) diff(i1#-i4#)~N(0,0.01);
```

Figure 3. M*plus* input

Table 3
Bayesian CFA Model Fit (N = 1,002)

	Prior	DIC	PPP	95% CI
Exact MI				
Configural		32976	0.350	[−28.4, +40.5]
Metric		32972	0.306	[−26.0, +42.5]
Scalar		33069	0.000	[+76.0, +144.1]
Approximate MI				
Scalar	*N*(0,0.001)	33014	0.002	[+17.4, +91.2]
Scalar	*N*(0,0.005)	32979	0.174	[−18.4, +51.8]
Scalar	*N*(0,0.010)	32974	0.249	[−23.4, +45.9]
Scalar	*N*(0,0.100)	32972	0.302	[−26.0, +43.0]
Scalar	*N*(0,0.500)	32973	0.293	[−25.4, +43.6]

DIC=deviance information criterion; PPP=posterior predictive p-value; CI=credibility interval; MI=measurement invariance.

we specified zero-mean, small-variance priors for the differences of all intercepts across time. We began the Bayesian analyses with a very small prior variance and successively increased it while we monitored DIC, PPP, and 95% credibility intervals for the difference between observed and replicated chi-square value. Although this strategy is not a test of the adequacy of the prior variances (Hoijtink & Van de Schoot, 2017), it is possible to decide if model fit substantially improves with larger prior variances. An extract of the M*plus* input for the approximate MI model with prior variance $\upsilon = 0.010$ is given in Figure 3.

The results show that the smallest prior variance $\upsilon = 0.001$ does not fit the data, i.e. υ is too close to zero. By gradually adjusting υ we were able to achieve good model fit while maintaining convergence and identifiability. The less strict prior variances $\upsilon = 0.100$ and $\upsilon = 0.500$ reveal no major improvement of model fit compared to $\upsilon = 0.010$. Moreover, models with $\upsilon = 0.100$ and $\upsilon = 0.500$ are very slow to converge, need much more iterations, and the quality of model identification diminishes.[8] Further, the DIC, PPP, and limits of 95% credibility interval for the difference between observed and replicated chi-square values for $\upsilon = 0.010$ do not substantially differ from the exact metric MI model. Thus, $\upsilon = 0.010$ is deemed sufficient to consider minor deviations from exact intercept equivalence and we assume that approximate scalar MI holds. Thus, it is reasonable to compare the latent means of hedonism across time.

Figure 4 shows a part of the M*plus* results output for the model with prior variance $\upsilon = 0.010$. It can be seen that the factor loadings are held exactly equal across time and small differences exist for the intercepts across time. The latent means of hedonism across time imply that a hedonistic orientation declines in the process of maturation from youth to young adulthood.

The latent means for hedonism across time can be compared with regard to the use of exact or approximate MI (Table 4). The decline of hedonism can be observed regardless of the model.[9] However, two differences are noteworthy.

[8]The model with $\upsilon = 0.500$ has poor convergence even after 500.000 iterations. 100.000 were sufficient for $\upsilon = 0.010$. 100.000 iterations were used for all other models, except 200.000 were used for $\upsilon = 0.100$.

[9]In this specific case (but not necessarily in other cases), the

Table 4
Hedonism Latent Mean Comparison (n=1,002)

	Exact MI (ML)		Exact MI (Bayes)		Appr. MI (Bayes)	
Age	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.
14	0.000	0.000	0.000	0.000	0.000	0.000
16	–0.161	0.049	–0.167	0.051	–0.144	0.113
18	–0.574	0.056	–0.600	0.062	–0.556	0.116
20	–0.777	0.061	–0.813	0.070	–0.738	0.123

```
MODEL RESULTS
                              Posterior  One-Tailed         95% C.I.
                   Estimate       S.D.      P-Value   Lower 2.5%  Upper 2.5%  Significance

hedo1      BY
    y11               0.668      0.044      0.000      0.581      0.754      *
    y21               0.523      0.037      0.000      0.450      0.596      *
    y31               0.514      0.035      0.000      0.444      0.583      *

hedo2      BY
    y12               0.668      0.044      0.000      0.581      0.754      *
    y22               0.523      0.037      0.000      0.450      0.596      *
    y32               0.514      0.035      0.000      0.444      0.583      *

hedo3      BY
    y13               0.668      0.044      0.000      0.581      0.754      *
    y23               0.523      0.037      0.000      0.450      0.596      *
    y33               0.514      0.035      0.000      0.444      0.583      *

hedo4      BY
    y14               0.668      0.044      0.000      0.581      0.754      *
    y24               0.523      0.037      0.000      0.450      0.596      *
    y34               0.514      0.035      0.000      0.444      0.583      *

Means
    hedo1             0.000      0.000      1.000      0.000      0.000
    hedo2            -0.144      0.113      0.099     -0.369      0.075
    hedo3            -0.556      0.116      0.000     -0.792     -0.334      *
    hedo4            -0.738      0.123      0.000     -0.991     -0.511      *

 Intercepts
    y11               2.986      0.037      0.000      2.913      3.058      *
    y21               2.448      0.035      0.000      2.380      2.516      *
    y31               3.125      0.035      0.000      3.057      3.194      *
    y12               2.901      0.076      0.000      2.753      3.052      *
    y22               2.486      0.062      0.000      2.364      2.608      *
    y32               3.195      0.061      0.000      3.076      3.315      *
    y13               2.865      0.076      0.000      2.717      3.016      *
    y23               2.425      0.063      0.000      2.305      2.549      *
    y33               3.306      0.061      0.000      3.187      3.426      *
    y14               2.793      0.080      0.000      2.642      2.954      *
    y24               2.397      0.064      0.000      2.274      2.524      *
    y34               3.421      0.062      0.000      3.303      3.544      *
```

Figure 4. M*plus* model results output ($\upsilon = 0.010$; point estimates are medians)

```
DIFFERENCE OUTPUT

       Average     Std. Dev.       Deviations  from  the  Mean
                                   I11          I21       I31      I41
1        2.886        0.053        0.099*       0.014    -0.021   -0.092*

                                   I12          I22       I32      I42
2        2.439        0.044        0.009        0.047    -0.013   -0.042

                                   I13          I23       I33      I43
3        3.262        0.042        0.136*      -0.067     0.044    0.159*
```

Figure 5. M*plus* difference output (υ = 0.010)

First, the decline appears to be less pronounced in the approximate MI solution. Second, the standard errors are larger in the approximate MI solution. For example, the change in latent means of hedonism between age 14 and 16 is considered not significant with approximate MI, but significant with both exact MI solutions.

Finally, the non-invariant parameters can be identified based on the approximate MI solution (B. Muthén & Asparouhov, 2013). Non-invariance is determined as the difference of a particular parameter at a time point from the average of estimates for the particular parameter across time. If a difference of zero is outside of the 2.5% and 97.5% quantiles of the posterior distribution of differences, the difference is assumed to be significant and the parameter is assumed non-invariant. Figure 5 shows the difference output produced by M*plus* for the Bayesian CFA model with prior variance υ = 0.010. The column "Average" refers to the average of estimates for parameters across time. Significant deviations from the average (i.e., non-invariant parameters) are labeled with an asterisk in the columns "Deviations from the Mean". The results indicate that the intercepts of items y_1 and y_3 significantly deviate from the cross-time average at time points 1 and 4.

6 Conclusion

The analyses reported in this paper used Bayesian approximate MI to test the comparability of students' preferences to maximize pleasure and enjoy life ("hedonism"). Since comparability is important for longitudinal research and the study of processes and change of latent variables, a careful assessment of MI has to be conducted before drawing substantive conclusions. The Bayesian approach offers a flexible alternative to the exact approach where all model parameters are held exactly equal across time points. Approximate MI allows for a small degree of deviation from exact zero constraints while the differences across time points are kept at a minimum.

Regardless of using maximum likelihood or Bayesian estimation, the results of testing the longitudinal MI of hedonism showed that exact scalar MI may not be the best solution. However, the use of zero-mean, small-variance priors lead to support for approximate scalar MI. A small prior variance (υ = 0.010) was deemed sufficient to account for minor differences of item intercepts across time. These minor differences already led to a substantial deterioration of fit of the exact scalar MI model and substantive conclusions may be not trustworthy. The Bayesian approach showed that the latent means are comparable at least based on approximate MI. Although the differences in the latent means from exact and approximate MI models were small, they may be more severe in other studies. In any case, a careful test of invariance properties is recommended over alternative strategies, such as using observed composite scores (Steinmetz, 2013).

For practical researchers it may not be easy to decide when and how to use the Bayesian approach. However, considering what can be achieved by using small variance priors for testing MI may open up a new perspective. When model fit of more restrictive (e.g., scalar) models is worse than that of less restrictive (e.g., metric) models, researchers may take a closer look at the actual parameter differences that elicit the misfit. Sometimes already very small parameter differences across time may lead to a decision against a particular model in the traditional (maximum likelihood) approach. In this case, using the Bayesian approach to reduce the stringency of parameter equality may be very reasonable and comparability can be justified, even if the parameter differences precluded a comparison using the traditional approach. However, if the differences are actually large and not regarded as negligible, too much leeway for the parameter differences is not advisable. Very large prior variances may be too vague and the estimation is dominated entirely by the data. Consequently, assuming comparability may be reasonable, even if model fit is good. In contrast, very small prior variances

same development can be observed when mean differences of composite scores are compared (not displayed). However, this similarity does not imply that testing for MI is unnecessary. Composite mean differences have been shown to be trustworthy only when full scalar MI holds for the underlying measures (Steinmetz, 2013).

may be too close to exact zero differences and result in poor model fit. A rule of thumb for limits of the prior variance does not seem to be useful. Many times a balance is needed between the desired condition (i.e., exact parameter equality) and the situation found in the data. This balance can be achieved by the Bayesian posterior distribution of parameter differences: while the irregularities in the data are considered the differences are kept as close to zero as possible.

The results presented in this paper do not rely on a strict test of the hypothesis that parameter differences defined by the prior variance υ are approximately zero. We applied the strategy outlined by Asparouhov et al. (2015) where no single prior is preferred over another. A test of the particular hypothesis that parameters are approximately zero is the prior-posterior predictive p-value (Asparouhov & Muthén, 2017; Hoijtink & Van de Schoot, 2017). However, M*plus* does not yet provide the prior-posterior predictive p-value for approximate MI.

Further, the choice of a different than a zero-mean prior may be considered in cases where different types of response bias may give rise to assumptions about systematic reporting tendencies that contradict measurement equivalence (Billiet & Davidov, 2008; Cheung & Rensvold, 2000; Welkenhuysen-Gybels, Billiet, & Cambré, 2003). If, for example, a shift in the intercept or factor loading of a particular item is expected across measurement occasions, this could be expressed with a non-zero mean for the prior.

Van de Schoot et al. (2013) pointed out that approximate MI may not be appropriate in all situations. They suggest partial approximate MI as a further approach that may lead to even better solutions. Partial approximate MI allows minor differences only for those parameters that are non-invariant. Further, in the presence of large parameter differences, non-invariant parameters tend to be pulled toward the average of the parameter estimates which causes bias in latent variable means (Van de Schoot et al., 2013). A method to compensate for this bias is alignment (Asparouhov & Muthén, 2014; Marsh et al., 2017; B. Muthén & Asparouhov, 2014) where a component loss function (Jennrich, 2006) is used to generate solutions with many approximately invariant parameters and very few non-invariant parameters. However, using strict, partial, or approximate MI, or even applying the alignment procedure must be judged on a case-by-case basis. The equality constraints imposed by any of these approaches are sensible to the unique data set that is analyzed.

Acknowledgements

The authors would like to thank the directors of the "Crime in the modern city (Crimoc)" study, Prof. Klaus Boers and Prof. Jost Reinecke, for granting us access to the data.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second international symposium on information theory.* (pp. 267–281). Budapest: Akademiai Kaido.

Ando, T. (2010). *Bayesian model selection and statistical modeling*. Boca Raton, FL: CRC Press.

Asparouhov, T. & Muthén, B. (2010a). *Bayesian analysis of latent variable models using Mplus*. Muthén & Muthén. Technical report (www.statmodel.com). Los Angeles, CA.

Asparouhov, T. & Muthén, B. (2010b). *Bayesian analysis using Mplus: Technical implementation*. Muthén & Muthén. Technical report (www.statmodel.com). Los Angeles, CA.

Asparouhov, T. & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*, 1–14.

Asparouhov, T. & Muthén, B. (2017). Prior-posterior predictive p-values.

Asparouhov, T., Muthén, B., & Morin, A. J. S. (2015). Bayesian structural equation modeling with cross-loadings and residual covariance: Comments on Stromeyer et al. *Journal of Management*, *41*(6), 1561–1577. doi:10.1177/0149206315591075

Billiet, J. & Davidov, E. (2008). Testing the stability of an acquiescence style factor behind two interrelated substantive variables in a panel design. *Sociological Methods & Research*, *36*(4), 542–562.

Boers, K., Reinecke, J., Seddig, D., & Mariotti, L. (2010). Explaining the development of adolescent violent delinquency. *European Journal of Criminology*, *7*(6), 499–520.

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

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2.). New York, NY: The Guilford Press.

Bujacz, A., Vittersø, J., Huta, V., & Kaczmarek, L. D. (2014). Measuring hedonia and eudaimonia as motives for activities: cross-national investigation through traditional and Bayesian structural equation modeling. *Frontiers in Psychology*, *5*, 984.

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

Chan, D. K. C., Ivarsson, A., Stenling, A., Yang, S. X., Chatzisarantis, N. L. D., & Hagger, M. S. (2015). Response-order effects in survey methods: A randomized controlled crossover study in the context of sport injury prevention. *Journal of Sport and Exercise Psychology*, *37*, 666–673.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504.

Cheung, G. W. & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equation modelling. *Journal of Cross-Cultural Psychology*, *31*, 187–212.

Chiorri, C., Day, T., & Malmberg, L.-E. (2014). An approximate measurement invariance approach to within-couple relationship quality. *Frontiers in Psychology*, *5*, 983.

Cieciuch, J., Davidov, E., Algesheimer, R., & Schmidt, P. (2017). Testing for approximate measurement invariance of human values in the European Social Survey. *Sociological Methods & Research*.

Cieciuch, J., Davidov, E., Schmidt, P., Algesheimer, R., & Schwartz, S. H. (2014). Comparing results of an exact vs. an approximate (Bayesian) measurement invariance test: a cross-country illustration with a scale to measure 19 human values. *Frontiers in Psychology*, *5*, 1–10.

Davidov, E., Cieciuch, J., Meuleman, B., Schmidt, P., Algesheimer, R., & Hausherr, M. (2015). The comparability of measurements of attitudes toward immigration in the European Social Survey. *Public Opinion Quarterly*, *79*, 244–266.

Davidov, E., Dülmer, H., Schlüter, E., Schmidt, P., & Meuleman, B. (2012). Using a multilevel structural equation modeling approach to explain cross-cultural measurement noninvariance. *Journal of Cross-Cultural Psychology*, *43*(4), 558–575.

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

De Bondt, N. & Van Petegem, P. (2015). Psychometric evaluation of the overexcitability questionnaire-two applying Bayesian structural equation modeling (BSEM) and multiple-group BSEM-based alignment with approximate measurement invariance. *Frontiers in Psychology*, *6*, 1963.

Ellison, A. M. (2004). Bayesian inference in ecology. *Ecology letters*, *7*(6), 509–520.

Falkenström, F., Hatcher, R. L., Skjulsvik, T., Holmqvist Larsson, M., & Holmqvist, R. (2015). Development and validation of a 6-item working alliance questionnaire for repeated administrations during psychotherapy. *Psychological Assessment*, *27*(1), 169–183.

Ferrer, E., Balluerka, N., & Widaman, K. F. (2008). Factorial invariance and the specification of second-order latent growth models. *Methodology*, *4*(1), 22–36.

Finkbeiner, C. (1979). Estimation for the multiple factor model when data are missing. *Psychometrika*, *44*(4), 409–420.

Finney, S. J. & DiStefano, C. (2013). Nonnormal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: a second course* (2nd ed., pp. 439–492). Charlotte, NC: Information Age.

Fox, J.-P. & Verhagen, A. (2017). Random item effects modeling for cross-national survey data. In E. Davidov, P. Schmidt, & J. Billiet (Eds.), *Cross-cultural analysis: methods and applications* (2nd ed.). New York, NY: Routledge.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). Boca Raton, FL: Chapman & Hall.

Gelman, A. & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*, 457–472.

Gelman, A. & Shirley, K. (2011). Inference from simulations and monitoring convergence. In S. Brooks, A. Gelman, G. Jones, & X.-L. Meng (Eds.), *Handbook of markov chain monte carlo* (pp. 163–174). Boca Raton, FL: Chapman & Hall.

Geyer, C. J. (2011). Introduction to markov chain monte carlo. In S. Brooks, A. Gelman, G. Jones, & X.-L. Meng (Eds.), *Handbook of markov chain monte carlo* (pp. 3–48). Boca Raton, FL: Chapman & Hall.

Hertzog, C. & Nesselroade, J. R. (2003). Assessing psychological change in adulthood: An overview of methodological issues. *Psychology and Aging*, *18*, 639–657.

Hitlin, S. & Piliavin, J. A. (2004). Values: Reviving a dormant concept. *Annual Review of Sociology*, *30*, 359–393.

Hoijtink, H. & Van de Schoot, R. (2017). Testing small variance priors using prior-posterior predictive p values. *Psychological Methods*. Retrieved from http://dx.doi.org/10.1037/met0000131

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

Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.

Jackson, B., Gucciardi, D. F., & Dimmock, J. A. (2014). Toward a multidimensional model of athletes' commitment to coach-athlete relationships and interdependent sport teams: A substantive-methodological synergy. *Journal of Sport and Exercise Psychology*, *36*(1), 52–68.

Jak, S., Oort, F. J., & Dolan, C. V. (2014). Measurement bias in multilevel data. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(1), 31–39.

Jennrich, R. I. (2006). Rotation to simple loadings using component loss functions: The oblique case. *Psychometrika*, *71*, 173–191.

Jöreskog, K. G. (1979). Statistical estimation of structural models in longitudinal-developmental investigations. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development*. New York, NY: Academic Press.

Kaplan, D. (2014). *Bayesian statistics for the social sciences*. New York, NY: Guilford Press.

Kaplan, D. & Depaoli, S. (2012). Bayesian structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 650–673). New York, NY: The Guilford Press.

Kruschke, J. K., Aguinis, H., & Joo, H. (2012). The time has come: Bayesian methods for data analysis in the organizational sciences. *Organizational Research Methods*, *15*(4), 722–752.

Lee, S.-Y. (2007). *Structural equation modeling: A Bayesian approach*. Chichester: Wiley.

Leitgöb, H., Paseka, A., Bacher, J., & Altrichter, H. (2012). Geschlechtsspezifische Leistungsdifferenzen in österreich. Vertiefende Analysen der PISA 2009-Daten. In F. Eder (Ed.), *PISA 2009. Nationale Zusatzanalysen für österreich* (pp. 308–336). Münster: Waxmann.

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: Guilford Press.

Lynch, S. M. (2007). *Introduction to applied Bayesian statistics and estimation for social scientists*. New York, NY: Springer.

Lynch, S. M. & Western, B. (2004). Bayesian posterior predictive checks for complex models. *Sociological Methods & Research*, *32*(3), 301–335.

Marsh, H. W., Guo, J., Nagengast, B., Asparouhov, T., Muthén, B., Parker, P. D., & Dicke, T. (2017). What to do when scalar invariance fails: The extended alignment method for multi-group factor analysis comparison of latent means across many groups. *Psychological Methods*. Retrieved from http://dx.doi.org/10.1037/met0000113

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525–543.

Meredith, W. & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care*, *44*(11), S69–S77.

Muthén, B. & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, *17*(3), 313–335.

Muthén, B. & Asparouhov, T. (2013). BSEM measurement invariance analysis.

Muthén, B. & Asparouhov, T. (2014). IRT studies of many groups: The alignment method. *Frontiers in Psychology*, *5*, 978.

Muthén, L. & Muthén, B. O. (1998-2017). *Mplus user's guide* (8.). Los Angeles, CA: Muthén & Muthén.

Reinecke, J. (2013). Growth curve models and panel dropouts: Applications with criminological panel data. *Netherlands Journal of Psychology*, *67*, 122–131.

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.

Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*(3), 354–373.

Scheines, R., Hoijtink, H., & Boomsma, A. (1999). Bayesian estimation and testing of structural equation models. *Psychometrika*, *64*(1), 37–52.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Advances in Experimental Social Psychology*, *25*, 1–65.

Schwartz, S. H. (1994). Are there universal aspects in the structure and contents of human values? *Journal of Social Issues*, *50*(4), 19–45.

Schwartz, S. H., Cieciuch, J., Vecchione, M., Davidov, E., Fischer, R., Beierlein, C., … Konty, M. (2012). Refining the theory of basic individual values. *Journal of Personality & Social Psychology*, *103*, 663–688.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

Seddig, D. (2014). Peer group association, the acceptance of norms and violent behaviour: A longitudinal analysis of reciprocal effects. *European Journal of Criminology*, *11*(3), 319–339.

Seddig, D. (2016). Crime-inhibiting, interactional and co-developmental patterns of school bonds and the acceptance of legal norms. *Crime & Delinquency*, *62*(8), 1046–1071.

Song, X.-Y. & Lee, S.-Y. (2012). *Basic and advanced Bayesian structural equation modeling. With applications in the medical and behavioral sciences*. Chichester: Wiley.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B*, *64*, 583–639.

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

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

Stoel, R. D., van den Wittenboer, G., & Hox, J. (2004). Methodological issues in the application of the latent growth curve model. In K. van Montfort, J. Oud, & A. Satorra (Eds.), *Recent developements on structural equation models: Theory and applications* (pp. 241–261). Amsterdam: Kluwer Academic Publishers.

Van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & van Aken, M. A. G. (2014). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development*, *85*(3), 842–860.

Van de Schoot, R., Kluytmans, A., Tummers, L., Lugtig, P., Hox, J., & Muthén, B. (2013). Facing off with Scylla and Charybdis: A comparison of scalar, partial, and the novel possibility of approximate measurement invariance. *Frontiers in Psychology*, *4*, 1–15.

Van de Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., & Zondervan-Zwijnenburg, M. (2015). Editorial: Measurement invariance. *Frontiers in Psychology*, *6*, 1064.

Van de Schoot, R., Winter, S., Ryan, O., Zondervan-Zwijnenburg, M., & Depaoli, S. (2016). A systematic review of Bayesian papers in psychology: The last 25 years. *Psychological Methods*, *21*(4).

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.

Ward, E. J. (2008). A review and comparison of four commonly used bayesian and maximum likelihood model selection tools. *Ecological Modelling*, *211*(1), 1–10.

Welkenhuysen-Gybels, J., Billiet, J., & Cambré, B. (2003). Adjustment for acquiescence in the assessment of the construct equivalence of Likert-type score items. *Journal of Cross-Cultural Psychology*, *34*, 702–722.

Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*, *4*(1), 10–18.

Zercher, F., Schmidt, P., Cieciuch, J., & Davidov, E. (2015). The comparability of the universalism value over time and across countries in the European Social Survey: Exact vs. approximate measurement invariance. *Frontiers in Psychology*, *6*, 1–11.

Zyphur, M. J. & Oswald, F. L. (2015). Bayesian estimation and inference: a user's guide. *Journal of Management*, *41*(2), 390–420.