Chapter 7
An Illustration of Local Structural Equation Modeling for Longitudinal Data: Examining Differences in Competence Development in Secondary Schools

Gabriel Olaru, Alexander Robitzsch, Andrea Hildebrandt, and Ulrich Schroeders

Abstract In this chapter, we discuss how a combination of longitudinal modeling and local structural equation modeling (LSEM) can be used to study how students' context influence their growth in educational achievement. LSEM is a nonparametric approach that allows for the moderation of a structural equation model over a continuous variable (e.g., socio-economic status; cultural identity; age). Thus, it does not require the categorization of continuous moderators as applied in multi-group approaches. In contrast to regression-based approaches, it does not impose a particular functional form (e.g., linear) on the mean-level differences and can spot differences in the variance-covariance structure. LSEM can be used to detect nonlinear moderation effects, to examine sources of measurement invariance violations, and to study moderation effects on all parameters in the model. We showcase how LSEM can be implemented with longitudinal of the *National Educational Panel Study* (NEPS) using the R-package *sirt*. In more detail, we examine the effect of parental education on math and reading competence in secondary school across three measurement occasions, comparing LSEM to regression based approaches and multi-group confirmatory factor analysis. Results provide further evidence of the

G. Olaru (✉)
Tilburg University, Tilburg, The Netherlands
e-mail: R.G.Olaru@tilburguniversity.edu

A. Robitzsch
Leibniz Institute for Science and Mathematics Education (IPN), Kiel, Germany
e-mail: robitzsch@leibniz-ipn.de

A. Hildebrandt
Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany
e-mail: andrea.hildebrandt@uol.de

U. Schroeders
University of Kassel, Kassel, Germany
e-mail: schroeders@psychologie.uni-kassel.de

strong influence of the educational background of the family. This chapter offers a new approach to study inter-individual differences in educational development.

7.1   Introduction

Research on education as a lifelong process often deals with questions addressing the trajectories of abilities and competencies across the lifetime of individuals (longitudinal design) or differences between individuals of different ages (cross-sectional design). The *National Educational Panel Study* (NEPS) combines both approaches in a multi-cohort sequence design providing access to high quality, nationally representative, longitudinal data on educational careers and on the developing competencies of preschoolers, students, and adults in Germany (Blossfeld et al., 2011). Educational studies are often concerned with identifying contextual factors (e.g., Hattie, 2009; Sirin, 2005; Watermann & Baumert, 2006) that might promote or impede learning beyond factors that can be identified on the individual level (e.g., prior knowledge, self-efficacy, grit).

To understand how such context variables moderate learning, it is vital to incorporate them adequately into longitudinal data analysis techniques. However, broadly applied traditional data analysis approaches for examining the influence of context variables in educational research (multiple regression, differences between extreme groups, etc.) have several major drawbacks. Regression analytic approaches only focus on mean-level differences across the covariate. Moderating effects are often studied categorically by comparing a small number of artificially created groups (e.g., with low vs. high socio-economic status). Unfortunately, in such multi-group confirmatory factor analysis, it is the statistical method and not the nature of the observed context variable that determines the way in which the data analysis is performed. To enrich the methodological toolbox of social and behavioural scientists, including researchers analysing the intensive longitudinal data of NEPS, we describe in this chapter a recently developed statistical data analysis technique that is suitable to examine moderation effects of continuous background variables—*local structural equation models* (LSEM; Hildebrandt et al., 2009, 2016)—and apply this technique to longitudinal data.

7.2   Longitudinal Local Structural Equation Modeling

7.2.1   Longitudinal Structural Equation Models

To examine the effects of educational and familiar context on educational trajectories in a longitudinal structural equation modeling framework, we first aim to set the methodological ground for the upcoming explanation. We neither elaborate on issues of assessment such as the need to develop and compile theoretically sound

and age-appropriate measures (for this purpose, see e.g., Coaley, 2014; Embretson & Reise, 2013), nor we detail core principles of structural equation modeling (see Hoyle, 2012; Kline, 2015). Also, we refer the interested reader to excellent and comprehensive textbooks and articles topic (e.g., Little, 2013; McArdle, 2009; Mund & Nestler, 2019), when it comes to in-depth discussions and applications of structural equation modeling with longitudinal data. Nonetheless, we want to mention that any longitudinal data analysis within the SEM framework should start by establishing and scrutinizing the measurement models within each measurement occasion. The aim is to probe the stability of the measured construct and to spot potential fluctuations in the factorial structure, which is commonly referred to as measurement invariance testing (Meredith, 1993; Little et al., 2007). In a subsequent step, the model is extended by specifying relations across measurement occasions. Structural equation modeling with longitudinal data has to tackle several modeling decisions, which will be explained in more detail in the following, including (a) the longitudinal measurement invariance, (b) the scaling of latent factors, and (c) the choice among different structural models to depict change.

From a measurement point of view, the basic question in longitudinal research is whether the same construct is being assessed over time. This is known as longitudinal measurement invariance. Similar to the cross-sectional case (Cheung & Rensvold, 1999; Meredith, 1993; Vandenberg & Lance, 2000), longitudinal measurement invariance requires the specification of several parameter constraints (e.g. Little et al., 2007; Liu et al., 2016). In general, the procedure for testing measurement invariance consists of a sequence of models with increasingly restrictive constraints on the measurement parameters. As a baseline model, a model without any constraints is specified, in which only same structure across time is estimated (i.e., configural invariance). Next, a model with equal factor loadings across time (i.e. metric invariance) is tested. Finally, in addition to the constrained factor loadings, the item intercepts are also constrained to equality across time (i.e., scalar invariance). The scalar level of measurement invariance is required to answer questions concerning mean level change across time. If introducing additional equality constraints on parameters were to result in a substantial deterioration of the model fit (e.g., Chen, 2007; Cheung & Rensvold, 2002), the assumption of measurement invariance would have to be discarded.

Factor scaling (also called factor identification) means that a metric needs to be established for the latent variable (or factor). There are several options for scaling latent variables. Preferably, the choice of scaling is led by considerations related to parameter interpretation according to the scientific hypotheses to be addressed. The factor identification method in longitudinal modeling also determines the metric in which changes in parameters across time are expressed and have to be interpreted (see Little et al., 2006). For instance, when using the reference variable method, in which the factor loading and the item intercept of a single indicator per factor is constrained to 1 and 0, respectively, the metric of the latent variable is equivalent to that of the chosen reference indicator. In the case of constraining the variance of the factor to 1 and its mean to 0 at the first measurement occasion (i.e., reference-group scaling), factor variances and means at subsequent measurement occasions are

identified and scaled relative to the first measurement occasion. Both scaling methods have some disadvantages: Differences cannot be interpreted in the original item metric and constraining the factor mean at the first measurement time point to 0 discards the possibility of examining factor mean differences across the moderator at baseline. One potential way to overcome these disadvantages is the so called effects coding method for scaling latent variables (Little et al., 2006). According to this approach, factors are taken to reflect a weighted composite of all items (i.e., weighted by the factor loadings). This is implemented by constraining factor loadings of a common factor to an average of 1 and item intercepts belonging to the same factor to an average of 0. This procedure allows researchers to estimate factor means and variances that correspond to the metric of the items at every measurement occasion.

Finally, there is a wide range of longitudinal modeling approaches from which researchers are expected to select the one that best fits their analysis objectives (for overviews, see McArdle, 2009; Mund & Nestler, 2019; Usami et al., 2019). These include autoregressive models (Selig & Little, 2012), cross-lagged panel models (Mund & Nestler, 2019), change score models (Ferrer & McArdle, 2010), latent growth curve models (McArdle & Bell, 2000), and their variants. These modeling approaches differ in how they conceptualize and assess sources of variance (i.e., between-person variance, within-person variance, and error variance; see Bainter & Howard, 2016). Thus, depending on the specific research question and the number of time points available, researchers have to select the most appropriate model: For example, autoregressive models are suitable for testing rank-order stability and variability across time, whereas change score models are suitable for investigating general developmental trajectories and individual differences therein. Some models incorporate both within- and between-persons differences, as well as inter-individual differences in intra-individual change (e.g., autoregressive latent trajectory model with structured residuals; Mund & Nestler, 2019).

For the application described in this chapter, we used a bivariate latent growth curve model (LGCM; see Fig. 7.1), because we aimed to examine academic achievement and growth and co-development in two core competencies (math and reading) from 5th to 9th grade. The focus is on modeling the influence of a contextual variable (educational background) on the structural parameters. A LGCM allows differentiating between the initial level of academic competencies (the intercept) and its growth (the slope) across the study period. Moreover, they are suitable to examine how the initial level is related to subsequent growth, or how initial values and growth on one competence are associated with the other competence and its growth. However, the data-analytic methods with respect to the moderator variable we describe in this chapter can be similarly applied to other families of longitudinal structural equation models.
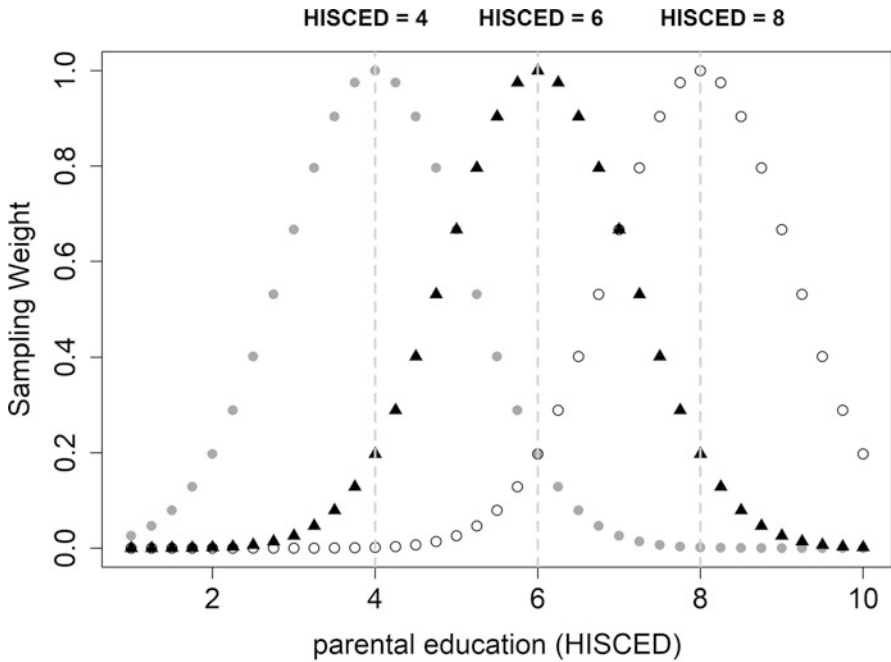
Fig. 7.1 Weighting functions for parental education (HISCED)

7.2.2 Including Covariates in a Longitudinal Structural Equation Model

The influence of background or context variables on parameters in a longitudinal model can be examined in various ways. The most broadly used approach is to include the context variable (e.g., parental SES) as a predictor of all latent variables. Thus, the (linear) relation of the context variable is accounted for, and factor residuals are interpreted as latent variables that were adjusted for the influence of the context variable. The downside of this approach is that it estimates only mean differences in the factor across the covariate. However, the covariates may also modulate other model parameters such as factor variances or factor covariances. In many applications, it is highly relevant to examine how individual differences in covariates are associated with the constructs and their growth, because this will help understand the processes of development more comprehensively than by examining a simple mean difference.

To examine the effect of a covariate on other model parameters than the mean, the covariate needs to be modelled as a moderator, which is often done with multi-group confirmatory factor analysis (MGCFA). In MGCFA, differences in model parameter are tested across a categorical moderator such as gender. For this purpose, model parameters are typically fixed to equality across groups, and deterioration in model

fit is tested following a straightforward procedure (for a detailed explanation, see Schroeders & Gnambs, 2018). MGCFAs are widely used and accepted for investigating model parameter differences across categorical context variables. However, to employ this method for continuous context variables such as SES, MGCFAs require one to first artificially categorize the context variable (e.g., into low vs. high SES groups by median split). But, artificially categorizing a continuous moderator has several disadvantages (see MacCallum et al., 2002; Preacher et al., 2005). First, nonlinear trends and complex patterns of moderation effects might be overlooked if too few groups have been analysed (e.g., Hildebrandt et al., 2016). Second, categorization results in a loss of any information on individual differences within a given moderator group. That means, when observations that differ across the range of a continuous variable are grouped, variation within these groups can no longer be detected. Third, setting cut-offs to split the distribution of a moderator into several parts is often arbitrary and might severely affect the results (e.g., Hildebrandt et al., 2009; MacCallum et al., 2002).

7.2.3 Local Structural Equation Modeling

In the following, we extend a recently developed method, *local structural equation modeling* (LSEM; Hildebrandt et al., 2009, 2016; Olaru et al., 2019), to longitudinal data aiming to overcome the aforementioned methodological issues. LSEM does not require an artificial categorization of moderators, renounces a pre-analytical specification of the relationship between moderator and psychological constructs, and can moderate both the mean and covariance structure. For these reasons, LSEM provides a very powerful approach with which to examine educational development across a wide range of background variables.

Next, we explain LSEM along an empirical example, demonstrating how researchers can examine contextual effects across a wide range of continuous moderators such as socio-economic status, years of formal education, or cultural embeddedness and a wide variety of models. LSEM has already been applied successfully to cross-sectional data to examine structural and mean-level differences in cognitive abilities across age or years of education (e.g., Gnambs & Schroeders, 2020; Hartung et al., 2018; Hülür et al., 2011; Schroeders et al., 2015) or to study age-related differences in personality (Olaru et al., 2019; Wagner et al., 2019). For instance, Wagner et al. (2019) and Olaru and Allemand (2022) used a combination of longitudinal models and LSEM to examine differences in the stability of personality traits and correlated change across the adult lifespan, respectively. In contrast to the current gap in the literature for such applications, combining LSEM and longitudinal SEMs is particularly important in educational research in which a wide range of different contexts (e.g., class, schools, peers, families) are theorized to have an important impact on the academic and extracurricular development of students and adults.

To achieve sufficiently stable parameter estimations, LSEM needs sufficiently large samples at each potential moderator value. Note that sample size restrictions are often the reason why naturally continuous moderators are categorized for MGCFA. This is because estimating a model at each moderator value (e.g., for each SES level) is not possible if only very few observations are available along single moderator values. As an alternative to achieve sufficiently stable parameter estimates, LSEM uses a sample weighting function to include observations from the neighbouring values on the moderator, albeit with smaller weights. The samples are weighted so that persons close to the targeted moderator value are weighted more strongly than persons farther away from this point. More specifically, the weighting function follows a Gaussian kernel function with a maximum of 1 at the focal point of the moderator considered (e.g., HISCED[1] = 6) and increasingly smaller weights for persons with a larger distance to the relevant moderator value (see Fig. 7.1). This approach assumes that observations close to each other on the moderator are more similar than distal observations. Figure 7.1 shows exemplarily (for three weight functions) that observations at the focal points receive a weight of 1, whereas observations with increasing distance from a focal point receive smaller weights. Because the Gaussian kernel function always attains values larger than zero, all observations will enter all models at each focal point in LSEM; but distant observations have very small values (below 0.01) resulting in no practical influence on the model parameter estimation at a given focal point.

After introducing the general idea of LSEM (for more details, see Hildebrandt et al., 2016; Olaru et al., 2019), we shall now illustrate the usefulness and versatility of the approach for analysing educational achievement outcomes in combination with contextual factors. More precisely, we apply LSEM to investigate mean, variance, and covariance differences in math and reading competencies[2] from the 5th to 9th grade of school (Starting Cohort 3; Blossfeld et al., 2011; https://doi.org/10.5157/NEPS:SC3:9.0.0) across educational levels of the family. To model mean-level performance and growth in the two domains as well as their interaction, we apply a bivariate latent growth curve model (see Fig. 7.2). Subsequently, we used LSEM to study the moderating effects of parental education within this model. We also compared the findings to a model in which the HISCED was included as a linear predictor of the factors, and to a model in which the HISCED was included as a categorical moderator (i.e., a multi-group confirmatory factor analysis across a low and high parental education group).

---

[1]HISCED is an acronym for highest international standard classification of education.

[2]For similarities and possible differences between the terms ability, skill, competence, and so forth, please see Schroeders (2018). In the present case, we use the terms synonymously.
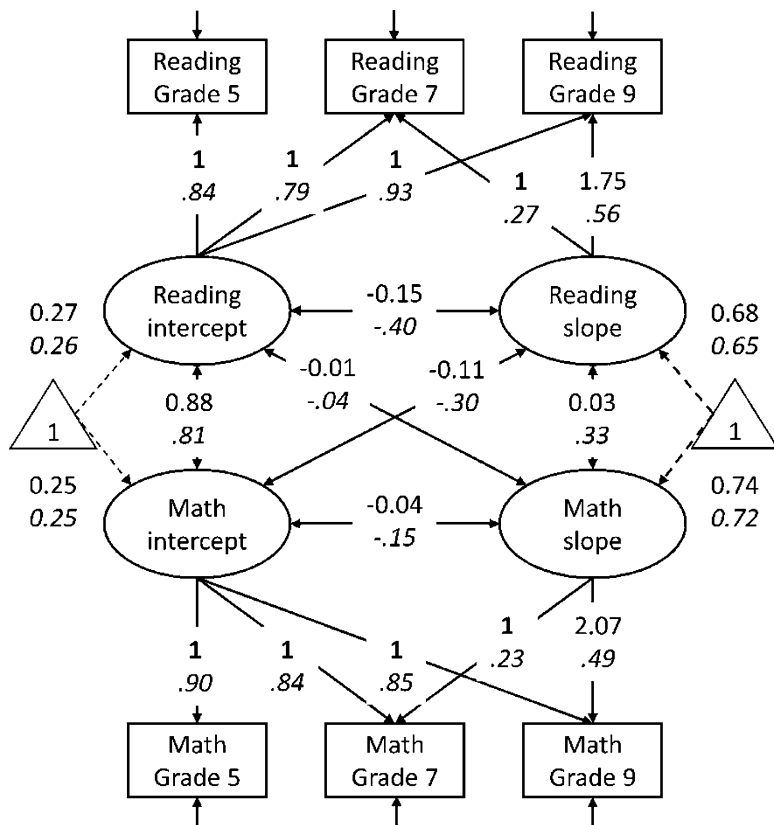
Fig. 7.2 Bivariate latent growth curve model of reading and math competence from Grade 5 to 9. Numbers show the estimated factor loadings, covariances, and means (triangles) on the full sample. Numbers in italics represent the standardized parameters; those in bold, constrained parameters

7.3 Method

7.3.1 Sample

The following illustration is applied to data from the National Educational Panel Study (NEPS): Starting Cohort Grade 5 (Blossfeld et al., 2011; https://doi.org/10.5157/NEPS:SC3:9.0.0). From 2008 to 2013, NEPS data was collected as part of the Framework Programme for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, NEPS is carried out by the Leibniz Institute for Educational Trajectories (LIfBi) in cooperation with a nationwide network. Of the Starting Cohort Grade 5 sample, we used only the 2037 students who had provided complete data on math

and reading competencies across the three measurement occasions together with their parents' education. Gender was balanced (50% female students). The mean age was 10.75 ($SD = 0.51$) in 5th grade, 12.75 ($SD = 0.49$) in 7th grade, and 14.92 ($SD = 0.46$) in 9th grade. Note that LSEM requires moderator values for each case used for model estimation, but can account for missing values in the indicators using pairwise estimation, imputed datasets, or model-based imputation (e.g., full information maximum likelihood; for an overview, see Lüdtke et al., 2007).[3] Because missing values in the data used for this demonstration indicated that some students did not participate in one or more measurement occasions (thus not being missing at random), we used only cases with full data.

7.3.2 Measures

7.3.2.1 Mathematical Competence

Mathematical competence in NEPS is a measure of mathematical literacy (OECD, 2009) requiring students to apply mathematics in realistic everyday situations. It combines content-related components (i.e., quantity, space, and shape; change and relationships; data and chance) with process-related components (i.e., applying technical skills, modeling, arguing, communicating, representing, and problem solving). For instance, the content-related facet of 'quantity' ranges from basic arithmetic operations (e.g., adding), over the use of different units, to simple equation systems. On the process-related side, the component 'technical skills' encompasses using known algorithms and calculation methods. The process 'representing' requires students to interpret tables, charts, or graphs, whereas 'problem solving' assesses students' ability to solve a problem with no obvious solution, typically by trying, generalizing, or examining exceptional cases.

7.3.2.2 Reading Competence

Reading competence is conceptualized in NEPS as competent handling of texts in different typical everyday situations. This operationalization of reading competence is based on the Anglo-Saxon literacy concept (also see OECD, 2009). The NEPS reading competence test combines different text forms, tasks, and response formats. Text forms consist of (a) factual texts (e.g., educational texts), (b) commenting texts (e.g., texts discussing a controversial question), (c) literary texts (e.g., short stories), (d) instructions (e.g., engineering manuals, cooking recipes), and (e) advertising

---

[3]Note that an imputation model has to be at least as flexible as the analysis model. For LSEM, the imputation of variables appearing in the SEM should allow relationships among variables to depend nonlinearly on moderators.

texts (e.g., job advertisements, recreational programmes) for which the lexical, semantic, and grammatical properties have been adapted to fit different age groups.

The reading comprehension tests require students to fulfil three types of tasks that were identified based on the reading comprehension literature (e.g., Kintsch, 1998; Richter & Christmann, 2002). These tasks are specified as (a) 'finding information in the text' (e.g., identifying information and recognizing statements), (b) 'drawing text-related conclusions' (e.g., relating several statements to each other in order to identify general propositions or the thoughts expressed in the text), and (c) 'reflecting and assessing' (e.g., deriving a situation model or understanding the central message of the text). Tasks and text forms are combined in a balanced manner to cover all possible text–task combinations.

7.3.2.3 Parental Education

We used the *international standard classification of education* (ISCED) as an indicator of parental educational levels. The ISCED provides information on educational attainment in terms of both the highest school certificate and the highest occupational qualification. The ISCED used in the NEPS study ranges from $0 = no$ *formal education* to $10 = doctoral degree$. We used the highest ISCED (HISCED) of both parents at the first measurement occasion as an indicator of educational levels in the family. If the ISCED was not measured in the first wave, we used the ISCED from subsequent measurement occasions. The average HISCED in the sample was 6.60 ($SD = 2.55$). It remained stable across the 4 years examined in this study (i.e., 95% of participants did not change in their value).

7.3.3 Statistical Analysis

7.3.3.1 Latent Growth Curve Model

As a starting point for our analyses, we used a bivariate latent growth curve model (LGCM; McArdle, 2009) on the math and reading competence ability estimates from an item response model linked across measurement occasions included in the NEPS SC3 dataset (Blossfeld et al., 2011). We modelled an intercept factor with loadings of 1 on all indicators. For the slope factor, we constrained the factor loadings to 0 and 1 for the first and the second measurement occasion respectively, while freely estimating the loading for the third measurement occasion. In contrast to other LGCM applications, the last slope loading was not constrained to 2 in order to allow nonlinear growth trajectories across time. All indicator intercepts were fixed to 0, so that factor means could be estimated. We allowed the intercept and slope factors of math and reading competence to covary. The model was estimated in *lavaan* (Rosseel, 2012) with maximum likelihood estimation. The *lavaan* code for

the model specification was as follows (please note that we use the original variable labels so that readers can replicate our example):

```
LGCM <- "
  # model intercept and slope factors
  math.inter =~ 1*mag5_sc1u + 1*mag7_sc1u + 1*mag9_sc1u
  math.slope =~ 0*mag5_sc1u + 1*mag7_sc1u + mag9_sc1u
  read.inter =~ 1*reg5_sc1u + 1*reg7_sc1u + 1*reg9_sc1u
  read.slope =~ 0*reg5_sc1u + 1*reg7_sc1u + reg9_sc1u

  # fix indicator intercepts to 0
  mag5_sc1u ~ 0*1
  mag7_sc1u ~ 0*1
  mag9_sc1u ~ 0*1
  reg5_sc1u ~ 0*1
  reg7_sc1u ~ 0*1
  reg9_sc1u ~ 0*1

  # estimate factor means
  math.inter ~ 1
  math.slope ~ 1
  read.inter ~ 1
  read.slope ~ 1"
```

7.3.3.2 Examining the Effect of Parental Education

We then compared three different approaches to examining the effect of parental education on math and reading competence: (a) a model with parental education as a linear predictor of the factors, (b) a MGCFA across two groups (constructed by median split), and (c) the LSEM approach. For the first approach, we regressed the intercept and slope factors on the HISCED (for the commented syntax, please see online supplement https://osf.io/vn297/). For the MGCFA approach aiming to examine differences in all model parameters, we split the sample into two groups around the median of 8 (participants with HISCED >7 were allocated to the group of individuals with higher education) and estimated the model simultaneously for the two groups without equality constraints across groups.

For LSEM, the lsem.estimate() function has been implemented in the *sirt* R-package (Robitzsch, 2019). We moderated the LGCM across HISCED values ranging from 3 to 9 in steps of 0.25 to provide a more nuanced picture than estimating the models only at full HISCED values. We excluded values at the borders of the distribution (0, 1, 2, and 10), because the effective sample size was low for these moderator values. Thus, the symmetric weighting function used in LSEM would create weighted samples skewed towards the middle of the distribution (because no participants can be found beyond the extremes; for an illustration see

Olaru et al., 2019). Based on suggestions in the literature (Hildebrandt et al., 2016), we used a bandwidth parameter of 2. The code used to run LSEM was as follows (for more information on the arguments of the function `sirt::lsem.estimate`, please refer to the manual or Olaru et al., 2019).

```
lsem.fit <- sirt::lsem.estimate(

  # data set
  data = mydata,

  # name of moderator
  moderator = 'hisced',

  # moderator levels to estimate model on
  # here from 3 to 9 in steps of 0.25
  moderator.grid = seq(3,9,0.25),

  # lavaan model syntax
  lavmodel = LGCM,

  # bandwidth parameter
  h = 2,

  # additional settings to estimate factor means
  residualize = FALSE,
  meanstructure = TRUE)

# return and plot the results
summary(lsem.fit)
plot(lsem.fit)
```

7.4 Results

The sample size for the baseline model used for the regression-based approach was $N = 2{,}037$. For the MGCFA approach, the sample was split into two groups with $n = 922$ (low education) and $n = 1{,}115$ (high education). In the LSEM approach, the weighted sample sizes ranged from $n = 479.15$ at HISCED = 3 to $n = 937.79$ at HISCED = 9 (the lowest weighted sample size was $n = 401.77$ at HISCED = 6).

Figure 7.2 shows the bivariate LGCM estimated on the full sample. Baseline performance in math and reading competence were strongly related ($\rho = .81$). The intercept factors were negatively related to growth, indicating that initially lower-performing students showed a higher increase in the competencies across school years. The growth of math competency was approximately linear (as indicated by the second slope factor loading of $\lambda = 2.08$), and the growth of reading competence was slightly smaller from Grade 7 to 9 (second slope factor loading of $\lambda = 1.72$).

7.4.1 Mean Level Differences

Figure 7.3 shows mean-level differences in the math and reading intercept and slope factors across parental education that can be displayed using the generic `plot()` function on the LSEM object (note that we also included the MGCFA and regression results in the figures). The LSEM plots of parameter estimates across parental education show that baseline math and reading competence are generally higher for students from families with a higher educational background providing a cognitively stimulating environment. Whereas math competence also shows a higher
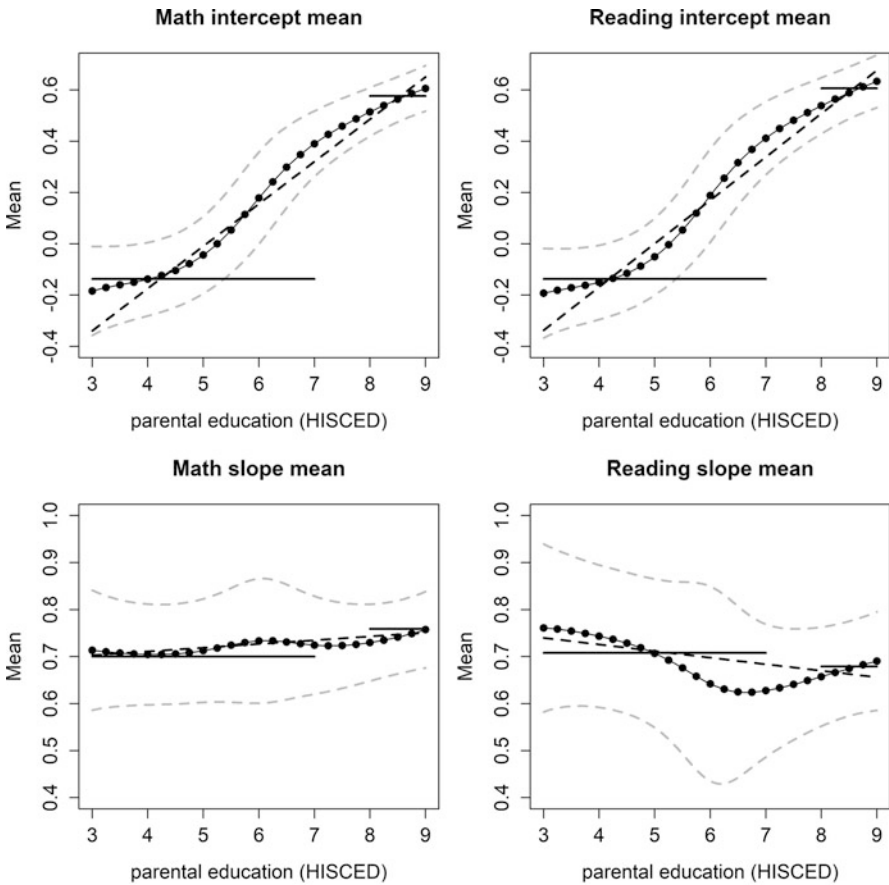
Fig. 7.3 Comparison of mean-level differences in mathematical and reading competence across three different methods. The dashed black line represents estimates based on the regression model. The black horizontal lines show estimates in the median-split MGCFA. The dotted black lines represent LSEM point estimates (i.e. each dot is the parameter estimate from a SEM at the corresponding moderator level). The dashed grey lines show the 95% confidence intervals for LSEM estimates

growth for these students, the effect seems to be negative for reading competence. Generally speaking, all three methods indicate the same pattern. However, the LSEM estimates show that the mean-level differences in the intercept factors are not perfectly linear across parental education, but have the steepest slope in the mid-range of the HISCED.

7.4.2 Differences in Variances and Covariances

The only moderation effect detected on the variance level was for the math intercept factor on which variance decreased across HISCED levels. Because of space restrictions here, we refrain from a detailed description, but point out that differences in factor variance might lead to biased results and should be investigated carefully. Additionally, they can also indicate meaningful differences in the distribution of inter-individual differences across the moderator. Concerning the correlations between the intercept and slope factors across parental education (see Fig. 7.4), the relation between math and reading growth decreases substantially across educational levels. This pattern suggests that growth trajectories in both competencies are more strongly related for students from a lower educational background. However, the large confidence intervals indicate that this effect might not be significant (for significance tests, see the section on 'Testing parameter equivalence'). The relationship between all other factor combinations remains stable across the HISCED. Again, the MGCFA and LSEM generally yield the same trends, but LSEM provides a much more detailed picture of the moderating effect.

7.4.3 Testing Moderation Effects

LSEM is primarily an exploratory method used to uncover potential effects across a continuous moderator. In general, examining potential moderations should start by examining the graphs provided by the `plot()` function on the output of the `lsem.estimate()` function. The plotted confidence intervals indicate whether parameter differences may be significant across the moderator. If point estimates at one moderator value are outside the confidence intervals at another moderator value, moderation effects can be concluded to be substantial. However, model parameter equivalence or measurement invariance cannot be tested by traditional means of inference testing (e.g., χ^2 difference tests) because the weighted samples used by LSEM overlap. Hence, alternative methods have been proposed to examine whether moderation effects are statistically significant: a permutation test (`lsem.permutationTest()` function) that has been used previously (Hildebrandt et al., 2016; Hülür et al., 2011; Schroeders et al., 2015) and joint estimation (setting the argument `est_joint = TRUE` within the `lsem.estimate()` function). The latter method is described for the first time in this chapter.
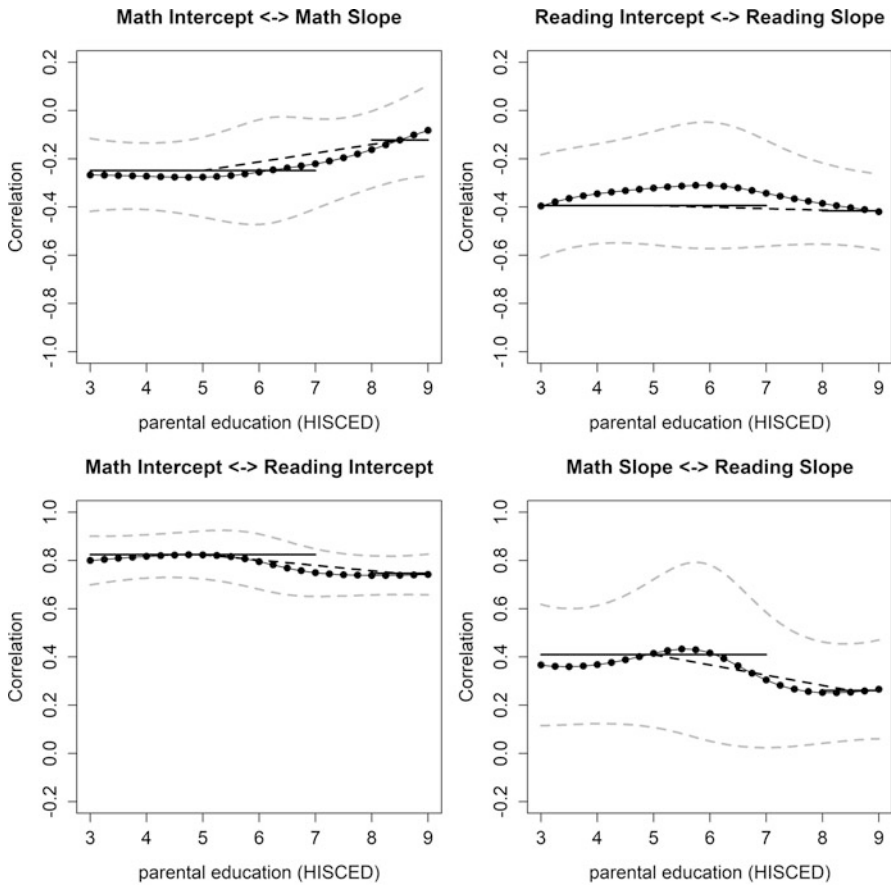
Fig. 7.4 Comparison of factor covariances across parental education for MGCFA and LSEM. Black horizontal lines show the estimates in the median-split MGCFA; dashed black lines, the linear approximation of the MGCFA differences; dotted black lines, the LSEM point estimates (i.e. each dot derives from a SEM). The dashed grey lines show the 95% confidence intervals for the LSEM estimates

The permutation test resembles traditional significance testing approaches in which the parameter values are tested against a distribution that can be expected to occur because of sampling error. To create such a distribution, the permutation test creates 1000 resampled copies of the dataset (on default settings). Within each dataset, the moderator values are shuffled around randomly across individuals (Hülür et al., 2011; Jorgensen et al., 2018). This removes all systematic moderation effects from the data. LSEM is then run on each dataset to derive the model parameters. This procedure results in a distribution of estimates for each parameter in which the parameter is independent of the moderator. The original LSEM parameter estimates are then compared to the corresponding distribution under the

null hypothesis. The permutation test function provides mean average distance, linear slope, and *p* values for each model parameter along the moderator. This allows users to identify which parameters change significantly across the values of the moderator, and whether the shape is linear or nonlinear. The permutation test can be run in R using the `lsem.permutationTest()` function on the `lsem.estimate()` object:

```
lsem.perm <- sirt::lsem.permutationTest(

  # lsem.estimate fit object
  lsem.object = lsem.fit,

  # number of permutations
  B = 1000,

  # allow mean-level differences
  residualize = FALSE)

# examine results
summary(lsem.perm)
```

The permutation test indicated that the reading and math intercept factor means differ significantly across parental education (see Table 7.1). As indicated by the significant linear slope value, the trajectories are approximately linear. The only other parameter that shows a significant moderation effect is the math intercept factor variance ($M = 0.970$; mean absolute distance $= 0.131$; mean absolute distance *p*-value $= .008$; linear slope $= -0.066$; linear slope *p*-value $= .006$) that decreases linearly across parental education. Whereas the decrease in the correlation between the growth factors from approximately $\rho = .70$ to .40 seems substantial, this effect is not significant, as also indicated by the large confidence intervals (Fig. 7.4).

Whereas the permutation test can be used to test moderation effects for each parameter separately, a more global approach of equivalence testing—similar to traditional MGCFA approaches—using a joint estimation procedure has been implemented recently in the *sirt* R-package (Robitzsch, 2019). The joint estimation procedure mirrors the approach used in MGCFA measurement invariance testing.

Table 7.1 Results of the permutation test for factors means

par	M	SD	SD_p	MAD	MAD_p	lin_slo	lin_slo_p
math.inter~1	0.235	0.320	0.000	0.306	0.000	0.153	0.000
math.slope~1	0.730	0.016	0.532	0.014	0.547	0.007	0.480
read.inter~1	0.247	0.337	0.000	0.322	0.000	0.161	0.000
read.slope~1	0.687	0.042	0.146	0.036	0.185	−0.015	0.248

Note. M = average parameter estimate across models, SD = standard deviation of parameter estimates, MAD = average absolute distance of parameter estimates from overall average, *lin_slo* = linear slope of the parameter estimates across the moderator, *_p* = corresponding global significance values

More specifically, each weighted sample in LSEM is treated as an independent group. By using a common likelihood function across groups, parameter estimates can then be derived across all moderator values simultaneously. In contrast, in the regular LSEM application, models are estimated separately, and parameter values can be constrained to equivalence only by specifying the values manually in the model. The joint estimation function allows users to estimate one parameter value across the moderator instead (if invariance assumptions are desired). Rather than providing model fit indices for each model across the moderator (e.g., CFI at each moderator level), the joint estimation procedure will also provide global fit indices (e.g., one global CFI value). By constraining parameters and examining the resulting model fit differences between the constrained and unconstrained model, measurement invariance or parameter equivalence in general can be evaluated in a similar way to MGCFA procedures. To use the joint estimation instead, the corresponding argument within the `lsem.estimate()` function has to be set to `est_joint = TRUE`. The resulting output will then correspond to a model with configural invariance (i.e., all parameters are unconstrained across the weighted samples). To constrain parameters to equality across the moderator, these need to be specified in the `par_invariant` argument. Parameters can also be constrained to follow a linear pattern by specifying the respective parameters with the `par_linear` argument. To constrain a parameter, it has to be included in the aforementioned arguments with the *lavaan* terminology. For instance, `par_invariant = c("factor1=~item1", "factor1=~item2")` will constrain the loadings of Factor 1 on Item 1 and 2 to equality across the moderator. LSEM will then return only one value for these parameters. The following code shows how LSEM with joint estimation and invariant parameters can be run:

```
lsem.fit.joint <- sirt::lsem.estimate(

  # LSEM parameters (see first code example)
  data = mydata,
  moderator = 'hisced',
  moderator.grid = seq(3,9,.25),
  lavmodel = LGCM,
  h=2,
  residualize = FALSE,
  meanstructure = TRUE,

  # joint estimation options
  est_joint = TRUE,

  # parameter equality constraints (examples)
  par_invariant = c(
    # invariant loading
    "math.slope=~mag9_sc1u",
    # invariant mean/intercept
    "math.slope~1",
```

(continued)

```
    # invariant covariance
    "math.slope~~math.inter",
    # invariant variance
    "math.slope~~math.slope",
    # invariant residual variance
    "mag9_sc1u~~mag9_sc1u")
  )

# examine results
summary(lsem.fit.joint)
```

The `summary()` output resembles the standard LSEM output except for global model fit indices. Both the permutation test and joint estimation can be used to investigate parameter equivalence, but the approach by which they do so differs between the methods. The strength of the permutation test is that it provides easy-to-use functionality for testing moderation effects on each parameter separately. The test results can be interpreted easily because they provide *p* values for each parameter moderation effect. The joint estimation procedure provides a global indication (e.g., CFI or RMSEA differences) of parameter equivalence that can be used to detect whether sets of parameters (e.g., all factor loadings) are equivalent across the moderator. Similar to MGCFA measurement invariance testing approaches, this can be done by comparing the model fit indicators across nested models (e.g., CFI differences between nested models should be below a value of .01; Cheung & Rensvold, 2002). Generally, it is advisable to run the permutation test first to identify which parameters are affected by the moderator. The joint estimation function can be used to impose constraints on the measurement model to investigate moderation effects in the structural model without bias—for example, by constraining all factor loadings before examining factor covariances. If the increase in misfit is too large as a result of the additional constraints, the most problematic parameters—as indicated by the permutation test—can be freed to achieve partial measurement invariance. Because both procedures can be used to test moderation effects on all model parameters, the two approaches can also be used to test invariance beyond traditional levels of measurement invariance that generally focus on factor loadings, item intercepts, and item residuals.

7.5 Discussion

This chapter illustrated different methodological approaches to the study the influence of contextual factors on educational achievement longitudinally. Traditional data analytic approaches—such as controlling for their influence by means of regressions or categorizing a continuous moderator and using MGCFA—are associated with a number of methodological limitations. LSEM, however, enables a detailed examination by providing nonlinear moderation effects on all parameters of

a SEM. The readily implemented functions of the *sirt* R-package allow educational researchers to scrutinize and test for measurement invariance. In the current example, we found that at Grade 5, students from families with higher education were better in math and reading than students from lower educational backgrounds. These differences due to parental education remained stable up to Grade 9, as indicated by the stable slope factor means. That is, the initial differences in the students' math and reading competencies across educational backgrounds remained stable in secondary school. Moreover, no moderation effect was found for the relation between initial competencies and growth. Formal education, however, seemed to help initially less capable students to catch up (see the stable negative correlations between the intercept and slope factors in both reading and math; Fig. 7.4), but this effect was similar across all educational backgrounds. On a more general stance, examining such structural differences in models of educational development is important to understand the processes underlying education and learning. For instance, one can assess whether the relation between mother-language competence and other academic competencies changes as a function of SES or cultural integration. Such an investigation would help to understand which students' language competence acquisition needs to be supported to improve knowledge in other academic fields.

7.5.1   Extensions of the LSEM Method

The nonparametric influence of moderators on parameters has a long tradition in varying coefficient regression models (Park et al., 2015). However, the principle of local estimation based on weights can be applied to any other model class that allows the use of sampling weights such as multilevel models (Wu & Tian, 2018), item response models, latent class models, mixture models, or survival models, to name a few. For longitudinal data, continuous-time models (Voelkle et al., 2012) are particularly attractive alternatives to the discrete-time models that were discussed in this chapter. In our empirical application, we also focused on only one moderator variable. Multiple moderator variables can be handled by replacing the unidimensional Gaussian kernel function for computing the weights with a multivariate Gaussian kernel function (see Hartung et al., 2018, for such an application). With many moderators, such an approach would lead to very sparse data, because only a few combinations of values would be available for multiple moderator values. Moreover, the interpretation of LSEM findings would be intricate in the multidimensional setting. One possibility would be to assume that only a subset of moderators affects a particular parameter. Essentially, this means that this parameter would be invariant with respect to the complementary set of moderators. This strategy can be implemented by using the joint estimation approach (see section 'Testing moderation effects'). Bolsinova and Molenaar (2019) discuss an LSEM application in which each item has its own moderator. They circumvent the problem of high dimensionality in the estimation by proposing an alternative estimation algorithm (Bolsinova & Molenaar, 2019). In their model, the set of parameters is

partitioned into subsets that depend on only one moderator variable (i.e. all parameters referring to an item depend only on the moderator corresponding to this item). The LSEM estimation is conducted by cycling through conditional estimation steps concerning the subsets of item parameters. Thereby only one subset of parameters is estimated, while holding all other parameters fixed. This principle can be generalized to LSEM applications with multiple moderators. This then results in an additive nonparametric model for the moderated parameter functions.

In the present demonstration, we used parental education (i.e., HISCED) at the first measurement occasion as a moderator that differs between participants but not within participants (i.e., across time). Because the HISCED values changed for only about 5% of the sample across the 4 years examined in this study, treating it as time-invariant was, in our opinion, a reasonable approximation. However, when using NEPS cohorts with younger participants (e.g., newborns and Kindergarten) and moderators with potentially stronger fluctuations across time (e.g., parental involvement; SES), the moderator values for each participant may change across time. It seems reasonable for model parameters referring to a particular time point to depend only on the moderator variable at this time point, as is done in the approach by Bolsinova and Molenaar (2019). For example, in a latent growth curve model, residual variances at a time point should depend only on the moderator assessed at this time point. However, it is less clear how intercept and slope variances depend on the time-varying moderator variables. One could argue that they depend only on the mean across time of the time-varying moderators, but they could alternatively depend on a measure of within-subject variability of the moderator or even depend on the moderator variables at all time points.

7.5.2   Alternative Modeling Approaches

Occasionally, the LSEM approach is computationally demanding, especially in cases with large models or more than one moderator variable. Alternatively, a computationally more parsimonious approach based on individual parameter contribution regressions (IPC) can be used to investigate relationships of model parameters to moderators (Arnold et al., 2019; Oberski, 2013). Both nonparametric approaches, LSEM and IPC, can be utilized to investigate whether a parametric approach such as moderated factor analysis (MFA) can be used (Hessen & Dolan, 2009; Molenaar et al., 2010; see also Hildebrandt et al., 2016, for a comparison). MFA allows for the inclusion of single or multiple parameter moderation effects in a structural equation model. For example, Molenaar et al. (2010) used it to study differentiation in a higher-order model of intelligence by examining moderation effects of age and ability levels on the factor and residual variances. MFA has the advantage that the test of moderation effects and model comparisons follows standard maximum likelihood or Bayesian theory. For example, moderation effects can be tested using χ^2-difference tests between nested models (e.g., by dropping or including single moderation effects in the model).

7.5.3 Conclusion

In our opinion, LSEM is an important tool for educational research because it can help to understand the underlying conditions of learning and to optimize education from the perspective of education policy. Uncovering which school, family, or child-related characteristics or backgrounds have a detrimental or favourable effect on learning is vital when it comes to identifying disadvantaged students and offering support that is targeted on the underlying mechanisms. Because the majority of these background variables are either continuous or are being understood increasingly as continuous concepts (e.g., cultural identity instead of categorical migration status), continuous moderation procedures are required to study these effects adequately. Whereas traditional measurement invariance approaches often focus only on the item level (i.e., factor loadings, item intercepts, and residuals), the procedures presented here provide equivalence tests for all model parameters that can be used to uncover differences across persons in the structure and mean levels of the latent variables as well.

References

Arnold, M., Oberski, D. L., Brandmaier, A. M., & Voelkle, M. C. (2019). Identifying heterogeneity in dynamic panel models with individual parameter contribution regression. *Structural Equation Modeling, 27*(4), 613–628. https://doi.org/10.1080/10705511.2019.1667240

Bainter, S. A., & Howard, A. L. (2016). Comparing within-person effects from multivariate longitudinal models. *Developmental Psychology, 52*(12), 1955–1968. https://doi.org/10.1037/dev0000215

Blossfeld, H. P., Rossbach, H. G., & von Maurice, J. (2011). The German National Educational Panel Study (NEPS). *Zeitschrift für Erziehungswissenschaft: Sonderheft 14*.

Bolsinova, M., & Molenaar, D. (2019). Nonlinear indicator-level moderation in latent variable models. *Multivariate Behavioral Research, 54*, 62–84. https://doi.org/10.1080/00273171.2018.1486174

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management, 25*, 1–27. https://doi.org/10.1177/014920639902500101

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

Coaley, K. (2014). *An introduction to psychological assessment and psychometrics*. Sage.

Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. Psychology Press.

Ferrer, E., & McArdle, J. J. (2010). Longitudinal modeling of developmental changes in psychological research. *Current Directions in Psychological Science, 19*, 149–154. https://doi.org/10.1177/0963721410370300

Gnambs, T., & Schroeders, U. (2020). Cognitive abilities explain wording effects in the Rosenberg Self-Esteem Scale. *Assessment, 27*, 404–418. https://doi.org/10.1177/1073191117746503

Hartung, J., Doebler, P., Schroeders, U., & Wilhelm, O. (2018). Dedifferentiation and differentiation of intelligence in adults across age and years of education. *Intelligence, 69*, 37–49. https://doi.org/10.1016/j.intell.2018.04.003

Hattie, J. (2009). The black box of tertiary assessment: An impending revolution. In L. H. Meyer, S. Davidson, H. Anderson, R. Fletcher, P. M. Johnston, & M. Rees (Eds.), *Tertiary assessment & higher education student outcomes: Policy, practice & research* (pp. 259–275). Ako Aotearo.

Hessen, D. J., & Dolan, C. V. (2009). Heteroscedastic one-factor models and marginal maximum likelihood estimation. *British Journal of Mathematical and Statistical Psychology, 62*, 57–77. https://doi.org/10.1348/000711007X248884

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.

Hildebrandt, A., Lüdtke, O., Robitzsch, A., Sommer, C., & Wilhelm, O. (2016). Exploring factor model parameters across continuous variables with local structural equation models. *Multivariate Behavioral Research, 51*, 257–258. https://doi.org/10.1080/00273171.2016.1142856

Hoyle, R. H. (Ed.). (2012). *Handbook of structural equation modeling*. Guilford Press.

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

Jorgensen, T. D., Kite, B. A., Chen, P. Y., & Short, S. D. (2018). Permutation randomization methods for testing measurement equivalence and detecting differential item functioning in multiple-group confirmatory factor analysis. *Psychological Methods, 23*(4), 708–728. https://doi.org/10.1037/met0000152

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford Press.

Little, T. D. (2013). *Longitudinal structural equation modeling*. Guilford Press.

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

Little, T. D., Preacher, K. J., Selig, J. P., & Card, N. A. (2007). New developments in latent variable panel analyses of longitudinal data. *International Journal of Behavioral Development, 31*, 357–365. https://doi.org/10.1177/0165025407077757

Liu, Y., Mo, S., Song, Y., & Wang, M. (2016). Longitudinal analysis in occupational health psychology: A review and tutorial of three longitudinal modeling techniques. *Applied Psychology, 65*, 379–411. https://doi.org/10.1111/apps.12055

Lüdtke, O., Robitzsch, A., Trautwein, U., & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung: Probleme und Lösungen. *Psychologische Rundschau, 58*, 103–117. https://doi.org/10.1026/0033-3042.58.2.103

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*(1), 19–40. https://doi.org/10.1037/1082-989x.7.1.19

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology, 60*, 577–605. https://doi.org/10.1146/annurev.psych.60.110707.163612

McArdle, J. J., & Bell, R. Q. (2000). An introduction to latent growth models for developmental data analysis. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples* (pp. 69–107, 269–281). Lawrence Erlbaum Associates Publishers.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525–543. https://doi.org/10.1007/BF02294825

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

Mund, M., & Nestler, S. (2019). Beyond the cross-lagged panel model: Next-generation statistical tools for analyzing interdependencies across the life course. *Advances in Life Course Research, 41*, 100249. https://doi.org/10.1016/j.alcr.2018.10.002

Oberski, D. (2013). Individual differences in structural equation model parameters. *arXiv*: 1304.3608.

OECD. (2009). *PISA: Take the test*. OECD Publications.

Olaru, G., & Allemand, M. (2022). Correlated personality change across time and age. *European Journal of Personality, 36*(5), 729–749. https://doi.org/10.1177/08902070211014054

Olaru, G., Schroeders, U., Hartung, J., & Wilhelm, O. (2019). Ant colony optimization and local weighted structural equation modeling: A tutorial on novel item and person sampling procedures for personality research. *European Journal of Personality, 33*, 400–419. https://doi.org/10.1002/per.2195

Park, B. U., Mammen, E., Lee, Y. K., & Lee, E. R. (2015). Varying coefficient regression models: A review and new developments. *International Statistical Review, 83*, 36–64.

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*(2), 178–192. https://doi.org/10.1037/1082-989X.10.2.178

Richter, T., & Christmann, U. (2002). Lesekompetenz: Prozessebenen und interindividuelle Unterschiede. In N. Groeben & B. Hurrelmann (Eds.), *Lesekompetenz: Bedingungen, Dimensionen, Funktionen* (pp. 25–58). Juventa.

Robitzsch, A. (2019). *sirt: Supplementary item response theory models*. R package version 3.6-21. https://CRAN.R-project.org/package=sirt

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling and more Version 0.5–12 (BETA). *Journal of Statistical Software, 48*(2), 1–36.

Schroeders, U. (2018). Ability. In M. H. Bornstein (Ed.), *The SAGE encyclopedia of lifespan human development* (pp. 1–5). SAGE. https://doi.org/10.4135/9781506307633.n8

Schroeders, U., & Gnambs, T. (2018). Degrees of freedom in multigroup confirmatory factor analyses: Are models of measurement invariance testing correctly specified? *European Journal of Psychological Assessment, 36*, 105–113. https://doi.org/10.1027/1015-5759/a000500

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. https://doi.org/10.1016/j.intell.2014.10.006

Selig, J. P., & Little, T. D. (2012). Autoregressive and cross-lagged panel analysis for longitudinal data. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods* (pp. 265–278). Guilford Press.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*, 417–453. https://doi.org/10.3102/00346543075003417

Usami, S., Murayama, K., & Hamaker, E. L. (2019). A unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods, 24*, 637–657. https://doi.org/10.1037/met0000210

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4–70. https://doi.org/10.1177/109442810031002

Voelkle, M. C., Oud, J. H. L., Davidov, E., & Schmidt, P. (2012). An SEM approach to continuous time modeling of panel data: Relating authoritarianism and anomia. *Psychological Methods, 17*, 176–192. https://doi.org/10.1037/a0027543

Wagner, J., Lüdtke, O., & Robitzsch, A. (2019). Does personality become more stable with age? Disentangling state and trait effects for the Big Five across the life span using local structural equation modeling. *Journal of Personality and Social Psychology, 116*(4), 666–680. https://doi.org/10.1037/pspp0000203

Watermann, R., & Baumert, J. (2006). Entwicklung eines Strukturmodells zum Zusammenhang zwischen sozialer Herkunft und fachlichen und überfachlichen Kompetenzen: Befunde national und international vergleichender Analysen. In J. Baumert, P. Stanat, & R. Watermann (Eds.), *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit* (pp. 61–94). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-90082-7_3

Wu, C. O., & Tian, X. (2018). *Nonparametric models for longitudinal data*. Chapman and Hall. https://doi.org/10.1201/b20631