MULTIVARIATE BEHAVIORAL RESEARCH
http://dx.doi.org/10.1080/00273171.2016.1142856

Exploring Factor Model Parameters across Continuous Variables with Local Structural Equation Models

Andrea Hildebrandt[a], Oliver Lüdtke[b,c], Alexander Robitzsch[b,c], Christopher Sommer[d], and Oliver Wilhelm[e]

[a]Department of Psychology, Ernst-Moritz-Arndt-Universität Greifswald; [b]Leibniz Institute for Science and Mathematics Education, Kiel University; [c]Centre for International Student Assessment, Technische Universität München; [d]Department of Psychology, Humboldt-Universität zu Berlin; [e]Department of Psychology and Education, Universität Ulm

ABSTRACT
Using an empirical data set, we investigated variation in factor model parameters across a continuous moderator variable and demonstrated three modeling approaches: multiple-group mean and covariance structure (MGMCS) analyses, local structural equation modeling (LSEM), and moderated factor analysis (MFA). We focused on how to study variation in factor model parameters as a function of continuous variables such as age, socioeconomic status, ability levels, acculturation, and so forth. Specifically, we formalized the LSEM approach in detail as compared with previous work and investigated its statistical properties with an analytical derivation and a simulation study. We also provide code for the easy implementation of LSEM. The illustration of methods was based on cross-sectional cognitive ability data from individuals ranging in age from 4 to 23 years. Variations in factor loadings across age were examined with regard to the age differentiation hypothesis. LSEM and MFA converged with respect to the conclusions. When there was a broad age range within groups and varying relations between the indicator variables and the common factor across age, MGMCS produced distorted parameter estimates. We discuss the pros of LSEM compared with MFA and recommend using the two tools as complementary approaches for investigating moderation in factor model parameters.

KEYWORDS
Local structural equation model; moderated factor analysis; multiple-group mean and covariance structures; age differentiation of cognitive abilities; WJ-III tests of cognitive abilities

Moderator variables are of great interest in the social sciences. They have been broadly defined in the seminal work by Baron and Kenny (1986) as “a third variable that affects the zero-order correlation between two other variables” (p. 1174). Moderators can be qualitative (e.g., sex, self-reported categorical ethnicity) or quantitative (e.g., age, socioeconomic status [as expressed, e.g., by the HISEI index], cognitive ability, or acculturation; see Szapocznik, Scopetta, Kurtines, & Aranalde, 1978). In several social science applications, quantitative moderator variables have unfortunately been recorded as categorical variables and have been analyzed as such—for example, age groups, high versus low socioeconomic status, high versus low ability levels, group membership based on social categories (e.g., “minorities,” “Asians”—instead of as continuous measures of acculturation (e.g., Berry, 2003).

In structural equation modeling (SEM) as generally used in empirical research, testing whether factor model parameters vary according to group membership (the moderator variable) is most commonly examined with multiple-group mean and covariance structure (MGMCS; e.g., Little, Card, Slegers, & Ledford, 2007) analyses. In MGMCS analyses, SEMs are simultaneously computed for several groups with parameters that are either set to be equal for all groups or estimated freely. If the overall model fit deteriorates significantly when parameters are constrained to equality, then the hypothesis that those particular parameters are invariant is not supported. MGMCS analysis is a widely used and accepted approach for investigating factorial invariance across categorical context variables (i.e., variables that define groups). However, often the context variable of interest is not categorical per se but can and ought to be conceptualized as a continuous variable. Because the number of observations for each measured value of a continuous variable is usually too small for estimating separate factor models that require large samples for each value of the continuous moderator, the common procedure is to pool participants into larger groups. Thus, in substantive research, the originally continuous variable is commonly treated as if it were, in fact, categorical.

Methodological problems related to the categorization of continuous variables are well known (for a review, see

CONTACT Andrea Hildebrandt andrea.hildebrandt@uni-greifswald.de Ernst-Moritz-Arndt-Universität Greifswald, Department of Psychology, Franz-Mehring-Strasse 47, 17487 Greifswald, Germany.
†The first three authors contributed equally to this manuscript.
Supplemental data for this article can be accessed on the publisher's website

MacCallum, Zhang, Preacher, & Rucker, 2002; Preacher, Rucker, MacCallum, & Nicewander, 2005). However, problems related to the categorization of continuous context variables in MGMCS have not been sufficiently evaluated. Obviously, some of the issues presented by MacCallum et al., (2002) apply here as well—for example, losing information about individual differences within groups and an increased risk of overlooking nonlinear relations. Among others, Hildebrandt, Wilhelm, and Robitzsch (2009) criticized the use of cutoff scores on a continuous moderator to build categories because those cutoffs are inevitably arbitrary. Hildebrandt et al. argued that categorization is problematic also because it exacerbates the detection of a change onset. When observations that differ across the range of a continuous variable are grouped, variation within those groups cannot be detected. This is because within groups, observations are treated as if they were equal regarding group differences across variables of interest. Therefore, continuous moderators should be treated as continuous variables, not as categorical variables.

As alternatives to MGMCS analyses, several authors (Curran et al., 2014; Hildebrandt et al., 2009; Liu, Magnus, & Thissen, 2015; Merkle & Zeileis, 2013; Molenaar, Dolan, Wicherts, & van der Maas, 2010; Tucker-Drob, 2009) have proposed approaches that do not require the categorization of continuous moderator variables. In this article, two of these approaches, local structural equation modeling (LSEM; Hildebrandt et al., 2009; Hülür, Wilhelm, & Robitzsch, 2011) and moderated factor analysis (MFA; Bauer & Hussong, 2009; Curran et al., 2014) will be presented in detail. We used empirical data to compare the two approaches with each other and with MGMCS analyses. The main aim was to illustrate the methods and provide code that substantive researchers can use to implement LSEM because such information has been lacking in the available literature that introduced the LSEM approach. With the empirical example, we investigated age-related changes in intelligence. In addition, we conducted a small simulation to investigate the statistical behavior of the LSEM approach.

The age differentiation hypothesis: Introducing the data example

We conducted the demonstration in the current artice to test a theoretically and methodologically important controversy in developmental psychology—Garrett's (1946) *age differentiation hypothesis*—which proposes that the relations between cognitive abilities monotonically decrease from childhood to young adulthood. Thus, the age differentiation hypothesis postulates that an increase in a cognitive specialization across childhood may be caused by individual differences in noncognitive factors, including motivation and interests. The rationale behind the hypothesis is that children develop different motivational levels for diverse aspects of cognition and achieve different intellectual interests until young adulthood. Diverging motivations and interests thus lead to differentiated aspects of cognition because they influence the investments children make in one but not another type of cognition. Different levels of investment and practice lead to higher efficiency in one as compared with another cognitive domain, thus to the increasing independence of cognitive ability factors.

We will describe analyses of cross-sectional data that were collected from an age-heterogeneous subsample from the normative sample of the Woodcock-Johnson III (WJ-III) Tests of Cognitive Abilities (Woodcock, McGrew, & Mather, 2001, 2007). We tested the differentiation hypothesis by examining the variation in *g*, or the common-factor saturation of cognitive abilities, across age. Thus, in this example, age was the moderator variable.

The WJ-III Tests of Cognitive Abilities (Woodcock, McGrew, Schrank, & Mather, 2001, 2007) are conceptualized on the basis of the Cattell-Horn-Carroll (CHC) theory of cognitive abilities, which is a fusion of Cattell's (1941) and Horn's (1965) theories of *fluid* and *crystallized* intelligence and Carroll's (1993) *three-stratum theory* of cognitive abilities. CHC theory postulates a hierarchically organized structure of human intelligence. Numerous narrow abilities are located on the lowest factor level in this model; these abilities are then grouped into broad second-order abilities. The broad abilities include, among others, fluid and crystallized intelligence as the most important intellectual capacities. Then, a common factor called *general intelligence*, or simply *g*, is postulated to be located at the highest level (McGrew, Schrank, & Woodcock, 2007). Consistent with earlier applications (Tucker-Drob, 2009), all of our analyses were based on composite variables representing second-order factors of cognitive abilities from the WJ-III. With regard to the psychological content, our main goal was to model the relations between these abilities and their common factor *g*.

The general model of interest, which we aimed to approximate with three different analytical approaches (MGMCS, LSEM, and MFA), is a one-dimensional common factor model (see Figure 1). However, please note that all modeling approaches discussed here can be applied to investigate more complex factor analytic structures, such as multiple factor confirmatory factor analyses (CFA), multifactorial SEMs, or hierarchical factor models (see Hildebrandt et al., 2009, for an example of LSEM applied to a hierarchical structure model).

Molenaar, Dolan, and Verhelst (2010) investigated differentiation in the higher order factor model and

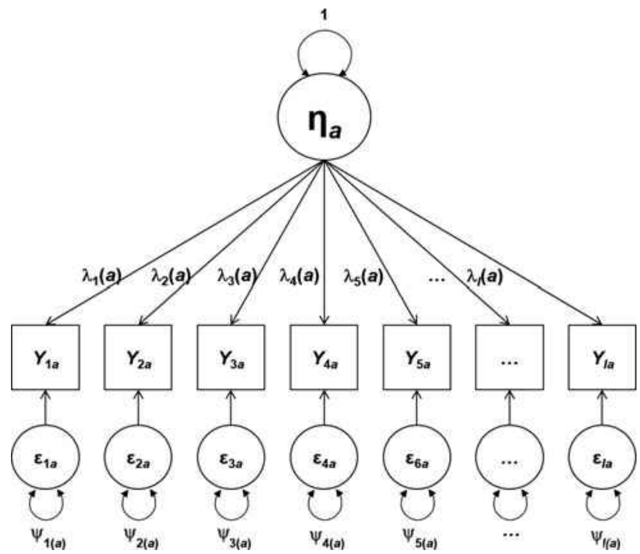

Figure 1. Path diagram of the one-factor model. η_a = common factor; $\lambda_i(a)$ = factor loadings depending on age; Y_i = indicators; ε_{ia} = residual variables depending on age; $\psi_i(a)$ = residual variances depending on age.

identified five possible sources of differentiation: (a) residual variances of the indicators, (b) loadings on the first-order factor, (c) first-order factor residual variances, (d) first-order factor loadings on the general factor, and (e) variance of the general factor. In the current study, we investigated differentiation in the one-factor model, using broad abilities as indicators. Thus, the possible sources of differentiation consisted of the residual variances of the scaled broad abilities, the factor loadings, and the factor variance. For factor identification purposes, we standardized the factor and freely estimated all factor loadings. Thus, differentiation was explored on the level of factor loadings and the residual variance of indicators. If heteroscedastic residuals were to occur, the conclusion would be that the measurement precision of broad abilities varies across age. If age-related changes in factor loadings were to occur, the conclusion would concern the *g*-saturation of indicators and thus the differentiation or dedifferentiation of cognitive abilities as previously defined.

Introducing the model to investigate age differentiation

Assume for each person a vector of variables $(A, Y_1 \ldots, Y_i, \ldots, Y_I)$, where A denotes a person's age and Y_i $(i = 1, \ldots, I)$ the person's performance for each of the first-order abilities. In general, at the population level, we study the conditional means $\mu_i(a) = E(Y_i \mid A = a)$ and the conditional covariances $\sigma_{ii'}(a) = Cov(Y_i, Y_{i'} | A = a)$ of the first-order abilities, where a denotes specific age values. The conditional covariance matrix $\Sigma(a)$, including the conditional variances and covariances $\sigma_{ii'}(a)$, should be represented by a one-dimensional common factor model as follows:

$$\Sigma(a) = \Lambda(a)\Lambda(a)^T + \Psi(a), \qquad (1)$$

where $\Lambda(a)$ is a column vector of loadings (at a specific age point a) and $\Psi(a)$ is an $I \times I$ matrix of age-conditional error variances and covariances (usually assumed to be diagonal). Note that the age differentiation hypothesis is a hypothesis about the age-conditional covariances and that restrictions are imposed only on the covariances and not on the conditional means. Please also note that age-related mean changes may be associated with age-related changes in the covariance structure. The one-factor model for each indicator and for each age value a can thus be written as

$$Y_{ia} = \nu_i(a) + \lambda_i(a) \cdot \eta_a + \varepsilon_{ia}, \qquad (2)$$

where $\upsilon_i(a) = E(Y_{ia})$ are age-specific intercepts and ε_{ia} are age-specific residuals, which are assumed to be uncorrelated with the common factor η_a and with the age-specific variances $\psi_i(a) = Var(\varepsilon_{ia})$. Researchers may also specify that the residual variances vary across the moderator. As we have outlined, in previous work, residuals have also been considered a source of differentiation (Molenaar, Dolan, & Verhelst, 2010). The common factor η_a is (locally) identified by setting its conditional mean to 0 and fixing its variance to 1:[1] $E(\eta_a) = 0$ and $Var(\eta_a) = 1$. As pointed out before, in the present study, we were interested in only the (age-conditional) covariance structure, and thus, the intercepts $\nu_i(a)$ were not restricted. By not restricting the intercepts, the mean structure of the model in Equation (2) became saturated and was completely determined by the age trends in the indicator variable Y_{ia}; that is, the indicator-specific means equal the age-conditional means: $\nu_i(a) = \mu_i(a)$. Loadings on the common factor quantify how much variance the broad abilities share with each other. Therefore, if age differentiation occurs, the factor loadings should systematically vary with age in that they should be higher in childhood and decrease monotonically with increasing age.

This model should be estimated for each possible value of the moderator variable, and the course of the parameter estimates would be inspected. However, this is rarely feasible because only a finite subset of all the possible values of the continuous context variable is usually available in empirical research. In addition, a substantial part of today's research settings usually involve relatively small sample sizes and sometimes broad ranges on continuous moderators. It is uncommon and often unfeasible to sample hundreds of observations for all different values of a conceivable moderator variable such as age. Thus, the model needs to be approximated for single values of

[1] Note that different ways of identifying the common factor are possible—for example, by setting the loading of one indicator to 1 and specifying the intercept to be 0 (Bollen, 1989).

the moderator. MGMCS, LSEM, and MFA are approaches that can be used to approximate the model given in Equation (2).

In his recent work, Molenaar (2015) set up a taxonomy of latent trait models for continuous and categorical moderator variables. This taxonomy is helpful for integrating the approaches discussed in the present article. If the data contain continuous trait indicator variables, and the moderator is an observed categorical one, the resulting latent trait model that can be used to consider moderation effects is a multiple-group factor model (see the MGMCS analyses in the present article). For the same continuous indicator variables, when entered in a factor model (note, however, that the parameter may vary across a continuous observed moderator), the resulting model is a moderated factor model (see the MFA analyses in the present article). An additional model category, again for continuous factor indicators and a continuous latent moderator variable, was introduced as the heteroscedastic factor model by Molenaar (2015). Tucker-Drob (2009) proposed latent interaction models (so-called latent moderated structures; Klein & Moosbrugger, 2000) for this category to investigate the age- and ability-related differentiation–dedifferentiation of intelligence. In our empirical example, the moderator (age) was an observed variable. Thus, we used MFA to illustrate factor-model-parameter variation across age. Furthermore, we will introduce a complementary modeling approach to MFA, the so-called LSEM, and propose it as a valuable exploratory method that offers advantages over MFA and that can be used to study variation in factor model parameters across a continuous observed moderator variable.

Thus, the aim of this article is (a) to assess and discuss the pros and cons of three competing, but also in some cases, as we will show, complementary, analytical approaches (MGMCS, LSEM, and MFA); (b) to further formalize LSEM as compared with descriptions in previous work (Hildebrandt et al., 2009) and investigate its statistical properties using an analytical derivation and a simulation study; and (c) to use three different methods to investigate whether differentiation occurs in our empirical example. Furthermore, we provide code for the easy implementation of LSEM.

Multiple-group mean and covariance structure analyses

In MGMCS analyses, the continuous moderator variable *age* is collapsed into a set of discrete age groups. For each group a_g ($g = 1, \ldots, G$), the common factor model is based on the following decomposition of the age-group-specific covariance matrix:

$$\Sigma(a_g) = \Lambda(a_g)\Lambda(a_g)^T + \Psi(a_g), \quad (3)$$

where $\Lambda(a_g)$ is a column vector of age-group-specific loadings for a specific age group a_g and $\Psi(a_g)$ is an $I \times I$ matrix of age-conditional error variances and covariances (usually assumed to be diagonal). The one-factor model for each indicator is then given for each age group a_g as follows:

$$Y_{ia_g} = \nu_i(a_g) + \lambda_i(a_g) \cdot \eta_a + \varepsilon_{ia_g}, \quad (4)$$

where the common factor η in the reference group is identified by setting the mean in each group to 0 and its variance to 1. The intercepts $\nu_i(a_g)$ and residual variances $\psi_i(a_g) = Var(\varepsilon_{ia_g})$ are freely estimated in each group. In addition, the residuals are assumed to be uncorrelated. The age differentiation hypothesis can then be tested in a two-step procedure (Little, 1997). In a first step, all model parameters (intercepts, factor loadings, and residual variances) in MGMCS analyses are allowed to differ across groups. A second step is applied to test whether the model fit deteriorates if all factor loadings are set to equality across the G age groups; that is, $\Lambda(a_1) = \ldots = \Lambda(a_g) = \ldots = \Lambda(a_G)$. This condition is known as weak factorial invariance in the context of measurement invariance tests (Meredith, 1993; Widaman, Ferrer, & Conger, 2010). By fixing the loadings to be equivalent across groups, factor variances can be freely estimated in all groups except for the reference group (Little, 1997). Note that model testing with MGMCS analyses is generally used as a test of whether a parameter varies across the groups a_g. However, the MGMCS analysis framework is flexible enough to implement cross-group constraints to investigate specific parametric variations in factor model parameters across groups (see Tucker-Drob & Salthouse, 2008).

It is important to note that if observed data are scarce at single values of the moderator, researchers often collapse observations into groups that span relatively wide intervals of the moderator. In most cases, these categorical boundaries are arbitrary. Because the observations within a group are treated as equal with respect to the variance-covariance matrix of the indicators, it is obvious that categorization will lead to a loss of information and may result in only a poor approximation of the model of interest given in Equation (2).

Local structural equation models

Instead of grouping participants who fall within a given range of the moderator as in MGMCS, in local structural equation models (LSEMs), observations are weighted around *focal points* (i.e., specific values of the continuous moderator variable). For every focal point, SEMs are sequentially estimated on the basis of weighted samples of observations. Ideally, SEMs are estimated in steps that are as narrow as possible on the scale of the continuous variable. How narrow these steps can be depends on the available sample size but also on the weighting function that is applied. In principle, there are many suitable weighting functions (for examples, see Fox, 2000; Wu & Zhang, 2006). For example, in the present study, each whole year—the granularity with which age was recorded—is an appropriate focal point.

Why the sample size at target focal points has to be considered before choosing the weighting function and its bandwidth can be illustrated as follows: Consider a study in which 15 observations were sampled at each focal age point measured in years, ranging from 4 to 23. By using weights for observations below and above the focal point, the effective sample size (N at focal points, achieved by weighting; see the following for elaborations) for LSEMs may reach a number that is generally suitable for estimating SEMs. However, the available N at the focal points is just one aspect to be considered. Decisions about how narrow the steps should be—if not restricted by the available sample size—are also guided by the researcher's considerations of how smooth the estimations of the parameter plots should be across the moderator.

A Gaussian kernel function can be used to weight observations around focal points (Gasser, Gervini, & Molinari, 2004; see also Hildebrandt et al., 2009; Hülür et al., 2011). The assumption behind this weighting procedure is that observations that are close to each other on a continuous scale are more similar than more distal observations. For every focal point (in our case, age in years), participants with that specific focal value receive the highest weight, whereas observations farther away from the focal point receive lower weights. Therefore, each estimated model is most strongly influenced by focal observations but is also influenced by observations near the focal point and influenced less by observations farther from the focal point. Repeating this procedure across the entire range of the moderator variable provides an approximation of the results that would occur if separate SEMs were fit to each focal point. As the name *Gaussian* suggests, this weighting procedure is based on the standard normal density function. The resulting weights are normally distributed around each focal point (see Figure 2). Observations at focal points receive a weight of 1 (see the focal points for ages 10, 15, and 22 in Figure 2, represented by small triangles, and the dashed vertical lines tracking the sizes of those points' weights). Observations with moderator values higher or lower than the focal point values receive weights smaller than 1. Because the normal density function has no limits on its lower and upper sides, all observations enter all models at each focal point, but observations that are far away from the focal points have very small values (below 0.01) as shown in Figure 2.

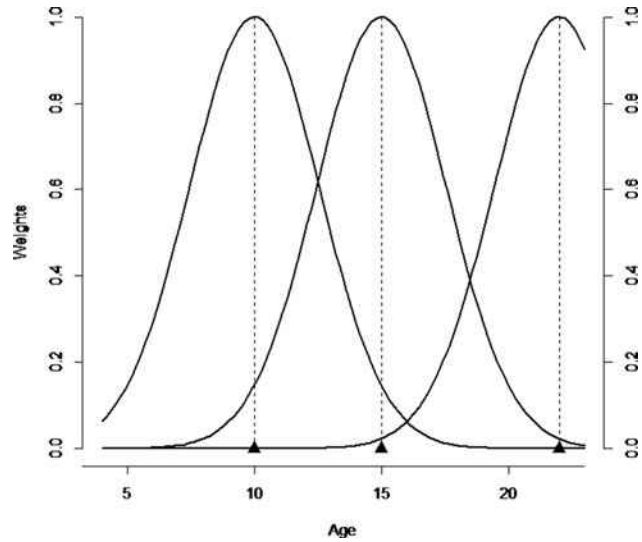

Figure 2. Gaussian kernel weighting functions for three selected focal age points used in LSEM. For the clarity of the illustration, only the courses of the weights for focal ages 10, 15, and 22 are shown. Focal age points are marked with small triangles along the *x*-axis, and the size of the weighting at the focal points (see the value of 1 on the *y*-axis) is tracked with straight dashed lines.

Computing weights for LSEM

Weights for LSEM are computed by the following formulas. The bandwidth (bw) is calculated by the formula

$$bw = \frac{h \cdot SD_A}{\sqrt[5]{N}}, \tag{5}$$

where h is the bandwidth factor; SD_A denotes the standard deviation of the moderator variable (A for age in our example); and N is the size of the total sample being analyzed. The bandwidth thus represents the standard deviation of the normal density function around each focal value of the moderator (denoted as a_0 in our example that uses age as a moderator). For example, if a bandwidth factor of $h = 2$ is chosen, observations farther than 2 times the bandwidth from the focal point receive only very small weights. Different bandwidth factors can be used: The factor $h = 1.1$ has been proposed in the nonparametric density estimation literature (e.g., Silverman, 1986), but simulation studies are needed to determine the optimal bandwidth for the application of LSEM to data that are characteristic of psychological research (see the simulation study that follows).

For each observation, a z value is calculated:

$$z(a, a_0) = \frac{a - a_0}{bw}. \tag{6}$$

Weights are then calculated using the Gaussian kernel function:

$$K(z) = \exp(-z^2/2). \tag{7}$$

As z values increase, the resulting weights get smaller. Focal observations receive the largest weight because they will have z values of 0. At a z value of 0, the density function reaches its maximum of $K(0) = 1$. The weights that vary between 0 and 1 are then calculated as follows:

$$W(a, a_0) = K(z(a, a_0)). \quad (8)$$

In general, the larger the bandwidth is, the smoother the resulting parameter function (the plot of the estimated model parameter along age) will be.[2] Thus, the bandwidth is essentially a smoothing parameter.

It should be mentioned that the MGMCS analysis is essentially based on a weighting scheme in which several focal points along the scale of the moderator—as many as are included in one group in the MGMCS analysis—are fully weighted, and weights of 0 are allocated to all other observations. Thus, the case in which MGMCS is fit to all of a moderator's focal points could be considered a special case of LSEM. However, comprehensive data sets that allow MGMCS modeling at all focal values of a continuously measured moderator are extremely rare. Thus, MGMCS usually includes many values of a moderator in one group and considers observations at different focal points within one group to be equal.

In LSEM, because the Gaussian kernel function utilizes the observations around a focal point, the number of observations for that model is higher than it would be if only the observations with that exact focal age value were included (see Figure 2). This is beneficial if the number of observations that are available is small somewhere in the distribution of the continuous moderator variable. See Table 1 for an example of how a model's sample size changes as a function of its weights. Usually weights are not standardized to the sum of N in nonparametric regression analyses and we follow the standard procedure practiced in this literature (e.g., Wu & Zhang, 2006). The effective N provides an estimate on how informative is the sample size at every focal point.

Note that focal points in the middle of the range of the moderator receive substantial weights from both lower and higher values, whereas focal values at the boundaries of the range of the moderator receive weights that are considerable in their magnitude only from either higher or lower values (see the focal value of 22 at the right boundary in Figure 2). This occurs because observations are de facto not available to the left versus to the right of the moderator values at the two boundaries, and consequently, there is no observation to be weighted by. Thus, the models estimated for focal points at the boundaries of the range of the moderator incorporate observations only with focal values that are either larger or smaller than the focal value of interest. When observations are scarce and LSEM is used to estimate local models, results at the left and right edges of the parameter functions should be interpreted with caution.

In our example, although not fully weighted, participants whose ages differ from the focal age point are entered into the models in an LSEM approach. As a consequence, there is variance in age in each of the estimated local models. When inspecting, for example, weights around the focal age of 10 in the current sample (see Figure 2), it can be seen that observations 3 points below and above the focal age are weighted notably above 0. However, if age-related changes in indicator means are present, the loadings and residual variances for an individual focal point will implicitly reflect these age trends. As the model in Equation (2) is aimed at modeling only the conditional covariance structure (conditioned on age), estimates of factor loadings will be distorted by indicator-specific age trends. The distortion will be due to mean-induced covariation between indicators and the moderator (Figure 3 shows the nonlinear relations between age and two broad abilities chosen for illustrative purposes).

Therefore, if such trends are present, all indicators should be detrended with regard to the moderator (i.e., age) and potentially the squared value of the moderator to control for linear and quadratic mean trends. The detrending needs to be conducted locally (Fan & Gijbels, 1996), before the LSEM analyses, and at the level of the observed data, using the same weighting function that will subsequently be used for conducting LSEM. This

Table 1. Calculation of the effective N at focal age 10 ($h = 2$).

No. of observations $N = 7$	Age $SD_A = 1.11$	Only weighting focal observations	LSEM weights based on Gaussian kernel function
1	9	0	.87
2	9	0	.87
3	10	1	1
4	10	1	1
5	11	0	.87
6	11	0	.87
7	12	0	.56
Sum of observations = Effective N		2	6.04

Note. In traditional analyses, only focal observations would receive full weights (see the third column of the table). In LSEM, based on the Gaussian kernel function, proximal observations, albeit not fully weighted, are also included in the model calculations at focal points. The sum of the weights equals the effective sample used in LSEM; h = bandwidth factor.

[2] The reason for this is that as the bandwidth increases, the overlap between the highly weighted observations increases. In each model, at a focal point, the overlap between the sample and neighboring samples becomes larger, and as a consequence, the parameter estimates of the different models become more similar. If the bandwidth approximated infinity, then all observations would be fully weighted in every model. All parameter plots would consequently show straight lines because the parameter estimates would be identical in every model. By contrast, a bandwidth that approximates 0 would use a weighting function that gives full weight to observations at that particular focal point while giving all other observations a weight of 0.

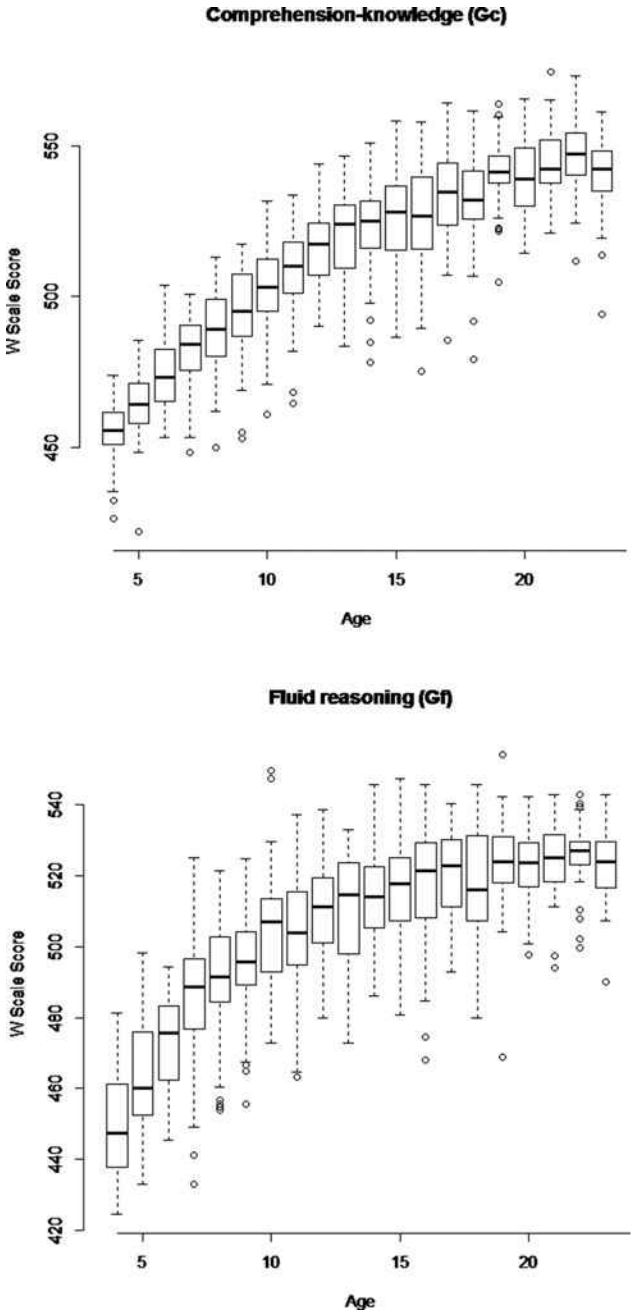

Figure 3. Boxplot showing the relations between two exemplary broad abilities (Gc and Gf) and age.

detrending procedure is also known as local polynomial smoothing (see Fan & Gijbels, 1996).

Following a reviewer's suggestion, we provide an analytical derivation of the statistical behavior of LSEM using techniques from nonparametric statistics. The Appendix shows that LSEM estimates of loadings and residual variances are asymptotically normally distributed if the bandwidth factor h is chosen proportional to $N^{-1/5}$. On the basis of this finding, it can be concluded that estimates are consistent and asymptotically unbiased. The proof of asymptotic normality relies on two facts. First, using nonparametric estimation techniques, a conditional covariance matrix $\hat{\Sigma}(a)$ is calculated, and it is shown that $\hat{\Sigma}(a)$ provides a consistent estimate of $\Sigma(a)$ (Yin, Geng, Li, & Wang, 2010). Second, $\hat{\Sigma}(a)$ is plugged into a confirmatory factor analysis from which the estimates of the parameter functions (for loadings and residual variances) are obtained. Using the implicit function theorem and applying the delta method (Wasserman, 2004), one can derive the asymptotic normality of the distribution for these parameter functions.

Permutation test for assessing the significance of parameter variations in LSEM

With LSEM, it is not possible to conduct a direct inferential test of the magnitude of variations in model parameters as in traditional model comparisons that are based on the likelihood function in MGMCS. However, parameter variations can be tested for statistical significance by implementing a permutation test (Good, 2005; see Briley, Bates, Harden, & Tucker-Drob, 2015; Hülür et al., 2011; Liu et al., 2015, for applications). More specifically, this is the test of the null hypothesis that a parameter γ (e.g., factor loading, residual variance) is constant across values of a continuous moderator variable.

For this test, first, a large number of permutation data sets are generated. Each of these data sets contains all variables from the original data and a permuted version of the original moderator variable. The permuted variable is obtained by repeatedly and randomly allocating values of the moderator variable to all observations. It can thus be assumed that in each of the permutation data sets, there is no systematic relation between the permuted version of the moderator variable and any parameter from the estimated SEM.

In a first step, for each of the permutation data sets, the series of locally weighted models are estimated on the basis of the permuted moderator variable. Then, in the next step, the parameter estimates from each permutation data set are used to test the null hypothesis that the parameter γ is constant across the moderator. Before calculating the test statistic, a weighted average for the parameter function needs to be computed with the formula

$$\bar{\gamma} = \sum_{k} \gamma\left(a_k\right) \cdot w\left(a_k\right), \tag{9}$$

where k runs across the set of defined focal values on the moderator; $\gamma\left(a_k\right)$ is the parameter estimate at the focal value a_k of the moderator; and $w\left(a_k\right)$ is a vector of weights that sum to 1. If there is equal weighting and 10 focal values on the moderator variable, the weights will each take a value of 1/10. The test statistic of the permutation test now quantifies how strongly $\gamma\ (a_k)$ varies around the average of the parameter function $\bar{\gamma}$, and is written as

$$SD_\gamma = \sqrt{\sum_k (\gamma(a_k) - \bar{\gamma})^2 \cdot w(a_k)}. \tag{10}$$

The distribution of SD_γ across the permutations is then used to test the null hypothesis that $\gamma\ (a)$ is constant across the continuous moderator a (age in our case). Given a certain significance level (e.g., $\alpha = .05$), the null hypothesis is rejected if the empirical value of SD_γ is greater than the corresponding percentile of the permutation distribution (e.g., 95th percentile).

Alternatively, pointwise hypotheses can be tested. For example, the hypothesis that a parameter γ at a specific focal point a_k does not deviate from the average of the parameter function $\bar{\gamma}$ could be tested by computing the following test statistic:

$$T_{\gamma\,(a_k)} = \gamma(a_k) - \bar{\gamma}. \tag{11}$$

Again, the null hypothesis is rejected if the empirical value of $T_{\gamma(a_k)}$ is greater than the corresponding percentiles of the permutation distribution (e.g., 2.5th and 97.5th percentiles). However, when testing several pointwise hypotheses, researchers need to consider the risk of an increase in the Type I error rate (e.g., Maxwell & Delaney, 2004). The test statistic of the permutation test could also be adapted to test more complex and informative hypotheses, including specific contrasts.

Moderated factor analysis

Moderated Factor Analysis (MFA; e.g., Bauer & Hussong, 2009; Curran et al., 2014) is an alternative approach that allows for the examination of factor model parameters along an observed continuous variable and is less exploratory than LSEM. As mentioned previously, in the taxonomy suggested by Molenaar (2015), latent trait models that are moderated by an observed continuous variable are subsumed under the category MFA. LSEM, as has been described, can be seen as a special, exploratory modeling approach in the MFA model category.

The main idea of MFA is that the factor model parameters (loadings, intercepts, and residual variances) are allowed to vary across a set of between-person covariates. In our case, age is the between-person covariate, and each parameter varies deterministically as a parametric function of age. To illustrate the use of MFA in our example data, we allow each factor model parameter to vary as a quadratic function of age. More specifically, in the factor model $Y_{ia} = \nu_i(a) + \lambda_i(a) \cdot \eta_a + \varepsilon_{ia}$, the factor loading of an indicator is predicted by the following relation:

$$\lambda_i(a) = \lambda_{i0} + \lambda_{i1} \cdot a + \lambda_{i2} \cdot a^2, \tag{12}$$

where the parameter λ_{i1} represents the linear regression weight of age (a); λ_{i2} represents the quadratic regression weight of age; and λ_{i0} is the regression intercept. The interpretation of the model parameters is similar to that of a regression analysis involving quadratic terms (Cohen, Cohen, West, & Aiken, 2003). Positive values indicate that as age increases, the factor loadings for the broad abilities also increase. As age was centered, the regression intercept λ_{i0} can be interpreted as the expected loading at the mean of the variable age. Similarly, the intercepts of the broad abilities also vary as a function of age:

$$\nu_i(a) = \nu_{i0} + \nu_{i1} \cdot a + \nu_{i2} \cdot a^2, \tag{13}$$

where the parameters ν_{i0}, ν_{i1}, and ν_{i2} are the regression coefficients. In addition, the residual variances are allowed to vary as a function of age:

$$\psi_i(a) = \exp[\psi_{i0} + \psi_{i1} \cdot a + \psi_{i2} \cdot a^2], \tag{14}$$

where the parameters ψ_{i0}, ψ_{i1}, and ψ_{i2} are the regression coefficients of a log-linear relation given that the variances are bounded by zero.

In the original formulation, the MFA model parameters vary deterministically as a parametric function of the moderator variable *age*. However, in a natural extension of MFA, the variation that is explained by the moderator variable is separated from the unexplained variation for each model parameter. This can be done by including a residual term in the regression equation for each factor model parameter. More specifically, for each loading, intercept, and residual variance, Equations (12), (13), and (14) are extended as follows:

$$\lambda_i(a) = \lambda_{i0} + \lambda_{i1} \cdot a + \lambda_{i2} \cdot a^2 + u_{\lambda ia} \tag{15}$$

$$\nu_i(a) = \nu_{i0} + \nu_{i1} \cdot a + \nu_{i2} \cdot a^2 + u_{\nu ia} \tag{16}$$

$$\psi_i(a) = \exp[\psi_{i0} + \psi_{i1} \cdot a + \psi_{i2} \cdot a^2 + u_{\psi ia}], \tag{17}$$

where $u_{\lambda ia}$, $u_{\nu ia}$, and $u_{\psi ia}$ are uncorrelated and normally distributed residuals with zero means and indicator-specific variances $\tau^2_{\lambda i}$, $\tau^2_{\nu i}$, and $\tau^2_{\psi i}$. The indicator-specific residual variances describe the variation in a factor model parameter that is not explained by the moderator variable. For example, a large residual variance $\tau^2_{\lambda i}$ for the loading of indicator i implies that only a small part of the variation in the factor loading of indicator i across age can be attributed to the linear and quadratic effects of age.

MFA can be considered a parametric version of the nonparametric LSEM approach because MFA specifies (parametric) regression equations to explain variation in factor model parameters. In this respect, MFA is a more confirmatory approach than LSEM because the functional relation between the factor model parameter and the moderator needs to be specified.

Research question for the empirical illustration

The aim of the empirical illustration was to investigate whether the three modeling approaches described, and specifically LSEM, would lead to the same substantive conclusions regarding the moderating effects of age on the parameters of the one-factor measurement model of cognitive abilities. We showed that the statistical model specified in Equation (2) can be approximated with MGMCS, LSEM, or MFA. Throughout, we chose to use the variable age as an example of a moderator variable because age can be measured with varying levels of resolution and because research applications with age as a moderator are frequently found in the literature.

With our example, we aimed to illustrate a realistic scenario that frequently occurs in empirical studies in which data are rather scarcely sampled along a continuous moderator. Specifically, we aimed to illustrate the sensitivity of the three analytical approaches to detect differentiation between cognitive abilities across childhood. Thus, we estimated the moderating effect of age in MGMCS, MFA, and LSEM using a subset of the available data ($N = 1{,}087$) from the normative sample of the Woodcock-Johnson III (WJ-III) Tests of Cognitive Abilities (Woodcock, McGrew, & Mather, 2001, 2007).

In addition, we conducted a small simulation study with two objectives. First, this study can be considered a first step in evaluating the statistical properties of the LSEM approach. Second, we use this study to provide researchers with further guidance for selecting the bandwidth factor h for calculating sample weights, a central part of the LSEM approach.

Data example

Indicators

Figure 1 illustrates a general cognition factor indicated by the seven broad cognitive abilities. Table 1S (in the Supplementary Material) provides a summary description of the measures and ability scores. *Comprehension-knowledge* and *Fluid reasoning* hereby correspond to crystallized and fluid intelligence, respectively. Each of the WJ-III second-stratum broad cognitive abilities is a cluster score obtained by averaging two W-scaled first-stratum cognitive ability tests (see Table 1S). Thus, each of the two W scales contributes equally to the cluster score. The W scale is a transformation of the Rasch model (Woodcock, 1999), which provides a common scale for the ability of a person and the difficulty of a task. The test's W scales are each centered on a value of 500, which constitutes the approximate average performance of 10-year-olds (McGrew et al., 2007).

Sample

The WJ-III normative sample consists of four cross-sectional subsamples (preschool, school, college, and adult subsamples) that are each representative of the U.S. population (McGrew et al., 2007). For the age range of 4 to 23, the original data set comprised $N = 5{,}475$ participants; 50.78% female; $M_{\text{age}} = 11.75$; $SD_{\text{age}} = 5.18$. We selected this age range because it is broad enough to detect potential differentiation effects that would be expected according to the age differentiation theory of cognitive abilities. To illustrate a scenario that more frequently occurs in empirical studies, we selected a smaller data set from this original sample for our illustrative analyses. This sample more closely reflects the prevalent sample sizes in the field. Thus, we randomly selected 40%–60% (depending on the density of the sampling in the original sample) of the available observations at each focal age point between 4 and 23 by completely excluding observations with missing values on any indicator. This resulted in focal age samples with the following numbers of persons: $N_{\text{A}=4} = 46$; $N_{\text{A}=5} = 49$; $N_{\text{A}=6} = 57$; $N_{\text{A}=7} = 63$; $N_{\text{A}=8} = 69$; $N_{\text{A}=9} = 67$; $N_{\text{A}=10} = 78$; $N_{\text{A}=11} = 72$; $N_{\text{A}=12} = 67$; $N_{\text{A}=13} = 64$; $N_{\text{A}=14} = 64$; $N_{\text{A}=15} = 52$; $N_{\text{A}=16} = 67$; $N_{\text{A}=17} = 56$; $N_{\text{A}=18} = 53$; $N_{\text{A}=19} = 43$; $N_{\text{A}=20} = 47$; $N_{\text{A}=21} = 39$; $N_{\text{A}=22} = 29$; $N_{\text{A}=23} = 30$. It is obvious that the sample sizes at several of the focal age values were too small to estimate a separate model for that focal age point, thus illustrating the necessity of an LSEM approach. For the MGMCS analysis, we merged these samples into six groups with the following age ranges: 4–6 ($N_{\text{A}=4\text{–}6} = 152$), 7–9 ($N_{\text{A}=7\text{–}9} = 199$), 10–12 ($N_{\text{A}=10\text{–}12} = 217$), 13–15 ($N_{\text{A}=13\text{–}15} = 197$), 16–18 ($N_{\text{A}=16\text{–}18} = 176$), and 19–23 ($N_{\text{A}=19\text{–}23} = 188$) years. We applied MFA to the entire random sample of $N = 1{,}129$ participants.

Software and model estimation

The MGMCS and LSEM approaches were estimated in R (R Development Core Team, 2008) using the structural equation modeling packages `lavaan` (Rosseel,

2012) and `lavaan.survey` (Oberski, 2014). For the LSEM approach, we programmed a wrapper function `lsem.estimate` (see the Supplementary Material) in the `sirt` package (Robitzsch, 2015). This function conducts all the data management and analysis steps (e.g., calculation of weights, model estimation in lavaan, management of output files, plots of parameter functions). The `lavaan.survey` package (Oberski, 2014) is also applied in the `lsem.estimate` wrapper function because the structural equation model in lavaan has to be extended to weighted data when estimating LSEM. An additional R function `lsem.permutation` performs the permutation tests (see the Supplementary Material for the R code). The number of permutations for testing the global and pointwise hypotheses was set to 1,000.

The parameters of the MFA were estimated using a Bayesian approach. We used the WinBUGS software (Windows version of Bayesian inference using Gibbs sampling; Spiegelhalter, Thomas, Best, & Lunn, 2003; see also Lunn, Jackson, Best, Thomas, & Spiegelhalter, 2012), which is a flexible program for Bayesian analyses of statistical models (see the Supplementary Material for the WinBUGS code). Diffuse or vague prior distributions were assigned to each model parameter. For the coefficients of the regressions predicting the loadings, intercepts, and residual variances, flat normal prior distributions with mean 0 and variance 1,000 were specified. Following recent recommendations for variance parameters in Bayesian analyses of hierarchical models (Gelman & Hill, 2007; Jackman, 2009), uniform distributions for standard deviations were used as prior distributions for the residual variation of the regressions. More specifically, the standard deviations of the residuals of the regressions were assumed to be uniformly distributed across the range (0, 100). Given the standardized metric of the W scales ($M = 500$, $SD = 50$), we think that the assumption regarding the standard deviation of the residuals that has been expected to be in the range 0 to 100, provided no information that was relevant to the inference. In WinBUGS, Markov chain Monte Carlo (MCMC) techniques are used to approximate the posterior distributions of the model parameters. We specified an MCMC chain of 5,000 iterations with a burn-in period of 2,500. To assess the convergence behavior of the MCMC algorithm, we inspected the trace plots of the univariate chains for each parameter. In addition, the Rhat statistic (Gelman & Hill, 2007) was calculated by dividing the MCMC chain into three subchains. The Rhat statistics of all parameters were smaller than 1.05, thus indicating sufficient convergence of the MCMC algorithm. The mean and standard deviation of the posterior distribution were used as the point and standard error estimates of the corresponding model parameter. A Wald test based on the MCMC output was specified to conduct single- and multiple-parameter tests.[3]

We decided to use the WinBUGS software and MCMC techniques to estimate the parameters of the MFA model for two reasons. First, the WinBUGS software is extremely flexible in its model specification facilities, and the user has complete control over the ability to define the parameter to be estimated. Second, WinBUGS and MCMC techniques can generally be used with diffuse or noninformative priors, and in this case, the results can be expected to be similar to those achieved by maximum likelihood estimations (Gelman et al., 2014). However, the use of MCMC techniques in a frequentist interpretation has the great advantage of flexible model specification. In our case, residual random effects for intercepts, loadings, and variances could be easily added to the model (see Equations [15], [16], and [17]). Another advantage of implementing MFA using WinBUGS is that WinBUGS is also freely available and easily accessed via R. Please note that the MFA model with age-specific residuals (see Equations [12], [13], and [14]) can be specified in widely used latent variable software (M*plus*, OpenMx) by using definition variables to define how factor model parameters are moderated by a continuously measured observed variable (see Cheung, Harden, & Tucker-Drob, 2015).

Results

We report unstandardized parameter estimates. This facilitates comparisons between the three approaches. Furthermore, the comparison of factor loadings needs to be based on unstandardized parameters because standardization is influenced by variable (residual) variances. The parameter functions obtained by applying MGMCS, LSEM, and MFA for all broad abilities are displayed in Figure 4.

Multiple-group mean and covariance structure analyses

In a first step, all model parameters (intercepts, factor loadings, and residual variances) were allowed to differ across the six age groups described. For this model, the comparative fit index (CFI) was .98, and the root mean square error of approximation (RMSEA) was .07. Browne and Cudeck (1993) suggested that RMSEA values of less

[3] First, from the output of the MCMC chain, a k-dimensional estimate $\hat{\delta}$ and its covariance matrix V can be calculated. For large sample sizes and noninformative prior distributions, statistical inference that is based on the posterior distribution (obtained from the MCMC output) approximates maximum likelihood inference (Gelman et al., 2014; see also Walker, 1969). Using the asymptotic multivariate normality of $\hat{\delta}$, the chi-square statistic $\chi^2 = \delta^{T} V^{-1} \delta$ of the Wald test for the null hypothesis $\delta = 0$ is formed; it is asymptotically χ^2 distributed with k degrees of freedom.

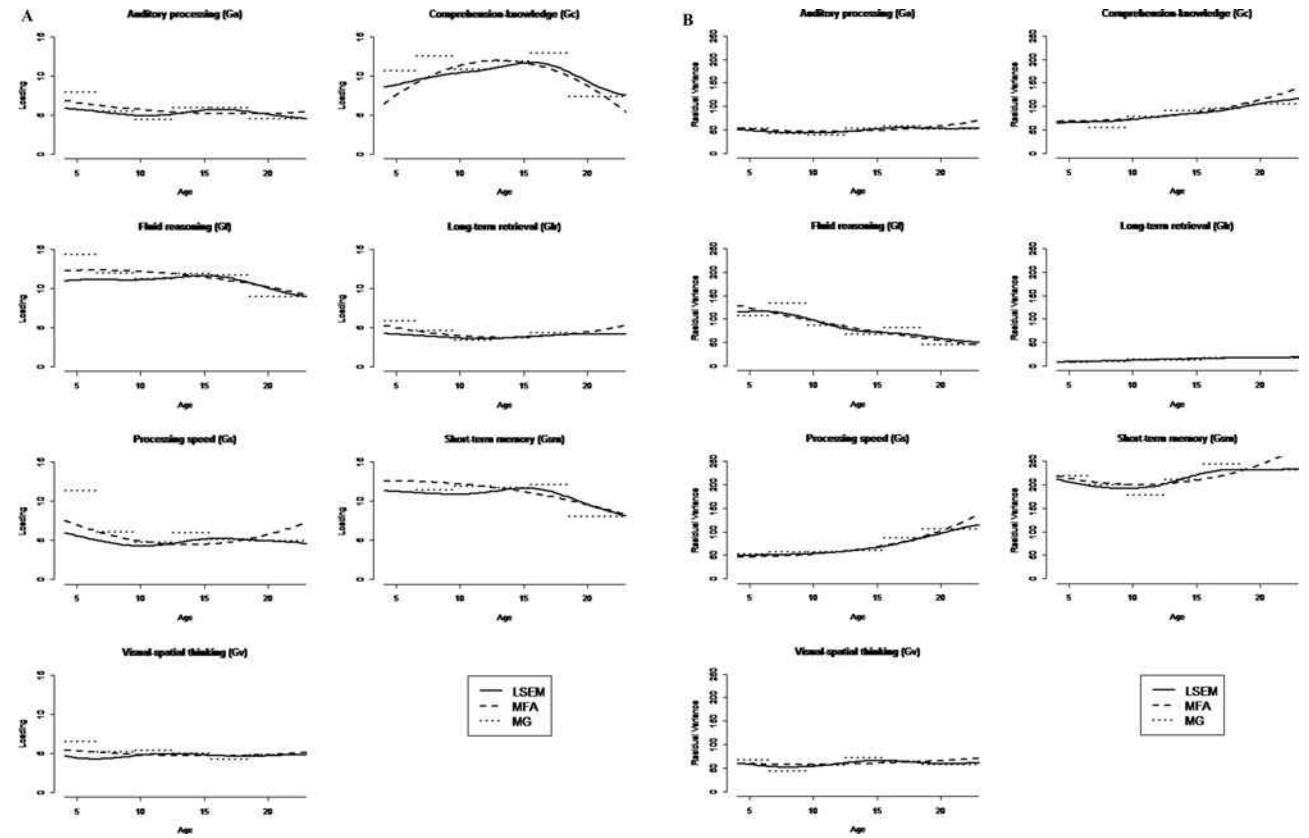

Figure 4. Parameter functions comparing the LSEM, MFA, and MG(MCS) results. Panel A depicts the functions for factor loadings; Panel B depicts the functions for residual variances of all investigated broad abilities.

than .05 should be considered to indicate good fit, and values between .05 and .08, fair fit. Hu and Bentler (1999) proposed that CFI values higher than .95 should be considered to indicate good fit. The fit of the model was $\chi^2(84) = 155.76$ (p < .001). We conducted model fit comparisons for factor loading variation in reference to this model. We first set the factor loadings of all broad abilities to equality across groups. As previously described, fixing the loadings to equality across groups allows to freely estimate the variance of *g* in all, but not the reference group. Thus, the latent variable is standardized in the first group, but it is freely estimated in all other groups (see Little, 1997). Fixing the loadings to equality and freely estimating the factor variances led to a significant deterioration in model fit ($\Delta\chi^2 = 99.46$, $\Delta df = 30$, $p < .001$). If additionally the variance of *g* is kept to be equal across groups in the model with invariant factor loadings, the likelihood ratio test in comparison with the reference model with invariant *g* variance across groups but different factor loadings, yielded the following result: $\Delta\chi^2 = 115.64$, $\Delta df = 35$, $p < .001$. Testing the cross-group equivalence of the *g* variance given equal loadings revealed that the variance of *g* significantly differs across groups: $\Delta\chi^2 = 16.17$, $\Delta df = 5$, $p < .001$. Thus, constraining the model to group-invariant loadings leads to differentiation that is manifested in nonequivalent *g* variance.

We then successively set the factor loadings to equality across groups—only the loadings of one indicator at a time—and compared each model separately with the unrestricted one. The restrictions of the single factor loadings yielded five more degrees of freedom as compared with the unrestricted model. Restricting the factor loadings led to a significant deterioration in model fit in all cases (*Comprehension-knowledge*, Gc, $\Delta\chi^2 = 11.19$, $\Delta df = 5$, $p < .05$; *Processing speed*, Gs, $\Delta\chi^2 = 45.58$, $\Delta df = 5$, $p < .05$; *Long-term retrieval*, Glr, $\Delta\chi^2 = 24.76$, $\Delta df = 5$, $p < .05$; and *Auditory processing*, Ga, $\Delta\chi^2 = 16.05$, $\Delta df = 5$, $p < .05$) with the exception of *Visual-spatial thinking*, Gv ($\Delta\chi^2 = 1.05$, $\Delta df = 5$, $p = .96$), *Fluid reasoning*, Gf ($\Delta\chi^2 = 5.45$, $\Delta df = 5$, $p = .36$), and *Short-term memory*, Gsm ($\Delta\chi^2 = 9.97$, $\Delta df = 5$, $p = .08$).

Parameter functions for all broad abilities are shown in Figure 4 by short dashed lines. Across age groups, a salient systematic trend reflecting a decrease in the parameters was triggered primarily by a decrease from Group 1 to Group 2 (ages 4–6 to 7–9) for all broad abilities except *Comprehension-knowledge* and *Short-term memory*. Thus, an MGMCS analysis would lead researchers to conclude that differentiation in cognitive abilities occurs in childhood (i.e., before adolescence) for four out of seven broad abilities.

Finally, as outlined earlier, differentiation could also be expected on the level of residual variances (Molenaar, Dolan, Wicherts, & van der Maas, 2010). To investigate differentiation in the residuals, pairwise models were again tested one by one. Models with fixed loadings for a given indicator were further restricted by introducing equality in the residual variances of the same indicator across groups. For each comparison, this again resulted in a difference of five degrees of freedom as compared with the model with only a fixed loading for a given indicator. Restricting the residual variances led to a significant deterioration in model fit for the following variables: *Fluid reasoning*, Gf, $\Delta\chi^2 = 29.71$, $\Delta df = 5$, $p < .01$; *Processing speed*, Gs, $\Delta\chi^2 = 20.95$, $\Delta df = 5$, $p < .01$; *Long-term retrieval*, Glr, $\Delta\chi^2 = 25.62$, $\Delta df = 5$, $p < .01$; and *Auditory processing*, Ga, $\Delta\chi^2 = 13.05$, $\Delta df = 5$, $p < .05$), but not for *Comprehension knowledge*, Gc ($\Delta\chi^2 = 10.60$, $\Delta df = 5$, $p = .06$), *Visual-spatial thinking*, Gv ($\Delta\chi^2 = 2.74$, $\Delta df = 5$, $p = .74$), or *Short-term memory*, Gsm ($\Delta\chi^2 = 1.04$, $\Delta df = 5$, $p = .96$).

Local structural equation models

The fit indices suggested a good fit for the locally estimated models according to the recommended cutoff scores mentioned previously. The average CFI value was $M = .958$ ($SD = .032$, Min = .836, Max = .988); the average RMSEA was $M = .075$ ($SD = .024$, Min = .044, Max = .144); and the average SRMR was $M = .031$ ($SD = .011$, Min = .019, Max = .064). With increasing age, a slight deterioration in model fit was suggested by the CFI and SRMR.

The parameter functions of the factor loadings depicted by straight solid lines in Figure 4 descriptively showed four different overall patterns: (a) increase-decrease—thus, dedifferentiation followed by differentiation (factor loadings for *Comprehension-knowledge*); (b) decrease-increase—thus, differentiation followed by dedifferentiation (factor loadings for *Auditory processing* and *Processing speed*); (c) straight-decrease—thus, stability followed by differentiation (factor loadings for *Fluid reasoning* and *Short-term memory*); and (d) straight—thus, continuous stability (factor loadings for *Long-term retrieval* and *Visual-spatial thinking*). As an inferential test of these descriptive patterns, we report the results of the permutation test for factor loadings as well as for residual variances in Table 2.

Only for the factor loading of *Comprehension-knowledge* did the permutation test reveal overall significant variation across age as compared with the weighted average Gc loading calculated with Equation (9) ($M = 10.25$, see Table 2). For the pointwise test as calculated with Equation (11), a more conservative significance criterion was applied ($\alpha = .01$). The pointwise tests showed that the estimated loadings at the focal ages 14 to 17 were significantly above the average loading (showing slight statistically significant dedifferentiation), whereas beginning at the age of 22, the loading of Gc was significantly lower than the average loading across age (showing slight statistically significant differentiation). The pointwise test for the variation in the Gc loadings is visually represented in Figure 5 with a p-value curve (see Silverman & Ramsay, 2005). Further pointwise tests for all other broad ability parameters are provided in the Supplementary Material.

Table 2 further shows that all other descriptively detected factor loading variations were not statistically significant overall or at specific age ranges or focal age values. However, LSEM detected significant variation in residual variances for all but two (*Visual-spatial thinking* and *Auditory processing*) broad abilities (see Figure 4 and the last four columns of Table 2). As Molenaar et al. (2010) suggested, differentiation and dedifferentiation may be detected in heteroscedastic residuals, as was also the case in the present data example. If the residual variances increase, the g-factor saturation of the broad abilities decreases. Slight differentiation then occurs as would be theoretically predicted for the age range we analyzed in our empirical illustration. Such an overall increase in residual variances, as compared with the average residual variance calculated with Equation (9), was salient for *Comprehension knowledge, Processing speed*, and *Long-term retrieval* (see Figure 4, Panel B). For *Fluid reasoning* and *Short-term memory*, however, heteroscedastic residuals estimated by LSEM revealed a slight dedifferentiation—thus, the opposite of what was theoretically predicted by the *differentiation hypothesis* for the age range that we analyzed.

To illustrate how the LSEM results depend on the selection of the bandwidth, we additionally provide an estimated loading function for *Comprehension knowledge* across age in Figure 6 with three different bandwidth factors (h) for calculating sample weights. Dotted lines track the loading in the condition in which $h = 3$. This trajectory is obviously very smooth and continuous, thus less sensitive to detecting change onsets. Dashed lines show the condition $h = 1.1$, which makes the loading functions wavy and most sensitive to noise. The solid lines depicting loading functions for $h = 2$ are not too smooth to miss change onsets in the plotted function but seem to smooth out noise.

Table 2. Permutation test for LSEM (conducted with $h = 2$).

	Loadings				Residual variances			
	M	*SD*	*p*(*SD*)	Pointwise sign. for age ranges	*M*	*SD*	*p*(*SD*)	Pointwise sign. for age ranges
Ga	5.27	0.33	0.36	—	48.41	6.27	.01	9
Gc	10.25	1.08	0.01	14–17; 22–23	81.21	16.67	.01	7–9; 20–21
Gf	11.03	0.61	0.20	—	84.44	24.85	.00	4–10; 14–23
Glr	3.90	0.21	0.25	—	13.51	3.25	.00	4–9; 16–22
Gs	4.87	0.40	0.29	—	66.39	20.98	.00	4–12; 17–23
Gsm	10.89	0.79	0.27	—	211.93	20.24	.00	—
Gv	4.64	0.20	0.84	—	58.24	8.03	.17	—

Note. Ga = Auditory processing; Gc = Comprehension-knowledge; Gf = Fluid reasoning; Glr = Long-term retrieval; Gs = Processing speed; Gsm = Short-term memory; Gv = Visual-spatial thinking; *M* = weighted average of the parameter function of interest (see Equation [9]); *SD* = test statistic of the permutation test (see Equation [10]); *p*(*SD*) = *p* value of the permutation test. For the pointwise tests, we set the significance level to .01. The choice of the bandwidth factor ($h = 2$) was justified by the results of the simulation study (see the section describing the simulation results).

Moderated factor analysis

Table 3 shows in detail the results obtained by MFA. Parametric functions estimated in MFA are depicted by long dashed lines in Figure 4. In accordance with LSEM, the factor loading of *Comprehension-knowledge* showed a linear increase and a quadratic decrease thereafter, with an overall moderation effect of $\chi^2 = 14.45$ ($df = 2$, $p < .01$). No further statistically significant moderation effects on factor loadings were salient according to MFA. In MFA, we also estimated the linear and quadratic moderation effects of the residual variances. These are also listed in Table 3. In line with LSEM, there was no overall moderation effect of the residual variances of *Auditory processing* and *Visual-spatial thinking*. One further moderation of a residual variance that had been salient in LSEM remained undetected by MFA—the parameter belonging to *Short-term memory* ($\chi^2 = 2.40$, $df = 2$, $p = .30$). Four residual variances linearly decreased across age (see also the long dashed lines in Figure 4). Whereas the residual variance of *Fluid-reasoning* decreased linearly, suggesting differentiation, the residual variances of *Comprehension-knowledge*, *Long-term retrieval*, and *Processing speed* increased linearly, suggesting dedifferentiation. These results agreed with the LSEM results. However, their quadratic counterparts, which became visible in the LSEM pointwise significance test that was based on the permutation test, remained undetected by MFA (see Table 3).

We will demonstrate how to interpret an MFA parameter (from Table 3) for the broad ability *Comprehension-knowledge*. The value of 11.99 for the estimated intercept is the expected loading of *Comprehension-knowledge* at

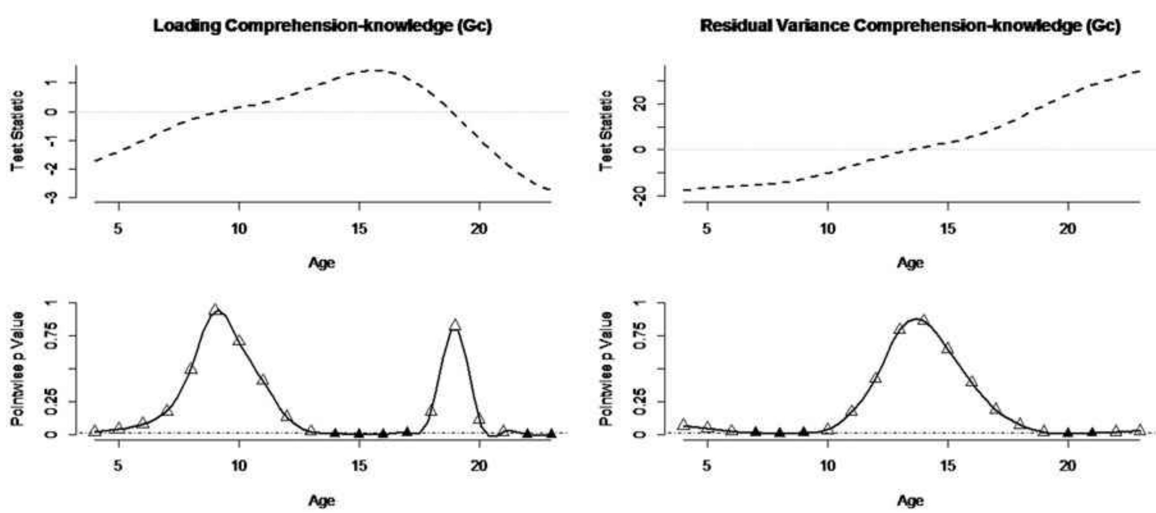

Figure 5. Pointwise *p*-value curves of estimated significance for parameter (loading—left; residual variance—right) variation across age in LSEM. The figure shows the significance tests conducted for Gc (*Comprehension-knowledge*). The upper figure represents the course of the test statistic across age on the basis of the permutation test calculated with Equation (11) for the factor loading of Gc. The lower figure shows the pointwise significance (below $p < .01$, see the *y*-axes in the figure) across age with solid triangles—the deviation of the estimated factor loading of Gc from its weighted average functions is significant between the ages of 14 and 17 and above the age of 22 (as also shown in Table 2). Equivalent plots for all other broad abilities are presented in the Supplementary Material.

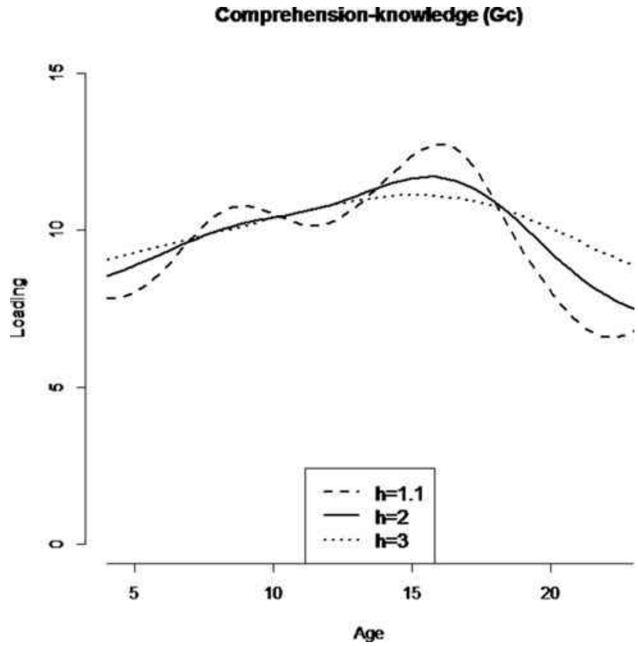

Figure 6. Estimated loading function for *Comprehension-knowledge* (Gc) across age with three different bandwidth factors for calculating sample weights.

$a = 0$ (i.e., $a = M_A$ because age is centered). The linear moderation effect, which is positive (0.01), shows an increase in the loading for a 1-unit increase in age. The quadratic decrease of –0.06 is expected for participants who are older than 12 (see also Figure 4, Panel A). Table 3 also shows the residual variances of the moderation effects of loadings and residual variances as formalized in Equations (15) and (17), along with their interval estimates (2.5th and 97.5th percentiles of the posterior distribution).

Comparing the LSEM and MFA results

LSEM showed that the *Comprehension-knowledge* factor loadings for focal ages 14 to 17 were significantly higher than the average loading computed for the entire age range and that the loading was significantly lower above the focal age of 22. This parallels the MFA finding of a positive but nonsignificant linear moderation effect of age and the negative quadratic moderation shown in Table 3 for *Comprehension-knowledge*. However, the analyses suggest that LSEM is more sensitive to detecting the change onsets due to the moderator.

MGMCS versus LSEM and MFA results

In MGMCS, restricting the factor loadings to equivalence across age groups led to a significant deterioration in model fit for three out of seven loading parameters. These results are different from the LSEM and MFA results, which suggested moderation of the loadings for only *Comprehension-knowledge*. However, the observed decrease in factor loadings between the estimates (generally for the first two age groups) and the relative stability thereafter (see Figure 4) were due to the indicator-specific age trends shown in Figure 3. These trends are specific for age-related changes in cognitive ability data in childhood. Because the MGMCS models are estimated for age groups, which specify particular age ranges, any remaining indicator-specific age trends within each age group will distort parameter estimates that are diagnostic of differentiation in the one-factor model, as presented in our example. LSEM and MFA, however, are more flexible and allow the researcher to control for these age trends across the range of the moderator variable. The results presented above emphasize the necessity of (a) studying mean trends

Table 3. Results of the moderated factor analyses (MFA): Estimates with WinBUGS.

	Intercept		Linear trend		Quadratic trend		*SD* of residuals / credibility interval			Test of moderating effects	
	Est.	*SE*	Est.	*SE*	Est.	*SE*	Est.	Q5	Q95	χ^2	*p*
Loadings											
Ga	5.315*	0.411	− 0.076	0.060	0.005	0.011	0.360	0.036	0.950	1.75	.42
Gc	11.988*	0.653	0.006	0.078	− 0.060*	0.016	0.681	0.090	1.610	14.45*	.00
Gf	11.947*	0.642	− 0.108	0.091	− 0.015	0.016	1.012	0.144	2.010	3.16	.21
Glr	3.737*	0.242	− 0.011	0.032	0.015*	0.006	0.238	0.050	0.510	5.37	.07
Gs	4.580*	0.507	− 0.065	0.075	0.027	0.014	0.733	0.105	1.640	4.85	.09
Gsm	11.698*	0.890	− 0.185	0.120	− 0.014	0.024	1.086	0.130	2.450	2.80	.25
Gv	4.809*	0.689	− 0.006	0.086	0.004	0.017	1.414	0.675	2.270	0.05	.98
Residual variances											
Ga	3.811*	0.092	0.006	0.013	0.003	0.002	0.198	0.073	0.340	2.72	.26
Gc	4.347*	0.113	0.040*	0.013	0.001	0.003	0.163	0.015	0.340	12.06*	.00
Gf	4.417*	0.092	− 0.056*	0.013	− 0.001	0.002	0.105	0.009	0.250	19.52*	.00
Glr	2.643*	0.088	0.053*	0.013	− 0.003	0.002	0.099	0.004	0.250	16.80*	.00
Gs	4.076*	0.071	0.052*	0.010	0.003	0.002	0.074	0.005	0.170	36.74*	.00
Gsm	5.310*	0.079	0.010	0.011	0.002	0.002	0.109	0.008	0.220	2.40	.30
Gv	4.055*	0.089	0.012	0.011	0.000	0.002	0.160	0.014	0.320	1.31	.52

Note. Ga = Auditory processing; Gc = Comprehension-knowledge; Gf = Fluid reasoning; Glr = Long-term retrieval; Gs = Processing speed ; Gsm = Short-term memory; Gv = Visual-spatial thinking; Est. = estimated parameter value; SE = standard error of the estimate; Q5 = 5% quantile; Q95 = 95% quantile. The χ^2 test has two degrees of freedom because the test of the moderation effects (see last two columns) tested linear and quadratic moderation simultaneously; significant effects ($p < .05$) are presented with asterisks.

at the level of indicators before investigating covariance structures because mean values may interact with a moderator and (b) adapting the flexibility of the model to control for such effects if age trends are apparent.

Simulation study: Evaluation of the LSEM approach

We conducted a simulation study to evaluate the statistical properties of the LSEM approach and to provide further guidance for the selection of the bandwidth factor h (see Equation [5]). The population model used to generate the data was the same as in the empirical example (one-dimensional model with seven indicator variables and a continuous moderator variable). We then slightly modified the parameter estimates of the empirical example to specify the parameters of the data-generating model (see the Supplementary Material, Tables 2S–5S, for a full list of the parameters used in the simulation). The distribution of the moderator variable (age) was fixed across each replication. That is, for each age value a_k, observed indicators Y_{ia_k} were generated by using Equation (2) and the loadings, intercepts, and residual variances from the empirical example. The sample sizes were manipulated to $N = 300$, 600, and 1,000. For each of the three sample size conditions, 1,000 data sets were generated. We analyzed each data set using LSEM with varying bandwidth factors ($h = 1.1, 1.3, 1.5, 1.75, 2, 2.3$, and 3). Two criteria (bias and root mean square error) were used to evaluate the statistical behavior of the LSEM estimates of the loadings and the residual variances. The pointwise (pw) bias for a factor loading at a specific focal value of an indicator was estimated as

$$\text{pwBias}(\hat{\lambda}_i(a_k)) = \bar{\hat{\lambda}}_i(a_k) - \lambda_i(a_k), \qquad (18)$$

where $\bar{\hat{\lambda}}_i(a_k)$ is the mean parameter estimate for a focal value a_k across the 1,000 generated data sets, and $\lambda_i(a_k)$ is the true population parameter. To assess the estimated bias for a factor loading across the age distribution, a weighted global (wg) bias was calculated as follows:

$$\text{wg Bias}(\hat{\lambda}_i) = \sum_{k=1}^{20} |\text{pw Bias}(\hat{\lambda}_i(a_k))| \cdot \ w(a_k), \qquad (19)$$

where $w(a_k)$ is a weight reflecting the fixed age distribution. Thus, the absolute pointwise biases at each focal value are averaged across the age distribution. The same definitions were used to estimate the pointwise and global biases for the residual variances. The pointwise mean square error (MSE), which combines bias and variability into an overall measure of accuracy, was estimated by taking the square of the mean square difference of the estimate and the true parameter at each specific focal value. Then the pointwise MSE estimates were averaged across the age distribution, and finally, the square root was taken to obtain a weighted global RMSE value.

Table 4 provides the estimated relative percentage pointwise bias for a selected loading (i.e., the pointwise bias was divided by the true pointwise parameter multiplied by 100). The percentage pointwise bias is shown as a function of the bandwidth parameter ($h = 1.1$, 2, and 3) and the sample size ($N = 300$, 600, and 1,000). Overall, the estimated relative percentage pointwise bias decreased when the sample size increased and the bandwidth was small. For a small bandwidth parameter ($h = 1.1$), the pointwise bias was below 5% at all focal values, with the exception of the boundary value of age $= 23$. Estimators with a relative percentage bias below 5% are often considered approximately unbiased (e.g., Boomsma, 2013). Thus, with a small bandwidth parameter and larger sample sizes, the LSEM approach provides approximately unbiased estimates of the factor loadings at specific focal points of the moderator. A similar picture emerged for the pointwise bias of the other six loading parameters (see Tables 6S–11S in the Supplementary Materials).

Table 5 provides the estimated weighted global bias and RMSE for the LSEM estimates of the loadings and residual variances for the small ($N = 300$) and large ($N = 1{,}000$) sample size conditions. The results are shown for LSEM estimates with bandwidth factors of $h = 1.1$ and 2. The column *Optimal h* indicates the bandwidth factor that was optimal with respect to the estimated weighted global bias or RMSE. The main results can be summarized as follows: First, the estimated bias for most parameters was small in magnitude. Using the (average) true parameter to calculate a relative bias, the largest estimated bias was observed for the residual variance of *Short-term memory* (9.92), which corresponds to an estimated relative percentage bias of 14%. Second, the estimated bias decreased with increasing sample size. The largest estimated bias for $N = 1{,}000$ was below 5% in relative terms. Third, the optimal bandwidth with respect to the estimated bias decreased slightly as the sample size increased. This indicates that a smaller bandwidth factor may be preferred in larger samples. A similar relation was observed for the estimated RMSE. Fourth, the estimated bias and RMSE suggested different optimal bandwidth factors. For minimizing bias, bandwidth factors between 1.1 and 1.5 seem to be optimal, particularly with larger samples. However, the RMSE was clearly reduced with a bandwidth factor between 2.5 and 3. Thus, researchers are in a conflicting situation in that they can either minimize bias by choosing a small bandwidth factor or increase the overall accuracy by selecting a larger factor.

Table 4. Results of the simulation study: Relative percentage pointwise bias for a selected factor loading (comprehension knowledge) as a function of the sample size and bandwidth factor.

	$h = 1.1$			$h = 2$			$h = 3$		
Age	$N = 300$	$N = 600$	$N = 1,000$	$N = 300$	$N = 600$	$N = 1,000$	$N = 300$	$N = 600$	$N = 1,000$
4	1.9	4.3	2.5	8.6*	9.2*	7.7*	13.1*	13.2*	11.9*
5	0.8	2.6	1.4	6.2*	6.7*	5.6*	10.1*	10.2*	9.2*
6	− 0.5	0.9	0.4	3.5	4.0	3.3	6.7*	6.9*	6.2*
7	− 1.6	− 0.4	− 0.2	1.1	1.6	1.3	3.5	3.9	3.4
8	− 2.3	− 1.0	− 0.5	− 0.5	0.1	0.1	1.2	1.7	1.5
9	− 2.3	− 1.0	− 0.4	− 1.1	− 0.4	− 0.2	− 0.2	0.5	0.5
10	− 1.7	− 0.6	0.0	− 0.9	− 0.1	0.2	− 0.9	0.0	0.3
11	− 1.0	− 0.1	0.4	− 0.8	0.2	0.7	− 1.6	− 0.4	0.1
12	− 0.9	0.0	0.5	− 1.3	− 0.2	0.4	− 2.8	− 1.5	− 0.8
13	− 1.5	− 0.5	0.0	− 2.8	− 1.4	− 0.8	− 4.9	− 3.4	− 2.6
14	− 2.7	− 1.5	− 1.0	− 4.8	− 3.3	− 2.6	− 7.2*	− 5.7*	− 4.8
15	− 3.9	− 2.5	− 1.9	− 6.5*	− 4.9	− 4.2	− 9.0*	− 7.5*	− 6.7*
16	− 4.6	− 3.0	− 2.5	− 7.1*	− 5.6	− 4.9	− 9.3*	− 8.0*	− 7.2*
17	− 4.3	− 2.8	− 2.4	− 6.1*	− 4.8	− 4.3	− 7.7*	− 6.5*	− 6.0*
18	− 3.0	− 1.7	− 1.6	− 3.2	− 2.3	− 2.1	− 3.8	− 2.9	− 2.6
19	− 0.7	0.0	− 0.2	1.7	2.0	1.6	2.5	3.0	2.9
20	2.1	2.0	1.5	7.9*	7.3*	6.4*	10.5*	10.5*	9.9*
21	4.8	3.8	3.4	14.2*	12.6*	11.1*	19.0*	18.3*	17.2*
22	6.4*	5.0	4.7	19.2*	16.7*	14.8*	26.4*	25.0*	23.2*
23	6.6*	5.2*	5.3*	21.9*	18.8*	16.6*	31.4*	29.2*	26.9*

Note. Pointwise bias estimates for all other broad abilities are provided in the Supplementary Material. Cells with absolute values larger than 5 are printed with asterisks.

The accuracy–bias trade-off can be further illustrated by inspecting the estimated pointwise bias and RMSE for the loading function of the indicator *Comprehension-knowledge* (see Figure 7). As can be seen, the estimated bias was smallest for a bandwidth factor of 1.1. It is also evident that for bandwidth factors of 1.1, 2, and 3, the bias did not completely vanish for the largest sample size of $N = 1{,}000$. Furthermore, the estimated RMSE clearly decreased for larger sample sizes and was considerably smaller with bandwidth factors of 2 and 3 than with a factor of 1.1. In addition, a bandwidth factor of 2 produced slightly more accurate estimates than a bandwidth factor of 3.

Discussion

Three approaches for testing the moderating effects of a context variable on factor model parameters were exemplified in this article with measures of cognitive abilities across childhood and young adulthood. We argued

Table 5. Results of the simulation study: Weighted global bias and RMSE for the factor loadings as a function of sample size and bandwidth factor.

			$N = 300$						$N = 1,000$					
	True parameter		wg Bias			wg RMSE			wg Bias			wg RMSE		
	M	*SD*	$h = 1.1$	$h = 2$	Optimal *h*	$h = 1.1$	$h = 2$	Optimal *h*	$h = 1.1$	$h = 2$	Optimal *h*	$h = 1.1$	$h = 2$	Optimal *h*
							Loadings							
Ga	5.27	0.00	0.10	0.07	3	0.92	0.70	3	0.05	0.04	2.5	0.56	0.43	3
Gc	9.98	1.26	0.25	0.45	1.1	1.40	1.21	2	0.12	0.32	1.1	0.83	0.76	1.75
Gf	10.86	0.77	0.20	0.23	1.1	1.39	1.10	3	0.08	0.17	1.1	0.83	0.68	2.5
Glr	3.95	0.22	0.09	0.13	1.1	0.55	0.43	3	0.04	0.08	1.1	0.33	0.26	3
Gs	4.91	0.43	0.21	0.32	1.1	1.89	1.48	3	0.13	0.26	1.1	1.18	0.95	3
Gsm	10.68	1.02	0.15	0.24	1.1	1.07	0.85	3	0.09	0.19	1.1	0.65	0.53	3
Gv	4.64	0.00	0.10	0.06	3	0.96	0.73	3	0.02	0.01	2.5	0.59	0.45	3
							Residual variances							
Ga	49.39	4.11	2.10	1.96	1.5	8.54	6.64	3	0.80	1.32	1.1	5.17	4.11	3
Gc	84.86	17.25	3.63	2.92	1.75	16.73	13.11	3	1.24	1.90	1.3	10.29	8.13	2.5
Gf	82.69	23.53	4.07	5.05	1.3	17.79	14.77	2.5	1.93	3.36	1.1	10.50	8.83	2
Glr	13.97	3.31	0.58	0.64	1.3	2.72	2.12	2.5	0.29	0.43	1.1	1.67	1.33	2.5
Gs	70.47	22.10	9.92	6.27	3	35.91	27.12	3	3.12	1.69	3	21.79	16.64	3
Gsm	211.35	0.00	3.39	3.48	1.5	12.41	10.69	2	1,29	2.33	1.3	7.43	6.65	1.75
Gv	59.17	4.46	2.68	2.55	1.5	9.88	7.86	3	1.25	1.95	1.1	5.96	4.93	2.5

Note. Gc = Comprehension-knowledge; Gv = Visual-spatial thinking; Gf = Fluid reasoning; Gs = Processing speed; Gsm = Short-term memory; Glr = Long-term retrieval; Ga = Auditory processing; *h* = bandwidth factor; RMSE = root mean square error; wg = weighted.

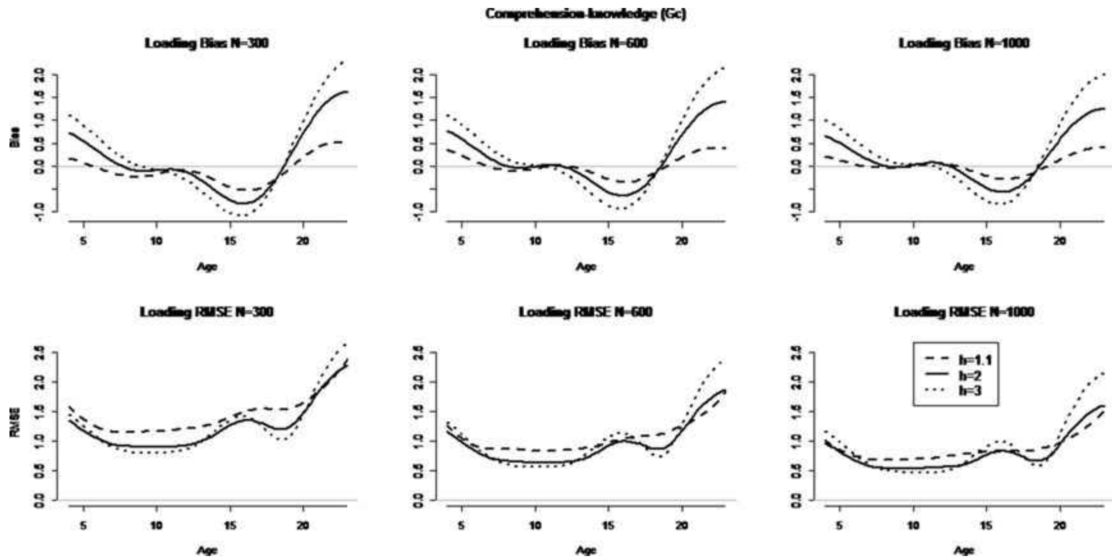

Figure 7. Illustration of the accuracy-bias trade-off for the example of the factor loading functions of *Comprehension-knowledge*.

that preferably, a factor model would be estimated on the basis of the conditional covariance structure at each value of a continuous moderator variable. In such an ideal case, the researcher can estimate the form of moderation without introducing strong assumptions because it is not necessary to specify whether the trend in the moderating effect is linear, quadratic, or a higher order polynomial. However, available data sets are often relatively small without hundreds of observations at the defined values of a continuous moderator variable. Thus, in most practical applications, the covariance structure—conditioned on the moderator variable—will have to be approximated.

In the MGMCS approach, observations are often arbitrarily merged to create sample groups that are large enough to estimate factor models simultaneously for groups of a given range on the moderator variable. Thus, the MGMCS approach is an approximation of a series of models estimated for grouped values of the moderator. We proposed LSEM as an extension of nonparametric regression to the SEM context (Hildebrandt et al., 2009). By weighting observations near the focal points of the moderator, the effective *N* for each focal point can be increased. The weighting scheme employed in LSEM allows for the application of SEM estimation across the entire range of the moderator. Finally, MFA approximates effects of a continuous moderator on factor model parameters by introducing linear and quadratic (or even more complex) effects that are fit to the whole range of the moderator variable in a single model.

In the present application, the results of MFA and LSEM were generally in accordance. Both approaches showed that the *g*-saturation of the first-order ability *Comprehension-knowledge* factor increased between ages 4 and 15 and decreased thereafter. The increase in this loading suggested dedifferentiation and was opposed to the differentiation hypothesis formulated for childhood age. However, the differentiation hypothesis was proposed for fluid abilities for which LSEM and MFA suggested no differentiation in our empirical example. This finding is in line with previous research (e.g., Molenaar, Dolan, Wicherts, & van der Maas, 2010; Tucker-Drob, 2009).

Molenaar, Dolan, Wicherts, and van der Maas (2010) investigated whether cognitive abilities could be differentiated by age and ability in the higher order factor model and identified five possible sources of differentiation: (a) residual variances of ability indicators, (b) loadings on the first-order factor (as also investigated by Tucker-Drob, 2009), (c) first-order factor residual variances, (d) first-order factor loadings on *g*, and (e) *g* variance. However, they showed that (a), (c), and (e) were statistically resolvable sources of differentiation. For this reason, we showed how LSEM and MFA could be used to investigate differentiation–dedifferentiation also at the level of residual variances. Because we used first-order abilities as indicators in the one-factor model, the present source of differentiation is related to source (c) as discussed by Molenaar, Dolan, Wicherts, and van der Maas (2010). The present results complement those findings because we considered a broader age range than Molenaar, Dolan, Wicherts, and van der Maas (2010) did. However, our findings are more heterogeneous than those found by Molenaar, Dolan, Wicherts, and van der Maas (2010), who reported no age differentiation. The LSEM (and MFA) analyses in our study suggest differentiation for three broad, first-order abilities (*Comprehension knowledge, Processing speed*, and *Long-term retrieval*) and dedifferentiation for two others (*Fluid reasoning* and

Short-term memory), whereas no change across age was detected for *Auditory processing*.

We showed that LSEM and MFA are complementary approaches. We described LSEM as a nonparametric variant of MFA and showed how LSEM has the advantage over MFA at exploratorily detecting the onsets of changes in parameter estimates across continuously observed moderator variables. For example, LSEM and MFA detected an increase in the factor loading of *Comprehension-knowledge* until approximately the middle of the age range we considered and a decrease thereafter. Although MFA and its effect plots (see Figure 4) allowed the user to visualize the symmetric bell-shaped function for the *Comprehension-knowledge* loading across age, LSEM revealed that the onset of the decrease was located specifically around the age of 16.

Furthermore, we showed that the MGMCS procedure, which groups cases across a broader age range arbitrarily, can lead to different conclusions from LSEM and MFA. In our empirical example, we were interested in whether the relation between an observed and a latent variable (represented by a factor loading) in a one-factor model depends on a third variable. The nature of a latent variable derived through factor analysis is highly dependent on the manifest variables that are included in the factor model. A latent variable (*g* in our examples) represents shared variance across all indicators (the broad abilities in our examples). Because there are relations between age (the moderator) and all WJ-III measures of cognitive ability, some of the variance shared between these abilities could be accounted for by age. However, in a factor analysis, the same variance could also be accounted for by a common factor. MFA takes into account the shared variance of the indicators that is due to common age trends by regressing the indicator intercepts on the moderator's linear and quadratic effects. Similarly, in LSEM, a detrending procedure (local polynomial smoothing) is applied to the indicators at the level of the observed data before the models are computed at each focal value. As a consequence, in both LSEM and MFA, the common factor, *g*, is derived only out of the portion of the shared variance of the indicators that is not due to age. This is different from the usual application of MGMCS, which incorporates shared variance that could be due to age by merging different age groups into multiple groups.

Future research on LSEM

LSEM has not yet been extensively investigated. In the present study, we included a first simulation study to provide some guidelines for the optimal bandwidth that should be chosen to calculate sample weights for LSEM. This first study suggested that a bandwidth factor of 2 was the optimal solution for sample sizes above 1,000 across the entire age range. However, we investigated only the bias and RMSE for the parameter estimates of LSEM. Additional simulations are needed if these recommendations also hold for other aspects of statistical inference (e.g., the Type I error rate and the power of hypothesis tests). Likewise, in addition to the permutation test presented, alternative significance tests for the moderation effects could also be explored.

There are a few data examples that have already suggested that LSEM offers advantages over other approaches (Briley et al., 2015; Hildebrandt et al., 2009; Hildebrandt, Sommer, Herzmann, & Wilhelm, 2010; Hülür et al., 2011; Schroeders, Schipolowski, & Wilhelm, 2015). The present work sheds further light on specific characteristics of model estimation based on LSEM. In the Supplementary Material, we provide R code implemented in the sirt package by Robitzsch (2015) to facilitate the use of LSEM.

Future research ought to more precisely investigate the performance of LSEM compared with competing approaches such as MFA. Simulation studies may be designed to test the performance of LSEM under certain circumstances—for example, for different sample sizes at focal points, different distributions of the moderator variable, or different effect size levels for the moderation effect. MFA might be expected to have more power to detect variation in factor model parameters if the form of the variation is correctly specified.

One advantage of LSEM is the flexibility with which it can be extended to assessing the effect of continuous moderators in very complex factor models. For example, LSEM can easily be adapted for binary, ordered-category, or count data. Furthermore, it would be interesting to apply LSEM to longitudinal or multilevel models (see Wu & Zhang. 2006). Researchers could explore how the variance decomposition of the longitudinal STARTS model (stable trait, autoregressive trait, and state; Kenny & Zautra, 2001) depends on a continuous moderator.

Conclusions and further applications

We presented two methods that can be applied to examine variation in factor model parameters along continuous variables—local structural equation modeling (LSEM) and moderated factor analysis (MFA)—along with multiple-group mean and covariance structure (MGMCS) analyses. We illustrated the nature of these analyses with the use of cognitive ability data, and we examined variation in the factor loadings with respect to the age differentiation hypothesis. The results of MFA and LSEM were mostly in accordance with each other. LSEM offers advantages for research questions regarding the form of variation in factor model parameters across a continuous moderator variable because LSEM does not require a priori specification of the variation function of the parameter estimates. If there are strong theoretical assumptions about the form of parameter variation, MFA is a reasonable choice because it allows traditional testing of the form of variation. However, the presented analyses showed that using both LSEM and MFA in conjunction might be a fruitful way to generate and test hypotheses.

The presented example can easily be expanded to include other potential continuous moderators of factor model parameters, such as socioeconomic status. For example, it has been hypothesized that personality traits are less correlated in high-ability than in low-ability groups, as well as at different levels of education (e.g., Mottus, Allik, & Pullmann, 2007; Rammstedt, Goldberg, & Borg, 2010; Toomela, 2003). In fact, in a given data set, it might be appropriate to use a broad variety of contextual variables to illuminate the issues and problems associated with the measurement instruments. The present work demonstrated how LSEM can be used to investigate such a hypothesis and provided code to facilitate the use of this modeling approach.

Article information

Conflict of interest disclosures

Each author signed a form for disclosure of potential conflicts of interest. No authors reported any financial or other conflicts of interest in relation to the work described.

Ethical principles

The authors affirm having followed professional ethical guidelines in preparing this work. These guidelines include obtaining informed consent from human participants, maintaining ethical treatment and respect for the rights of human or animal participants, and ensuring the privacy of participants and their data, such as ensuring that individual participants cannot be identified in reported results or from publicly available original or archival data.

Funding

No funding from a third party was available to complete this work.

Role of the funders/sponsors

None of the funders or sponsors of this research had any role in the design and conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Acknowledgments

The authors thank the Woodcock-Munõz Foundation for access to the normative data from the Woodcock-Johnson III Tests of Cognitive Abilities and Kevin McGrew for help with the data acquisition procedure. The ideas and opinions expressed herein are those of the authors alone, and endorsement by the authors' institutions is not intended and should not be inferred.

References

Amann, H. & Escher, J. (2008). *Analysis II*. Basel, Switzerland: Birkhäuser.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182. doi:10.1037/0022-3514.51.6.1173

Bauer, D. J., & Hussong, A. M. (2009). Psychometric approaches for developing commensurate measures across independent studies: Traditional and new models. *Psychological Methods*, *14*, 101–125. doi:10.1037/a0015583

Berry, J. W. (2003). Conceptual approaches to acculturation. In K. M. Chun, P. Balls Organista, & G. Marín, (Eds.), *Acculturation: Advances in theory, measurement, and applied research* (pp. 17–37). Washington, DC: American Psychological Association.

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

Boomsma A. (2013). Reporting Monte Carlo studies in structural equation modeling. *Structural Equation Modeling*, *20*, 518–540. doi:10.1080/10705511.2013.797839

Briley, D. A., Bates, T. C., Harden, K. P., & Tucker-Drob, E. M. (in press). Nonparametric estimates of gene × environment interaction using local structural equation modeling. *Behavior Genetics*, *45*, 581–596. doi:10.1007/s10519-015-9732-8

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York, NY: Cambridge University Press.

Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin*, *38*, 592.

Cheung, A. K., Harden, K. P., & Tucker-Drob, E. M. (2015). From specialist to generalist: Developmental changes in the genetic structure of early child abilities. *Developmental Psychobiology*, *57*, 566.533. doi:10.1002/dev.21309

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Curran, P. J., McGinley, J. S., Bauer, D. J., Hussong, A. M., Burns, A., Chassin, L., Sher, K., & Zucker, R. (2014). A moderated nonlinear factor model for the development of commensurate measures in integrative data analysis. *Multivariate Behavioral Research*, *49*, 214–231. doi:10.1080/00273171.2014.889594

Delaney, H. D., & Maxwell, S. E. (2004). *Designing experiments and analyzing data: A model comparison perspective*. New Jersey: Lawrence Erlbaum Associates Publishers.

Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications*. London, UK: Chapman and Hall.

Fox, J. (2000). *Nonparametric simple regression: Smoothing scatterplots*. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07–130. Thousand Oaks, CA: Sage.

Garrett, H. E. (1946). A developmental theory of intelligence. *American Psychologist*, *1*, 372–378. doi:10.1037/h0056380

Gasser, T., Gervini, D., & Molinari, L. (2004). Kernel estimation, shape-invariant modeling and structural analysis. In R. Hauspie, N. Cameron, & L. Molinari (Eds.), *Methods in human growth research* (pp. 179–204). Cambridge, UK: Cambridge University Press.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge, UK: Cambridge University Press.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis*. CRC Press.

Good, P. (2005). *Permutation, parametric, and bootstrap tests of hypotheses*. Springer.

Harville, D. A. (2008). *Matrix algebra from a statistician's perspective*. New York, NY: Springer.

Hildebrandt, A., Sommer, W., Herzmann, G., & Wilhelm, O. (2010). Structural invariance and age-related performance differences in face cognition. *Psychology and Aging*, *24*, 794–810. doi:10.1037/a0019774

Hildebrandt, A., Wilhelm, O., & Robitzsch, A. (2009). Complementary and competing factor analytic approaches for the investigation of measurement invariance. *Review of Psychology*, *16*, 87–102.

Horn, J. L. (1965). *Fluid and crystallized intelligence*. (Unpublished doctoral dissertation). University of Illinois, Urbana-Champaign.

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. doi:10.1080/10705519909540118

Hülür, G., Wilhelm, O., & Robitzsch, A. (2011). Intelligence differentiation in early childhood. *Journal of Individual Differences*, *32*, 170–179. doi:10.1027/1614-0001/a000049

Jackman, S. (2009). *Bayesian analysis for the social sciences*. Chichester, U.K.: Wiley.

Kenny, D. A., & Zautra, A. (2001). Trait–state models for longitudinal data. In L. M. Collins & A. G. Sayer (Hrsg.), *New methods for the analysis of change* (S. 243–263). Washington, DC: American Psychological Association.

Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, *65*, 457–474. doi:10.1007/BF02296338

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, *32*, 53–76. doi:10.1207/s15327906mbr3201_3

Little, T. D., Card, N. A., Slegers, D. W., & Ledford, E. C. (2007). Representing contextual effects in multiple-group MACS models. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 121–147). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Liu, Y., Magnus, B. E., & Thissen, D. (2015). Modeling and testing differential item functioning in unidimensional item response models with a single continuous covariate: A functional data analysis approach. *Psychometrika* [Epub ahead of print]. doi:10.1007/s11336-015-9473-x

Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2012). *The BUGS book: A practical introduction to bayesian analysis*. Chapman & Hall/CRC Texts in Statistical Science.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*, 19–40. doi:10.1037//1082-989X.7.1.19

McGrew, K. S., Schrank, F. A., & Woodcock, R. W. (2007). Technical Manual. *Woodcock-johnson III normative update*. Rolling Meadows, IL: Riverside Publishing.

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

Merkle, E. C., & Zeileis, A. (2013). Tests of measurement invariance without subgroups: A generalization of classical methods. *Psychometrika*, *78*, 59–82. doi:10.1007/S11336-012-9302-4

Molenaar, D. (2015). Heteroscedastic latent trait models for dichotomous data. *Psychometrika*, *80*, 625–644. doi:10.1007/s11336-014-9406-0

Molenaar, D., Dolan, C., & Verhelst, N. (2010). Testing and modeling non-normality within the one factor model. *British Journal of Mathematical and Statistical Psychology*, *63*, 293–317. doi:10.1348/000711009X456935

Molenaar, D., Dolan, C. V., Wicherts, J. M., & van der Maas, H. L. J. (2010). Modeling differentiation of cognitive abilities within the higher-order factor model using moderated factor analysis. *Intelligence*, *38*, 611–624. doi:10.1016/j.intell.2010.09.002

Mottus, R., Allik, J., & Pullmann, H. (2007). Does personality vary across ability levels? A study using self and other ratings? *Journal of Research on Personality*, *41*, 155–170. doi:10.1016/j.jrp.2006.03.007

Oberski, D. (2014). lavaan.survey: An R package for complex survey analysis of structural equation models. *Journal of Statistical Software*, *57*(1), 1–27. doi:10.18637/jss.v057.i01

Preacher, K. J., Rucker, D. D., MacCallum, R. C., & Nicewander, W. A. (2005). Use of the extreme groups approach: A critical reexamination and new recommendations. *Psychological Methods*, *10*, 178–192. doi:10.1037/1082-989X.10.2.178

R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from: http://www.R-project.org

Rammstedt, B., Goldberg, L. & Borg, I. (2010). The measurement equivalence of Big Five factors with different levels of education. *Journal of Research in Personality*, *44*, 53–61. doi:10.1016/j.jrp.2009.10.005

Robitzsch, A. (2015). *sirt: Supplementary item response theory models*. R package version 1.8-9. Retrieved from: http://cran.r-project.org/web/packages/sirt/

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. doi:10.18637/jss.v048.i02

Rotnitzky, A., Cox, D. R., Bottai, M., & Robins, J. (2000). Likelihood-based inference with singular information matrix. *Bernoulli*, *6*, 243–284. doi:10.2307/3318576

Schroeders, U., Schipolowski, S., & Wilhelm, O. (2015). Age-related changes in the mean and covariance structure of fluid and crystallized intelligence in childhood and adolescence. *Intelligence*, *48*, 15–29. doi:10.1016/j.intell.2014.10.006

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London, UK: Chapman and Hall.

Silverman, B. W., & Ramsay, J. O. (2005). *Functional data analysis*. New York, NY: Springer.

Spiegelhalter, D. J., Thomas, A., Best, N. G., & Lunn, D. (2003). *WinBUGS version 1.4 user manual*. MRC Biostatistics Unit, Cambridge, UK.

Szapocznik, J., Scopetta, M. A., Kurtines, W., & Aranalde, M. D. (1978). Theory and measurement of acculturation. *Revista Interamericana de Psicología*, *12*, 113–130.

Toomela, A. (2003). Relationships between personality structure, structure of word meaning, and cognitive ability: A study of cultural mechanisms of personality. *Journal of Personality and Social Psychology*, *85*, 723–735. doi:10.1037/0022-3514.85.4.723

Tucker-Drob, E. M. (2009). Differentiation of cognitive abilities across the life span. *Developmental Psychology*, *45*(4), 1097–1118. doi:10.1037/a0015864

Tucker-Drob, E. M. & Salthouse, T. (2008). Adult age trends in the relations among cognitive abilities. *Psychology and Aging*, *23*, 453–460. doi:10.1037/0882-7974.23.2.453

Walker, A. M. (1969). On the asymptotic behavior of posterior distributions. *Journal of the Royal Statistical Society, Series B*, *31*, 80–88.

Wasserman, L. (2004). *All of statistics. A concise course in statistical inference*. New York, NY: Springer.

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*, 10–18.

Woodcock, R. W. (1999). What can Rasch-based scores convey about a person's test performance? In S. E. Embretson & S. L. Hershberger (Eds.), *The new rules of measurement: What every psychologist and educator should know* (pp. 105–128). Mahwah, NJ: Lawrence Erlbaum.

Woodcock, R. W., McGrew, K. S., Schrank, F. A., & Mather, N. (2001, 2007). *Woodcock-Johnson III normative update*. Rolling Meadows, IL: Riverside Publishing.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001, 2007). *Woodcock-Johnson III tests of cognitive abilities*. Rolling Meadows, IL: Riverside Publishing.

Wu, H., & Zhang, J.-T. (2006). *Nonparametric regression methods for longitudinal data analysis*. Hoboken, NY: Wiley-Interscience.

Yin, J., Geng, Z., Li, R., & Wang, H. (2010). Nonparametric covariance model. *Statistica Sinica*, *20*, 469–479.

Appendix: Asymptotic normality of LSEM estimates

In the following Appendix, we provide a proof for the consistency and asymptotic normality of the LSEM estimates. The proof relies on two steps. First, asymptotic normality for the conditional covariance is based on result from the literature. Second, the estimated covariance matrix is plugged into the likelihood, and the delta method is used to derive the asymptotic normality of the model parameter estimates (i.e., loading and variance functions).

The term *vech* denotes the vectorization operator, which stacks all nonduplicated entries of a symmetric matrix Σ into a vector σ (i.e., $\sigma = vech(\Sigma)$; see Harville, 2008). The corresponding inverse transformation *avech* puts elements of a vector into a symmetric matrix, which is formalized by $\Sigma = avech(\sigma)$.

In the first step, the LSEM approach removes the mean structure with nonparametric estimation (e.g., polynomial smoothing). More specifically, weights are calculated for each case (depending on a bandwidth factor h). The weights are then used as sample weights to calculate the SEM for each focal point a. When the data are normally distributed, instead of using each individual case, the nonparametrically estimated covariance matrix $\hat{\Sigma}_a$ (which is a weighted covariance using sample weights specific to each focal point a) is used in the likelihood function as input data. Yin et al. (2010) showed that, if the bandwidth factor h is chosen proportional to $N^{-1/5}$, where N denotes the sample size, the estimated covariance matrix $\hat{\Sigma}_a$ is a consistent estimator for the true age-conditional covariance matrix Σ_a. In addition, Yin et al. (2010) also derived the asymptotic normality for the estimator $\hat{\sigma}_a = vech(\hat{\Sigma}_a)$ of the age-conditional covariance matrix $\sigma_a = vech(\Sigma_a)$:

$$\sqrt{Nh} \cdot \left(\hat{\sigma}_a - \sigma_a - h^2 B_a\right) \to N(0, V_a) \;\; (N \to \infty), \quad \text{(A1)}$$

where B_a and V_a denote fixed matrices that do not depend on N and h. The matrix B_a is an asymptotic bias term, and V_a can be used to compute the asymptotic covariance matrix of $\hat{\sigma}_a$.

In the second step of LSEM, the estimated covariance matrix $\hat{\Sigma}_a$ (or equivalently, the weighted observations using bandwidth factor h) is plugged into the likelihood function. Let Σ_a^* be the restricted covariance matrix implied by the structural model and Σ_a be the unrestricted covariance. Then the likelihood can be written as

$$l(\Sigma_a, \Sigma_a^*) = \log|\Sigma_a^*| - \log|\Sigma_a| + trace((\Sigma_a^*)^{-1}\Sigma_a). \quad \text{(A2)}$$

The likelihood function in Equation (A2) can be written equivalently as a function of the corresponding vectorized parameters σ_a and σ_a^* of the covariance matrices. The model-implied covariance matrix σ_a^* depends on a structural parameter θ_a containing the unknown loadings and variances of the LSEM model. Using a known function G (which maps the model parameters onto the covariance matrix), this dependency can be expressed as $\sigma_a^* = G(\theta_a)$. The likelihood in Equation (A2) can now be represented by a function F_0 such that $F_0(\sigma_a, \theta_a) = l(\Sigma_a, \Sigma_a^*)$. In LSEM, the estimated covariance matrix $\hat{\sigma}_a$ is plugged into F_0, resulting in a function $F_0(\hat{\sigma}_a, \theta_a)$, which is maximized with respect to the structural parameters

θ_a. This means that the derivatives of F_0 with respect to θ_a must be equal to zero. More generally, the derivatives of F_0 define a differentiable function $F_1(\sigma_a, \theta_a) = \partial F_0/\partial\theta_a(\sigma_a, \theta_a)$, which for nonsingular models (i.e., models with an invertible Fisher information matrix; see Rotnitzky, Cox, Bottai, & Robins, 2000) fulfills the assumptions of the implicit function theorem (Amann & Escher, 2008). Given that the LSEM model holds in the population, we have $\sigma_a = G(\theta_a)$, and thus $F_1(\sigma_a, \theta_a) = 0$. Using the implicit function theorem, there exists a linear approximation in the neighborhood of (σ_a, θ_a) with a differentiable function g mapping the vector σ_a to the vector θ_a such that

$$\hat{\theta}_a - \theta_a = g(\hat{\sigma}_a) - g(\sigma_a) \approx W_a(\hat{\sigma}_a - \sigma_a), \qquad \text{(A3)}$$

where the matrix W_a is defined by the derivative of g with respect to σ_a. According to the implicit function theorem, it holds that $W_a = (\partial F_1 \,/\, \partial\theta_a)^{-1}\,(\partial F_1 \,/\, \partial\sigma_a)$, where the derivatives are evaluated in $(\sigma_a,\ \theta_a)$. Using Equation (A3), we can approximate the distribution of $\hat{\theta}_a$ by

$$\sqrt{Nh}\cdot(\hat{\theta}_a - \theta_a) = \sqrt{Nh}\cdot W_a(\hat{\sigma}_a - \sigma_a - h^2 B_a) + \sqrt{Nh}\cdot W_a h^2 B_a. \qquad \text{(A4)}$$

As the last term in Equation (A4) is a constant vector, we obtain the following result by using Equation (A2) and the delta method:

$$\begin{aligned}\sqrt{Nh}\cdot(\hat{\theta}_a - \theta_a - h^2 W_a B_a) &= \sqrt{Nh}\cdot W_a(\hat{\sigma}_a - \sigma_a - h^2 B_a) \\ &\rightarrow N(0,\, W_a V_a W_a^T)\ (N \rightarrow \infty).\end{aligned} \qquad \text{(A5)}$$

This result shows that the LSEM estimates $\hat{\theta}_a$ are asymptotically normally distributed from which the consistency of the estimates follows. Yin et al. (2010; see Remark 3) argued that the magnitude of the bias of a parameter estimate is comparable to its standard error for an appropriately chosen bandwidth (h is proportional to $N^{-1/5}$).