Intraindividual Structural Equation Models for learning experiences

Lars-Erik Malmberg

University of Oxford, UK

Manuscript in progress for Special Issue on "The contribution of Multilevel Structural Equation Modeling to contemporary trends in Educational Research" in the International Journal of Research and Method in Education, version 24 June, 2020

Author Note

Acknowledgements: The data were collected during the author's Research Council's UK (RCUK) fellowship 2007-12. Address for correspondence: Lars-Erik Malmberg University of Oxford, Department of Education 15 Norham Gardens, Oxford OX2 6PY, U.K. Tel: +44-1865-274008, Fax: +44-1865-274144 e-mail: lars-erik.malmberg@education.ox.ac.uk

Abstract

With a growing interest in research on educational processes, there is a need to overview suitable latent variable models for students' learning experiences in real-time. This **tutorial** provides an **introduction** to intraindividual (multilevel) structural equation models (ISEM) for the analysis of process data (e.g., intensive longitudinal, intraindividual, diary, or person-period data) collected in educational settings. Using example data on 202 students' ecological momentary assessment of 10 to 24 reports ($M = 14.9$, $SD = 3.2$), of controlled (extrinsic) and autonomous (intrinsic) motivation the following models are presented: (1) measurement models and covariate effects models; (2) models for fixed, random, and moderator effects; and (3) models for reciprocal effects of chronologically ordered data. Step-by-step instructions for modelling, and substantive interpretations are given. Overall, ISEM establishes an important window into research on real-time educational processes.

Keywords: intraindividual; motivation; ecological momentary assessment; multilevel structural equation models; intensive longitudinal data

Intraindividual Structural Equation Models for learning experiences

Introduction

There is a growing interest in the analyses of process data (e.g., intraindividual, intensive longitudinal, diary, or micro-analytical data) in educational **research, and in the social sciences**. This trend follows a general surge in studies using ambulatory measurement (e.g., experience sampling, ecological momentary assessment) for collecting real-time self-reports (Hamaker & Wichers, 2017). Focus is on the sequence of events. These are multiple *learning experiences* that take place within a relatively short time-window. Situation-specific reports allow participants to report **on an event close in time** to the experience of **the** event. This reduces recall bias and enhances contextual closeness **of the resulting data**. Educational researchers interested in learning processes have collected process data in order to understand fluctuations of a multitude of variables, for example: interest, self-regulation, motivation, competence beliefs, and engagement. **The resulting data allows us to investigate how these variables are associated within and between time-points, and whether there are individual differences in such associations (for an overview see Murayama et al., 2017).**

The process-perspective was proposed by Schmitz (2006) according to which researchers, first, investigate the shape and form of changes in real-time (i.e., the process) and second, investigate the effects of instruction on these processes. Third, as proposed by Hamaker (2012) researchers would then investigate whether there are individual differences in the effects of instruction on the processes, and fourth, whether these differences are predictable by individual characteristics. In this **tutorial**, a series of intraindividual models relevant for educational research are presented. As example data we use controlled and autonomous motivation, which are two of many key dimensions for describing learning in classrooms. Controlled motivation represents effects of instruction in the form of demands, pressure and expectations in the classroom, and autonomous motivation represents the sense of self-direction, enjoyment, interest and choices during the learning process, both as experienced by students. This presentation complements other tutorials on multilevel modelling (MLM, e.g., Hox, 2002; Moskowitz

& Hershberger, 2002), MLM applied to diary data (Bolger & Laurenceau, 2013), and multilevel structural models (MSEM; e.g., Hox, 2013; Morin, Bollen, Nagengast & Scalas, 2014; Silva, Bosancianu & Littvay, 2020). The models presented here are suitable for data which includes time-points nested in students, for which researchers want to specify latent multiple-indicator constructs. These are referred to as intraindividual structural equation models or "ISEM". While educational researchers who have applied MLM and MSEM have typically focused on students nested in classrooms or schools, the sequence of models presented here can serve as a protocol for educational researchers who want to venture into **the emerging field of intraindividual research, focusing on time-points nested within students.**

An intuition for intraindividual data

According to self-determination theory motivation is conceptualized on a continuum from autonomous (or intrinsic) to controlled (or extrinsic) motivation (Ryan & Deci, 2000). Intrinsic motivation depicts the reasons for regulation as inherently internal, such as carrying out an action because one enjoys it, has the choice to do so, or one is interested in it. When actions are perceived as internally regulated individuals feel autonomous. In classrooms most learning goals are set by teachers, who can decide on the contents, sequence, and pace of the teaching, consistent with ideas of teacher-centered instruction (Cornelius-White, 2007), teacher-directed instruction (Connor, Morrison & Petrella, 2004), direct instruction (Flanders, 1970), or structure (Jang, Reeve & Deci, 2010; Skinner & Belmont, 1993). When teachers set expectations *and* support students at the same time, students sense of autonomy can be maintained even under pressure. If the expectations are upheld with threats, deadlines and negative repercussions for the individual, the sense of autonomy may be replaced by a sense of coercion.

In order to get an intuition for intraindividual data, let's start by looking at data from two fictional students. In Table 1, we see that students report their experienced controlled and autonomous motivation five times the second day of the study. There are

Table 1

Observed controlled and autonomous motivation of a fictional student, and two ways of creating lagged autonomous motivation

Day	Lesson	Subject	Cntr _T	Auto _T	Auto _{T-1}	Auto _{D-1}	Auto _{S-1}
2	1	Math	2.67	2.0	2.5	*	3.5
2	2	Eng	2.33	2.0	2.0	2.0	3.0
2	3	Sci	2.00	3.0	2.0	2.0	3.0
2	4	PE	2.00	3.0	3.0	3.0	*
3	1	Music	1.67	3.5	3.0	*	*
3	2	Computer	2.67	4.0	3.5	3.5	*
3	3	Sci	3.00	4.0	4.0	4.0	2.0
3	4	Math	3.33	3.5	3.5	4.0	2.0
3	5	Eng	2.33	3.0	3.5	3.5	2.0

Note: Auto_{T-1} is the lag, Auto_{D-1} is the within-school day lag, Auto_{S-1} is the within-school-subject lag. The first occurrence each day and school-subject is coded missing (Malmberg & Martin, 2019).

many ways in which we can make sense of these time-points nested in students, according to the organisational structure of school-life. The self-reports are nested within days, but can also be thought of as structured by school subject.

Figure 1. Controlled motivation (in blue) of the first fictive student, by day (left, day-means in red) and school subject (right, school-subject-means in red)

In Fig 1, we plot controlled motivation of Student 1 by day (to the left) and by

school subject (to the right). We can see that the self-reports vary below and above the daily means, the daily means are different from each other, highest on Wednesday and lowest on Thursday. Alternatively, when we group the ratings by school-subject we see differences in controlled motivation. It is low during the one Arts lesson and relatively highest during English lessons. These structures have been taken into account in modelling in different ways. For example, in their study spanning four weeks, (Martin et al., 2015) found that there was the most variance between lessons, the least variance between weeks, and moderate variance between students. Other studies have found small effects of school-subject (e.g., Tolvanen et al., 2011). In the present study we included time-of-day and day-of-the-week as predictors in the models.

To the left in Figure 2 we plot their controlled and autonomous motivation once each lesson during one week. First, we visually inspect the line-graphs (spanning time-points 1-22) which present ups and downs of situation-specific reports. At some points in time, controlled motivation was higher, at other times it was lower. The same happened with autonomous motivation. Note that both fictive students are here depicted to report the same amount of controlled motivation, **to demonstrate** how students can differ in their reactions to a teacher's actions. Second, we can disregard the chronological order of time, and estimate the association between controlled and autonomous motivation. As shown to the right in Fig 1 the association between controlled and autonomous motivation of the first student is moderately positive, the *intraindividual* correlation is $r = 0.32$), i.e., at points in time when controlled motivation is high, autonomous motivation is high. This student might interpret the teachers' actions as providing structure and guidance. For the second student, this correlation is negative ($r = -0.34$), i.e., when controlled motivation is high, autonomous motivation is low. This student might interpret the teacher's instruction as demanding, and to some extent coercive.

The difference in intraindividual correlations is important, as this suggests there might be *individual differences in associations* between controlled and autonomous motivation. Exploring such *differential* effects of controlled motivation on autonomous

Figure 2. Controlled motivation (black lines) and autonomous motivation (red/blue lines) of two fictive students: Student 1 time-trends (A) and intrapersonal correlation (B), Student 2 time-trends (C) and intrapersonal correlation (D)

motivation would be an important research question (i.e., individual differences in slopes). After that one would investigate if any individual characteristic could predict such differential effects (Hamaker, 2012), (i.e., moderators of the association between controlled and autonomous motivation).

Third, we can also think of the data-points in their chronological order, and estimate associations between each data-point (at Time T) and each previous time-point (the lagged time-point Time T-1) as shown in Table 1. This gives two *auto-regressive* correlations, the association between each observation of the same variable at T and the preceding time-point T-1. It gives two *cross-lagged* correlations,

Table 2

Intraindividual correlations: two fictive students

	Student 1			Student 2		
	Cntr _T	Auto _T	Cntr _{T-1}	Cntr _T	Auto _T	Cntr _{T-1}
Auto _T	.32			-.34		
Cntr _{T-1}	.42	-.01		.42	-.32	
Auto _{T-1}	.43	.58	.32	.06	.46	-.34

Note: Cntr = controlled motivation, Auto = autonomous motivation, T = concurrent time-point, T-1 = previous time-point.

the association between one of the variables at T and the other variable at T-1. We can then estimate further associations between these four variables. As shown in Table 2 we can see that both students have fairly stable impressions of controlled motivation over time. The association between controlled motivation at Time T and T-1 is $r = .42$. The stability of autonomous motivation is slightly higher for the first ($r = .58$) than for the second student ($r = .46$). A higher level of previous autonomous motivation is associated with a higher level of concurrent controlled motivation for the first student ($r = .43$), but not for the second student ($r = .06$). An opposite pattern is found when we look at the association between concurrent autonomous motivation (at T) and the preceding controlled motivation (T-1). We see that a higher level of previous controlled motivation is unrelated with the concurrent autonomous motivation for the first student ($r = -.01$), but negative for the second student, i.e., the higher level of controlled motivation the lesson before, the lower is autonomous motivation in the concurrent lesson ($r = -.32$). The directionality in the effects over time **is** different for the two students. Student 1 might perceive the teacher as tightening the grip the higher autonomous motivation is the previous lesson, i.e., the student has an *evocative* effect on the teacher (Nurmi, 2012). Student 2 seems to be more evoked by the teacher's action the previous lesson, i.e., an effect of the teacher on the student the subsequent lesson.

Models for intraindividual data

Models for intraindividual data are not new. The multilevel model for change (Hox, 2002; Raudenbush, 2002; Snijders & Bosker, 2012), and the latent growth model (Grimm, Ram & Estabrook, 2017), enable the estimation of individual differences in change over time. Multivariate time-series analysis enables the estimation of stability over time, and effects of one time-varying variable on another (Schmitz & Skinner, 1993). While growth models can be specified with as few as three time-points (or two time-points for a change model), time-series analyses typically require 100 data-points. Models for intraindividual studies, ranging between 20-40 observations per participant, has too many time-points for a growth model, and too few time-points for a time-series analysis, making the lagged-variable approach attractive (Musher-Eizenman, Nesselroade & Schmitz, 2002). Many intraindividual studies to date have made use of scale-scores or single-item response variables used with multilevel models. The structural equation model (SEM) was developed to use for multiple-indicator constructs in single-level models (Jöreskog, 1969; Jöreskog & Sörbom, 1982). Later developments allowed constructs at two levels to be fitted (B. O. Muthén, 1991, 1994) merging the best of two worlds MLM and SEM (Mehta & Neale, 2005; Rabe-Hesketh, Skrondal & Xiaohui, 2012). The *doubly latent* model has advantages over and above the use of scale-scores (Lüdtke et al., 2008; Marsh et al., 2009).

Research questions and models

A sequence of models which educational researchers could pose is summarized in Table 1. This sequence follows both the theoretical sequence of intraindividual research (Hamaker, 2012; Schmitz, 2006), the hierarchical nested structure of our data (here: time-points nested in students, nested in classes), and the logic of SEM. First, we need to ensure that the hypothesized factor structure of our latent constructs fit data. Then, we pursue two types of models with particular focus on the time-structure. The models presented here can be summarized as (1) measurement models and covariate effects models (Models 1a, 1b and 2); (2) models for individual differences in effects of

Table 3

Model descriptions

Model	Research Question	Description
Model 1a	Does the hypothesized factor structure fit data?	Configural invariance. Factor loadings are freely estimated at the within and between-levels.
Model 1b	Does the hypothesized factor structure fit data?	Strong invariance. Factor loadings are equated at the within and between-levels.
Model 2	Are there covariate effects on motivation?	Level-specific predictors included in model.
Model 3a	Does situation-specific controlled motivation predict situation-specific autonomous motivation	Random intercept, fixed effect model
Model 3b	Does the association between controlled and autonomous motivation vary across students	Random intercept, random slope model
Model 4	Does academic performance moderate the association between controlled and autonomous motivation	Cross-level interaction model, slopes as outcome model
Model 5	Is situation-specific autonomous motivation stable over time?	Autoregressive model.
Model 6	Which directionality is there between autonomous and controlled motivation?	Reciprocal effect model, including autoregressive and cross-lagged paths.

Note: Models 1 and 3 are described in two steps each.

instruction on processes (Models 3a, 3b and 4); and (3) models for the chronological order of time-lagged reciprocal effects between instruction and processes (Models 5 and 6).

Multilevel Confirmatory Factor Analysis. We start by investigating whether our hypothesized factor structure fits the data. We test our *a priori* idea that controlled motivation can be measured with two items and autonomous motivation with three items. Whether this model fits data or not is determined by the goodness of fit indices. We fit this model at both levels at once. At the within-level, we model how the reports of situation-specific experiences fit the pooled situation-specific constructs. At the between-level, we model how the students' average experiences (average across all the reports during the week) fit the student-level constructs. This will be done in two steps, the second step incrementally stricter than the first.

Upon establishing that the model fits data we can inspect how large proportion of variance there is at the between-level, by inspecting the intraclass correlations (ICC). This gives us an idea of how much students deviate from the grand-average (at the between-level) and how much students deviate from their own average (at the within-level).

Covariate effects. For researchers covariates can be important, either for answering a substantive research question (e.g., what is the effect of gender on motivation?), or for statistical control, to rule out that the sought effects are not due to extraneous factors (e.g., what is the association between controlled and autonomous motivation controlling for time of day?).

Important covariates to include could be features of the time-structure and characteristics of participants. In the present example, we illustrate the use of such level-specific covariate effects on motivation. At the within-level, we included time of day and day of the week, and gender and year-group at the between level. Depending on the role of the covariates, whether these answer specific research questions of the researcher or are used as statistical controls, their meaning could vary across projects. After inspecting covariate effects on key outcomes, researchers could then in their

specific design, decide to keep the covariates included in the model or to exclude them. In the sequence of modelling here, we for simplicity remove the covariate effects in the subsequent models. This model is called a *fixed* effect model in MLM literature.

In the following steps we pose Models 3a, 3b and 4, in which the chronological order of time is *not* explicit in the research questions instead modelling "individuals as their own controls", while in Models 5 to 6 we explicitly include this chronological order, by using lagged variables.

Individuals as their own controls. Upon establishing the first theoretical idea of intraindividual research, understanding the shape and form of processes (as done descriptively), researchers would be interested in whether instruction (controlled motivation) has an effect on processes (autonomous motivation). Phrased differently, we in Model 3a investigate whether a students' autonomous motivation in a particular situation is predicted by controlled motivation in that situation. In other words, when controlled motivation in a situation is higher than the students' average controlled motivation, is autonomous motivation then higher or lower? We only model this association at the within-level of Model 3a. This model is estimated so that the effect of controlled on autonomous motivation is the same for all students. This model is called a random intercept, *fixed* effect model in MLM literature, but in contrast to the the previous fixed effect model (Model 2) we include the predictor variable as a latent construct.

We then go on to investigate whether there are *individual differences* in the association between controlled and autonomous motivation, in Model 3b. Here we allow the regressions to look different for each student (individual differences in slopes), meaning that some students could be more autonomously motivated when controlled motivation is high, and others less motivated. As will be shown there were indeed individual differences in this association. This model is called a random intercept, *random* effect model in MLM literature. It means that the regression slope for each student is estimated to deviate from the average slope (the fixed slope in model 3a).

In Model 4 we include a student-characteristic as a moderator of the association

between controlled and autonomous motivation. Technically we include students' academic performance at the between-level in the model to predict the individual slopes. This model is called a cross-level interaction, or slopes-as-outcome model in MLM literature, analogous to a moderator in the single-level parlance. It means that we regress the individual regression slopes on the moderator.

Chronological order of time. When we want to take account of the chronological order of the data, we can conceptualize the time-structure in educational studies in many ways, e.g., time-points nested in days, nested in weeks, nested in students (Martin et al., 2015), fixed effects of time-structure variables (Tolvanen et al., 2011), as lagged models (Musher-Eizenman et al., 2002; Schmitz & Skinner, 1993), or Dynamic SEM (Asparouhov, Hamaker & Muthén, 2018). There are two important issues regarding the lagged structure of educational data (Malmberg & Martin, 2019), first, whether to lag within school-days (e.g., D-1 in Table 1) or within school-subjects (e.g., S-1 in Table 1), and second what to do with downtime, when students are away from school? In this example we use the within school-subject lag. In Model 5 we investigate whether autonomous motivation is stable between time S-1 and S (note that S is the same time-point as T), i.e., whether a previous learning experience in a particular school subject predicts the subsequent learning experience in that particular school subject, even though these might be a day apart, as shown in Table 1?

In Model 6 we are interested in the directionality in learning experiences, i.e., reciprocal effects (Guay, Marsh & Boivin, 2003), over time, particularly, does autonomous motivation in a previous lesson predict controlled motivation in the subsequent lesson, *or* does controlled motivation in a previous lesson predict autonomous motivation in the subsequent lesson?

Summary of models. In order to specify theory-informed models for intra-individual data (1. What is the shape and form of the processes? 2. How does instruction predict process(es)? 3. Are there individual differences in the predictions? and 4. Are there individual-level moderators of the predictions?), a set of models was presented and summarized in Table 3. Upon establishing that one's indicators fit the

hypothesized factor structure, and effects of any necessary covariates, the researcher can embark on modelling associations between constructs (i.e., "Individuals as their own controls"), or explicitly focus on the lagged chronological sequence of experiences. We will now illustrate these models using empirical example data.

Method

Sample and Procedure

Example data are part of the **Learning Every Lesson (LEL)** study (see Malmberg, Walls, Martin, Little & Lim, 2013). A subset of 202 students, in 14 classrooms, who had reported between 10 to 24 times during a calendar week, on average 14.9 time-points ($SD = 3.2$; in total 3,017 responses (see Malmberg & Martin, 2019). Of these 77 (38.1%) were in Year 5 and 125 (61.9%) in Year 6. There were 93 (46.0%) boys and 109 (54.0%) girls. Their mean age was 10.5 years ($SD = 0.65$). We used time-stamp data for creating covariates for the analyses: time of day, and day of the week (see Table 4). Student-reports of school subject (see Table 4) were used for creating the lagged school-subject variable S-1 (see Table 5) . Teachers reported the students' most recent academic performance, but imputed for five students. Otherwise the dataset was listwise deleted for simplicity.

Measures

Students' real-time experiences of learning were measured using the Learning Experience Questionnaire (LEQ; Malmberg, Woolgar & Martin, 2013). Autonomous motivation included three items in response to the opening phrase "why were you doing this task?...I enjoyed it", "I was interested" and "I chose it" (1 = not at all, 4 = very much). We report McDonald's ω (the proportion of shared variance between the items of each construct; McDonald, 1999; Revelle & Zinbarg, 2009) for each construct, $\omega_W = 0.83$; $\omega_B = 0.89$. We calculated the average and variation in internal consistencies across time-segments (cross-sections of the data): $M_a = .84$; $SD_a = .07$). Controlled motivation included two items "I had to do it" and "my teacher wanted me to do it" (1

Table 4

Time and school-subject

<u>Day of week</u>	n	%	<u>School-subject</u>	n	%
Monday	425	15.0	History	153	5.1
Tuesday	662	21.9	Geography	22	0.7
Wednesday	733	24.3	Art	177	5.9
Thursday	690	22.9	Mathematics	627	20.8
Friday	480	15.9	English	818	27.1
			Science	407	13.5
<u>Time of day¹</u>			Physical Education	291	9.6
9-10 a.m.	867	28.7	Music	162	5.4
10-11 a.m.	765	25.4	Internet Technology	121	4.0
11-12 a.m.	664	22.0	Modern Foreign Language	24	0.8
1-2 p.m.	469	15.5	Religious Education	47	1.6
2-3 p.m.	182	6.0	Health	23	0.8
3-4 p.m.	70	2.3	Other e.g., form-time	145	4.8

Note: ¹ = Times are approximate.

= not at all, 4 = very much; $\omega_W = 0.76$; $\omega_B = 0.94$; $M_a = .83$; $SD_a = .08$). Descriptives for the raw and lagged data are presented in Table 5. **The lagged variables have missing data. This data is missing by design, assumed to be Missing Completely at Random (MCAR; Schafer & Graham, 2002).**

Models and findings

The specification of and key findings from the models included in Table 1 are now presented. All models were fitted in Mplus (B. Muthén & Muthén, 1998-2017), and made available in the supplementary materials.

Table 5

Descriptives for raw data

	n	Min	Max	M	SD	Skew	Kurt	ICC
had	3017	1	4	3.21	0.92	-1.09	0.40	0.43
tea	3017	1	4	3.28	0.88	-1.25	0.91	0.47
enj	3017	1	4	3.00	1.04	-0.72	-0.70	0.29
cho	3017	1	4	2.88	1.09	-0.53	-1.04	0.45
int	3017	1	4	2.48	1.13	0.05	-1.38	0.29
had _{S-1}	1609	1	4	3.23	0.89	-1.14	0.65	
tea _{S-1}	1606	1	4	3.29	0.86	-1.29	1.18	
enj _{S-1}	1611	1	4	2.99	1.03	-0.70	-0.70	
cho _{S-1}	1607	1	4	2.90	1.08	-0.56	-0.99	
int _{S-1}	1607	1	4	2.48	1.11	0.06	-1.35	
Acad perf	202	-2.58	1.92	0.15	0.85	-0.53	-0.02	

Note: The items were: had = I had to do it, tea = teacher wanted me to do it, enj = enjoyment, int = interest, and cho = choice. The subscript $S-1$ indicates the within-school-subject lagged variable. Acad perf = teacher reported academic performance.

Models 1a and 1b. Two-level Confirmatory Factor Analysis

We started by inspecting descriptive data. As some of the indicators exceeded ± 1 for skewness, although none exceeded ± 3 for kurtosis (Hair Jr, Black, Babin & Anderson, 2014), we used Robust Maximum Likelihood (MLR) which adjusts the standard errors for minor deviations from normality in all models (see Table 5). Intraclass correlations showed that 29% to 47% of the variance was at the between-level, suggesting that taking account of the multilevel structure is warranted.

Figure 3. Multilevel confirmatory factor analysis of controlled and autonomous motivation.

The graphical representation of the MCFA follows the logic of placing the observed variables in the mid-section, with the between-part of the model above (with parameters indexed with "B") and within-part of the model below (with parameters indexed "W"). Please see the Supplementary Materials for the models represented in further detail in LISREL notation.

We specified MCFAs in Models 1a and 1b, to test whether the hypothesized factor structure would fit the data. Following the legend in Fig 3, we start with the raw data depicted in rectangles in the mid-part. Our five manifest (observed) variables provide the input information needed for the model: the variance-covariance matrix (see explanation in next paragraph). This information is partitioned into a within-part and a between part, giving a variance/covariance-matrix for each level. The smaller circles above and below the rectangles (linked with a directional arrow fixed at 1) depict this level-specific variance-covariance information. They are depicted as latent variables as it is not observed in the raw data. Latent constructs at the within ("W") and between levels ("B") are specified to predict the variance in the level-specific (latent) indicators. Arrows pointing from latent constructs to smaller circles are factor loadings, indicating this directionality: the construct predicts each manifest variables via regression-paths, e.g., $enj_B = intercept_B + factor-loading_B \times latent\ construct_B + a\ uniqueness_B$.

Although controlled motivation would not be identified on its own, as there are only three pieces of information (two variances and one covariance; $2 \times (2+1) / 2 = 3$), and we would want to estimate four parameters, and autonomous motivation is saturated (as there are $3 \times (4+1) / 2 = 6$) six pieces of information for estimating six parameters. For the two constructs, we have five indicators which give ($5 \times (5+1) / 2 = 15$) 15 elements in a variance-covariance information matrix to estimate the model on: 5 variances and 10 covariances. The factor structure is the same at the within-level (representing the time-points) and the between-level in the model (representing the students), but the factor loadings are allowed to be different at the two levels. This is called *configural invariance*. As the students were nested in 14 classes, we also need to take this **level of the hierarchical structure into consideration**. We do so by adjusting the standard errors in the model for the classroom-level clustering. In Model 1a the factor loadings are allowed to be different at the two levels.

In the next step, we continue to investigate whether the hypothesized factor structure fits our data, but impose a stricter measurement model, called *weak invariance* (Meredith, 1993), in order to establish **an equivalent** metric of the parameter

estimates (e.g., B , R^2) across the two levels (Morin et al., 2014). This is done by equating the factor loadings to be equal across the two levels. These are indicated with subscripts (a) to (e) in Fig 3.

We will assess if the measurement model (Model 1b) fits data by observing the Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR). Level-specific goodness-of-fit is given for the Standardized Root Mean Square Residual, for the within $SRMR_W$ and between-levels respectively $SRMR_B$. As the CFI and RMSEA were developed for single-level models, which are not sensitive to model-fit at different levels in a hierarchical structure we, in addition, hand-calculated the partially saturated Comparative Fit Index ($CFI_{PS,W}$ and $CFI_{PS,B}$), and the partially saturated Root Mean Square Error of Approximation ($RMSEA_{PS,W}$ and $RMSEA_{PS,B}$) following Ryu and West (2009) and Hox (2002). Thresholds for good fit are RMSEA below 0.06 and SRMR below 0.08, and CFA above 0.95 (Hu & Bentler, 1999). Both models 1a and 1b fitted data well as shown in Table 6. When we plot the factor scores of students' autonomous motivation in quadrant A of Fig 5, we see that students' differ in the range of experiences (some students deviate more, and others less from their mid experience) and student averages are also different from each other.

Model 2: Covariate effects. Two fixed effects predictors were included at the within-level, time of day and day of week. Time of day was centred at noon each day (-2 = first lesson, 0 = lesson before lunch, 3 = later afternoon lesson), and day of the week was centred at Wednesday (-2 = Monday, 0 = Wednesday, 2 = Friday). At the between level gender (0 = boy, 1 = girl) and year-group were included (0 = Grade 5, 1 = Grade 6). As shown in Fig 4 controlled motivation was higher later in the week ($\beta = 0.07$) but not during the day ($\beta = 0.04$). Autonomous motivation was higher later in the day ($\beta = 0.10$) and later in the week ($\beta = 0.09$). No effect of year-group or gender were found. For simplicity, we do not include the covariates in subsequent models, although researchers are advised to retain relevant covariates of their interest in models.

Figure 4. Key findings from models 2-6 (only structural part of models shown.

Notation is simplified).

Note: Coefficients are standardized beta-weights from Mplus (except for Model 3b which are unstandardized), with * = $p \leq .05$, ** = $p \leq .01$, *** = $p \leq .001$

Models 3a-3c: Fixed, random and cross-level interaction models.

Although it would be possible to jump directly to Model 4 for answering this particular research question, we here split up the steps for demonstrating transparency and good practice to incrementally build up models (see Fig 4). By going slowly step-by-step, we can keep an eye on whether model fit and parameter estimates become unreasonable at any stage. Details for all steps are available in the Supplementary Materials. We model autonomous motivation at both the within and between-levels, and controlled motivation at the within level only, as a predictor of autonomous motivation. Autonomous motivation is equipped with a disturbance term, which captures the residual in the regression equation: $auto_W = \beta c_ntr_W + \text{residual}$.

In Model 3a we are interested in the fixed effect of situation-specific controlled motivation on situation-specific autonomous motivation. We keep in mind this is the pooled average across all the situations. This effect was $\beta = 0.12$, $p = \text{n.s.}$, meaning

that students jointly experience somewhat (but not significantly) elevated enjoyment, interest and choice (autonomous motivation) in situations when teacher expectations (controlled motivation) were above their average controlled motivation (as the indicators of controlled motivation were centred within clusters; (Enders & Tofghi, 2007; Malmberg, Walls et al., 2013)). This effect is shown as parallel regression slopes in quadrant B of Fig 5, as all individual effects were estimated to be equal to each other.

In Model 3b we allowed both the intercept and the slope of each regression line to vary for each student. This is indicated as a dot in the within-part of the model, and we include a new construct "random slopes", at the between level. The average "random slopes" is the fixed effect of controlled on autonomous motivation ($B = 0.06$, $p = 0.39$), and the variance is the deviation of the individual slopes from this average slope ($\psi = 0.62$, $p \leq 0.001$). The steepness of the slope was unrelated to the student-average autonomous motivation ($\rho = 0.02$, $p = 0.49$). This means that some individual students feel inspired when demands are higher, while others get turned off when demands are higher, as shown in panel C of Fig 5.

In Model 4 we include a student-level moderator of the association between controlled and autonomous motivation. Both the individual slopes construct and student-average autonomous motivation were regressed on academic performance. We find that academic performance predicts the slope ($\beta = -0.23$; $p \leq 0.001$). This is referred to as a cross-level interaction, slopes-as-outcome model, or moderation effect. This effect is plotted in panel D of Fig 5. For relatively higher achievers an elevated controlled motivation was associated with lower autonomous motivation, and for relatively lower achievers with a higher autonomous motivation. This finding demonstrates how differently students interpret their teachers' actions, depending on whether they are low, mid-range, or high-achievers.

Figure 5. Illustrative results.

Note: **A** = Factor-scores of within-level autonomous motivation with superimposed medians, for 50 students (Model 1a); **B** = Predicted autonomous motivation from the random intercept and fixed slopes model (Model 3a); **C** = Predicted autonomous motivation from the random intercept and random slopes model (Model 3b); **D** = Estimated (cross level-interaction) effect of academic performance on the autonomous-on-controlled-motivation-slope (Model 4): full-drawn (red) line = academic performance -1 SD; dotted (green) line = academic performance 0 SD; dashed line (blue) = academic performance +1 SD.

[!htbp]

Table 6

Model fit

Model	χ^2	d.f.	p	CFI	CFI_{PS}	RMSEA	$RMSEA_{PS}$	$SRMR_W$	$SRMR_B$
1a. MCFA Configural invariance	10.36	8	0.24	0.999		0.010		0.008	0.030
1b. MCFA Weak invariance	32.923	12	0.001	0.991		0.024		0.015	0.077
MCFA within-level				0.999			0.012		
MCFA between-level				0.980			0.104		
2. 1b with covariates	54.347	24	≤ 0.001	0.989		0.020		0.014	0.065
3a. Fixed effect	12.31	7	0.091	0.997		0.016		0.011	0.083
3b. Random effect									
4. Moderation									
5. Auto-regressive effect	22.39	9	0.008	0.998		0.022		0.022	0.100
6. Reciprocal effects	55.915	34	0.010	0.996		0.015		0.026	0.071

Note: Please see Table 3 for an overview of the specified models.

Lagged-effect model

We now include constructs representing the lagged effects of the previous school-subject-lagged report (at time S-1)(see Models 5 and 6 in Fig 4). Thus, the lagged indicators had_{S-1} to cho_{S-1} were included to load on their respective controlled $S-1_W$ and autonomous $S-1_W$ constructs. Importantly, all the uniquenesses of like indicators, e.g., int_{S-1} and int_S , are allowed to correlate to account for the dependencies of the observations within each student (see Supplementary Materials). Only the concurrent time-point constructs are included at the between-level.

In order to investigate how stable students' autonomous motivation is from one lesson in a particular school-subject to the next, Model 5 included the lagged construct of autonomous motivation at time S-1 at the within level, as a predictor of autonomous motivation at time S. This is referred to as the auto-regressive path, interpreted as the rank-order stability of each student's experiences between the concurrent and previous time-point. **Positive** stability indicates that the level of autonomous motivation is consistently similar at each across the time-points. **Negative** stability indicates that a higher autonomous motivation at the previous time-point is associated with a lower autonomous motivation at the concurrent time-point. The stability of autonomous motivation was $\beta = 0.32$, $p \leq 0.001$, meaning that on average a higher level of autonomous motivation at time S-1 predicted a higher level of autonomous motivation at time S. It would be possible to investigate whether stability of autonomous motivation varies across the students (i.e., whether there are individual differences in stability between time S-1 and S) by specifying a model as in 3b.

We then, in Model 6 investigated the reciprocal effects between controlled and autonomous motivation over time, both the auto-regressive (i.e., from $controlled_{WS-1}$ to $controlled_{WS}$, and $autonomous_{WS-1}$ to $autonomous_{WS}$), and cross-lagged associations (i.e., from $controlled_{WS-1}$ to $autonomous_{WS}$, and from $autonomous_{WS-1}$ to $controlled_{WS}$) at the within-level of the model. Controlled motivation was grand-mean centred in order to interpret the effects of the controlled and autonomous motivations in the same metric across the constructs. Both motivations were stable over time: the

auto-regressive path for controlled motivation was $\beta = 0.30$, $p \leq 0.001$, and autonomous motivation was $\beta = 0.33$, $p \leq 0.001$. A higher level of autonomous motivation predicted a higher level of subsequent controlled motivation $\beta = 0.10$. Controlled and autonomous motivation were positively associated at the previous time-point $\rho = 0.26$, and the disturbance terms were associated at the concurrent time-point $\rho = 0.18$ (i.e., the association between controlled and autonomous motivation, controlling for the previous time-points).

Model summary and interpretations

To sum up the modelling, we first found that the hypothesized factor structure fitted the manifest items, also across the two levels of measurement, providing a solid base for interpreting situation-specific and individual differences in autonomous and controlled motivation. Applying an "individuals as their own controls" perspective on the modelling showed that students' experiences of controlled motivation were unrelated to their autonomous motivation. Then we found that some students were inspired by an elevated level of controlled motivation, and other were put off. It turned out that relatively lower-achievers were more inspired and relatively higher-achievers were somewhat put off. Applying a "lagged variable-perspective" showed that both autonomous and controlled motivation were quite stable from one subject-specific lesson to another. If a student experienced a higher level of autonomous motivation in the previous lesson, they experienced more controlled motivation in the concurrent lesson. The findings thus provide much support for teachers' perceptions that a teacher's "one size fits all" instruction for a task, can switch one student on, but deter another. A teaching strategy at the group-level, might not fit particular students in particular situations.

Discussion and conclusions

The aim of this tutorial was to provide an overview of ISEMs for educational researchers interested in the analysis of process data collected in educational contexts, using ambulatory techniques. Followed theoretical proposals by Schmitz (2006) and

Hamaker (2012), three broad types of models were proposed for investigating the associations between process (here: autonomous motivation) and instruction (here: controlled motivation): 1) measurement models and covariate effects models; (2) models for fixed, random, and moderator effects; and (3) models for reciprocal (auto-regressive and cross-lagged) effects of chronologically ordered data. Model specifications were given in the Supplementary Materials, and interpretations of key findings were given above. The discussion below focusses on considerations for (1) design of studies, (2) operationalization of measures, and (3) models and choices of software and estimators.

Design of intraindividual studies

The example data used in this presentation was correlational, and the modelling design two-level: time-points nested in students, adjusting standard errors for the classroom-level. This design works quite well for student groups in primary school, which stay relatively intact from one lesson to another. However, this design might not necessarily fit secondary school students, in which students belong to multiple student-groups depending on which subject they study. We know from intraindividual studies of teachers, that their experiences vary across student groups, with different student-group characteristics (Malmberg, Hagger & Webster, 2014). To accommodate such structures (see Mainhard, Oudman, Hornstra, Bosker & Goetz, 2018), future intraindividual studies could be planned as multiple membership or cross-classified designs. Situational student-reports and teacher-reports could be nested in multiple students, multiple student-groups, and their multiple teachers. If researchers were to include *both* student and teacher reports, i.e., cross-assessed data, this could reduce self-report bias. For example, all students in a class could report on their interaction with the teacher each lesson, while the teacher could report on the interaction with a select number of target-students each lesson. **While software for MLM enable modelling of multiple membership and cross-classified structures**, these structures are not at the time of writing well-implemented in SEM software. At the time of writing xxM (Mehta, 2013) allows specification of three or more hierarchical and cross-classified levels.

Second, while our data is correlational, the technique is amendable to intervention studies. Bolger and Laurenceau (2013) illustrated how to apply MLMs for experimental intraindividual data, by including experimental condition as a predictor of onset levels of an outcome variable (i.e., a point in time at which researchers expect no differences between groups), and differences in slopes (i.e., if the experimental group changes more in the expected direction than the control group), analogous to model 2 above (i.e., experimental condition would be a between-level predictor). The next generation intraindividual within-person designs for educational research could follow recommendations of Schmiedek and Neubauer (2019). In their "nudge" design, random allocation of encouragement could be given to individuals (e.g., "go on", "you're getting there", "don't give up"), so that effects of situation-specific encouragement on situational outcomes could be estimated. In this case, the experimental condition would be a within-level predictor.

Regardless of whether the study is correlational or experimental, the number of time-points, students, teachers, and classes need to be balanced so that substantive research questions can be answered. In the current example, no classroom-level effects were possible to estimate. Finding the optimal balance of structural elements of schooling is important. While some studies, like the current one, focussed on the complete coverage of all school subjects during a calendar week, other studies have focussed on a particular school subject, for example all science lessons over six weeks (Patall, Hooper, Vasquez, Pituch & Steingut, 2018). Finding the optimal number of self-reports that are feasible to collect also merits consideration. For example, if we collect data up to three times per lesson (e.g., beginning, middle, and end), or task (beginning, 5 minutes in, 10 minutes in, end) we can get closer to the processes we are studying. We can also capitalize on the rapidly advancing technology and smart-wear. Event-driven data-collection can be facilitated by e.g., being prompted to complete a brief survey every time we sit down in front of our computer.

How many items, situations, and persons are needed?

How many items do we need for capturing intraindividual constructs? Arguments are made for and against single item indicators, and multiple item indicators. Fewer items are faster to respond to and might reduce response fatigue. More items take longer to complete and might prompt greater reflection time, more recall bias, and elevate response fatigue (Goetz, Frenzel, Stoeger & Hall, 2010). Single-item measures typically correlate highly with corresponding multiple-item constructs. Multiple item-constructs can also be highly intercorrelated (Vasalampi et al., 2016). Finding the optimal balance between the number of items required and participants' response fatigue is an important issue to resolve for future simulation studies with implications for real-world data-collection.

How many observations do we need? Although some rules of thumb for multilevel models are around, e.g., at least 10 time-points of at least 30 persons, a recent simulation study (Schultzberg & Muthén, 2018) suggests that models with more persons and fewer time-points are generally more robust than models with fewer persons and more time-points. The number of time-points and persons required increases with the number of parameters researchers want to specify, e.g., random effects requiring more participants than fixed effects specifications.

Software and estimators for intraindividual studies

The models presented here would be possible to extend in multiple directions. While the number of time-points in an intraindividual study is typically larger than that of a growth-curve model, an alternative could be to apply differential equation SEM. In this specification, it is possible to estimate departures from a baseline and seeing how fast individuals can return to this baseline (Deboeck, 2013), and whether there are individual differences in the time required for returning to baseline. For example, how fast can a student return to **his or her** average level of autonomous motivation after a tough challenge which has lowered this motivation?

Another possible extension is the mixture model, a technique used for extracting

profiles across variables, typically at the individual level, e.g., "high motivation profile" vs. "low motivation profile". Two-level mixture models have started to emerge in the context of students nested in schools, in which clustering has been carried out at both levels (Collie, Martin, Papworth & Ginns, 2016). Also a profile study at the situation-level has been carried out (Neubauer, Dirk & Schmiedek, 2019). Applying this method to time-points within persons, enable researchers to create profiles of situations (i.e., at the within-level), and compare student profiles (i.e., at the between-level) based on proportions of situation-specific profiles.

Thirdly, there is a rapid expansion of available freeware for modeling MSEM, such as lavaan (Rosseel, 2012), xxM (Mehta, 2013), and OpenMx (Neale et al., 2016). Commercial software like Mplus (B. Muthén & Muthén, 1998-2017) can be used interactively through MplusAutomation (Hallquist & Wiley, 2018). Software with Bayesian estimation exist such as blavaan, stan and brms (Bürkner, 2017), which use the R-interface to communicate with C++. Space does not allow for elaborating on Bayesian models, but many introductions are available, e.g., (Van de Schoot et al., 2014; Kaplan & Depaoli, 2012). Importantly, Bayesian estimation can overcome overcome some of the problems researchers encounter in MSEM such as negative uniquenesses, and out-of-bound parameter estimates, (Zitzmann, Lüdtke, Robitzsch & Marsh, 2016). They also estimate random effects and parameters in cross-classified data efficiently (Asparouhov et al., 2018), and can provide robust parameter estimates for as few as 13 upper-level units of analysis (Hox, Van de Schoot & Matthijsse, 2012).

Conclusion: next generation intraindividual models for understanding learning processes

There is a surge in research on processes of learning experiences in real-time. Knowledge of how these processes are associated with instruction is important for the understanding of individual differences in learning. The models presented in this didactic piece serve a foundation for modelling multi-item, multiple-observation data nested in multiple students. These, and the next generation models for intraindividual

modelling provide an important window into learning that takes place in schools (and elsewhere), and the variability in the experiences within and between students.

References

- Asparouhov, T., Hamaker, E. L. & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 25, 359–388. doi: 10.1080/10705511.2017.1406803
- Bolger, N. & Laurenceau, J.-P. (2013). *Intensive longitudinal methods*. New York: Guilford.
- Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80. doi: 10.18637/jss.v080.i01
- Collie, R. J., Martin, A. J., Papworth, B. & Ginns, P. (2016). Students' interpersonal relationships, personal best (PB) goals, and academic engagement. *Learning and Individual Differences*, 45, 65–76. doi: 10.1016/j.lindif.2015.12.002
- Connor, C. M., Morrison, F. J. & Petrella, J. N. (2004). Effective reading comprehension instruction: Examining child x instruction interactions. *Journal of Educational Psychology*, 96, 682–698. doi: 10.1037/0022-0663.96.4.682
- Cornelius-White, J. (2007). Learner-centered teacher-student relationships are effective: A meta-analysis. *Review of Educational Research*, 77, 113–143. doi: 10.3102/003465430298563
- Deboeck, P. R. (2013). Dynamical systems and models of continuous time . In T. D. Little (Ed.), *The oxford handbook of quantitative methods in psychology: Vol. 2: Statistical analysis*. (pp. 411–431). Oxford, UK: Oxford University Press. doi: 10.1093/oxfordhb/9780199934898.013.0019
- Enders, C. K. & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, 12, 121–138. doi: 10.1037/1082-989X.12.2.121
- Flanders, N. (1970). *Analyzing Teacher Behavior*. Reading, MA: Addison-Wesley.
- Goetz, T., Frenzel, A. C., Stoeger, H. & Hall, N. C. (2010). Antecedents of everyday positive emotions: An experience sampling analysis. *Motivation and Emotion*, 34, 49–62. doi: 10.1007/s11031-009-9152-2
- Grimm, K. J., Ram, N. & Estabrook, R. (2017). *Growth modeling: structural equation*

- and multilevel modeling approaches*. New York, NY: Guilford.
- Guay, F., Marsh, H. W. & Boivin, M. (2003). Academic self-concept and academic achievement : Developmental perspectives on their causal ordering. *Journal of Educational Psychology*, *95*(1), 124–136. doi: 10.1037/0022-0663.95.1.124
- Hair Jr, J. F., Black, W. C., Babin, B. J. & Anderson, R. E. (2014). *Multivariate Data Analysis* (7th ed.). Harlow: Pearson Education.
- Hallquist, M. N. & Wiley, J. F. (2018). MplusAutomation: An R package for facilitating large-scale latent variable analyses in Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*, 621–638. doi: 10.1080/10705511.2017.1402334
- Hamaker, E. L. (2012). Why researchers should think "within-person": A pragmatic rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life*. (pp. 43–61). New York, NY: Guilford.
- Hamaker, E. L. & Wichers, M. (2017). No time like the present: Discovering the hidden dynamics in intensive longitudinal data. *Current Directions in Psychological Science*, *26*, 10–15. doi: 10.1177/0963721416666518
- Hox, J. (2002). *Multilevel analysis. Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum.
- Hox, J. (2013). Multilevel regression and multilevel structural equation modeling. In T. D. Little (Ed.), *The oxford handbook of quantitative methods in psychology: Vol. 2: Statistical analysis*. (pp. 281–294). Oxford, UK: Oxford University Press. doi: 10.1093/oxfordhb/9780199934898.013.0014
- Hox, J., Van de Schoot, R. & Matthijsse, S. (2012). How few countries will do? Comparative survey analysis from a Bayesian perspective. *Survey Research Methods*, *6*, 87–93. doi: 10.18148/srm/2012.v6i2.5033
- Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1–55. doi: 10.1080/10705519909540118
- Jang, H., Reeve, J. & Deci, E. L. (2010). Engaging Students in Learning Activities: It

is Not Autonomy Support or Structure but Autonomy Support and Structure.

Journal of Educational Psychology, 102, 588–600. doi: 10.1037/a0019682

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34, 183–202. doi: 10.1007/BF02289343

Jöreskog, K. G. & Sörbom, D. (1982). Recent developments in structural equation modeling. *Journal of Marketing Research*, 19, 404–416. doi: 10.1177/002224378201900402

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

Lüdtke, O., Marsh, H., Robitzsch, A., Trautwein, U., Asparouhov, T. & Muthén, B. (2008). The multilevel latent covariate model: a new, more reliable approach to group-level effects in contextual studies. *Psychological methods*, 13(3), 203–29. doi: 10.1037/a0012869

Mainhard, T., Oudman, S., Hornstra, L., Bosker, R. J. & Goetz, T. (2018). Student emotions in class: The relative importance of teachers and their interpersonal relations with students. *Learning and Instruction*, 53, 109–119.

Malmberg, L.-E., Hagger, H. & Webster, S. (2014). Teachers' situation-specific mastery experiences: teacher, student group and lesson effects. *European Journal of Educational Psychology*, 29, 429–451.

Malmberg, L.-E. & Martin, A. J. (2019). Processes of students' effort exertion, competence beliefs, and motivation: effects of lagged within-day and within-school-subject experiences. *Contemporary Educational Psychology*, 58, 299–309.

Malmberg, L.-E., Walls, T. A., Martin, A. J., Little, T. D. & Lim, W. H. (2013). Primary school students' learning experiences of, and self-beliefs about competence, effort, and difficulty: Random effects models. *Learning and Individual Differences*, 28, 54–65. doi: 10.1016/j.lindif.2013.09.007

Malmberg, L.-E., Woolgar, C. & Martin, A. J. (2013). Quality of measurement of the

- learning experience questionnaire for personal digital assistants. *International Journal of Quantitative Research in Education*, 1, 275–296.
- Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B. & Nagengast, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research*, 44, 764–802. doi: 10.1080/00273170903333665
- Martin, A. J., Papworth, B., Ginns, P., Malmberg, L. E., Collie, R. J. & Calvo, R. A. (2015). Real-time motivation and engagement during a month at school: Every moment of every day for every student matters. *Learning and Individual Differences*, 38, 26–35. doi: 10.1016/j.lindif.2015.01.014
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Hillsdale: Erlbaum.
- Mehta, P. D. (2013). *N-level structural equation modeling. xM user's guide. Version 1.0*. Los Angeles, CA.
- Mehta, P. D. & Neale, M. C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods*, 10(3), 259–284. doi: 10.1037/1082-989X.10.3.259
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58, 525–543. doi: 10.1007/BF02294825
- Morin, A. J. S., Bollen, M. H. W., Nagengast, B. & Scalas, F. S. (2014). Doubly latent multilevel analyses of classroom climate: an illustration. *The Journal of Experimental Education*, 82, 143–167. doi: 10.1080/00220973.2013.769412
- Moskowitz, D. S. & Hershberger, S. L. (2002). *Modelling intraindividual variability with repeated measures data: Methods and applications*. Mahwah, New Jersey: Laurence Erlbaum.
- Murayama, K., Goetz, T., Malmberg, L.-E., Pekrun, R., Tanaka, A. & Martin, A. J. (2017). Within-person analysis in educational psychology: Importance and illustrations. In D. W. Putwain & K. Smart (Eds.), *British journal of educational psychology monograph series ii: Psychological aspects of education — current*

trends: The role of competence beliefs in teaching and learning (pp. 71–87).

Oxford: Wiley.

Musher-Eizenman, D. R., Nesselroade, J. R. & Schmitz, B. (2002). Perceived control and academic performance: A comparison of high- and low-performing children on within-person change patterns. *International Journal of Behavioral Development*, *26*, 540–547. doi: 10.1080/01650250143000517

Muthén, B. & Muthén, L. (1998-2017). *Mplus User's Guide. Sixth Edition*. Los Angeles, CA: Muthén and Muthén.

Muthén, B. O. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement*, *28*(4), 254–338.

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research*, *22*, 376–398.

Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., ... Boker, S. M. (2016). OpenMx 2.0: Extended Structural Equation and Statistical Modeling. *Psychometrika*, *81*, 535–549. doi: 10.1007/s11336-014-9435-8

Neubauer, A. B., Dirk, J. & Schmiedek, F. (2019). Momentary working memory performance is coupled with different dimensions of affect for different children: A mixture model analysis of ambulatory assessment data. *Developmental Psychology*, *55*, 754–766. doi: 10.1037/dev0000668

Nurmi, J.-E. (2012). Students' characteristics and teacher-child relationships in instruction: A meta-analysis. *Educational Research Review*(7), 177–197.

Patall, E. A., Hooper, S., Vasquez, A. C., Pituch, K. A. & Steingut, R. R. (2018). Science class is too hard: Perceived difficulty, disengagement, and the role of teacher autonomy support from a daily diary perspective. *Learning and Instruction*, *58*(July), 220–231. doi: 10.1016/j.learninstruc.2018.07.004

Rabe-Hesketh, S., Skrondalh, A. & Xiaahui, Z. (2012). Multilevel structural equation models. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 512–531). New York: Guildford Press.

- Raudenbush, S. W. (2002). Alternative covariance structures for polynomial models of individual growth and change. In D. S. Moskowitz & H. S. L (Eds.), *Modelling intraindividual variability with repeated measures data: Methods and applications* (pp. 25–57). Mahwah, New Jersey: Laurence Erlbaum.
- Revelle, W. & Zinbarg, R. E. (2009). Coefficients Alpha, Beta, Omega, and the GLB: Comments on Sijtsma. *Psychometrika*, 74(1), 145–154. doi: 10.1007/S11336-008-9102-Z
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48, 1–36.
- Ryan, R. & Deci, E. (2000). Self-determination theory and the facilitation of intrinsic motivation. *American Psychologist*, 55, 68–78. doi: 10.1037/0003-066X.55.1.68
- Ryu, E. & West, S. G. (2009). Level-specific evaluation of model fit in multilevel structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 16, 583–601. doi: 10.1080/10705510903203466
- Schafer, J. L. & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7, 147–177. doi: 10.1037/1082-989X.7.2.147
- Schmiedek, F. & Neubauer, A. B. (2019). Experiments in the wild: introducing the within-person encouragement design . *Multivariate Behavioral Research*. doi: 10.1080/00273171.2019.1627660
- Schmitz, B. (2006). Advantages of studying processes in educational research. *Learning and Instruction*, 16, 433–449. doi: 10.1016/j.learninstruc.2006.09.004
- Schmitz, B. & Skinner, E. (1993). Perceived control, effort, and academic performance: Interindividual, intraindividual, and multivariate time-series analyses. *Journal of Personality & Social Psychology*, 64, 1010–1028. doi: 10.1037//0022-3514.64.6.1010
- Schultzberg, M. & Muthén, B. O. (2018). Number of subjects and time points needed for multilevel time-series analysis: A simulation study of dynamic structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(4), 495–515. doi: 10.1080/10705511.2017.1392862

- Silva, B. C., Bosancianu, C. M. & Littvay, L. (2020). *Multilevel structural equation modeling*. Quantitative Applications in the Social Sciences, 179. Thousand Oaks, CA: Sage.
- Skinner, E. A. & Belmont, M. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology*, 85, 571–581. doi: 10.1037/0022-0663.85.4.571
- Snijders, T. A. B. & Bosker, R. J. (2012). *Multilevel analysis: an introduction to basic and advanced multilevel modeling* (2nd ed.). London: Sage.
- Tolvanen, A., Kiuru, N., Leskinen, E., Hakkarainen, K., Inkinen, M., Lonka, K. & Salmela-Aro, K. (2011). A new approach for estimating a nonlinear growth component in multilevel modeling. *International Journal of Behavioral Development*, 35, 370–379. doi: 10.1177/0165025411406564
- Van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J. & van Aken, M. A. G. (2014). A gentle introduction to Bayesian analysis: applications to developmental research. *Child Development*, 85, 842–860. doi: 10.1111/cdev.12169
- Vasalampi, K., Muotka, J., Pöysä, S., Lerkkanen, M. K., Poikkeus, A. M. & Nurmi, J. E. (2016). Assessment of students' situation-specific classroom engagement by an InSitu Instrument. *Learning and Individual Differences*, 46–52. doi: 10.1016/j.lindif.2016.10.009
- Zitzmann, S., Lüdtke, O., Robitzsch, A. & Marsh, H. W. (2016). A Bayesian approach for estimating multilevel latent contextual models. *Structural Equation Modeling*, 23(5), 661–679. doi: 10.1080/10705511.2016.1207179