Revealing dynamic relations between mathematics self-concept and perceived achievement from lesson to lesson : An experience-sampling study

Niepel, Christoph, Marsh, Herbert W., Guo, Jiesi, Pekrun, Reinhard and Möller, Jens

This MS is the final prepublication (open access) version of the published article:

Niepel, C., Marsh, H. W., Guo, J., Pekrun, R., & Möller, J. (2022). Revealing Dynamic Relations Between Mathematics Self-Concept and Perceived Achievement From Lesson to Lesson: An Experience-Sampling Study. *Journal of Educational Psychology, 114*(6), 1380–1393. https://doi.org/10.1037/edu0000716

Revealing Dynamic Relations Between Mathematics Self-Concept and Perceived Achievement From Lesson to Lesson: An Experience-Sampling Study

Academic self-concept and achievement have been found to be reciprocally related across time. However, existing research has focused on self-concept and achievement scores that have been averaged over long time-periods. For the first time, the present study examined intraindividual (within-person) relations between momentary (state) self-concept and lesson-specific perceived achievement (i.e., self-reported comprehension) in students' everyday school life in real time using intensive longitudinal data. We conducted an experience-sampling (e-diary) study with 372 German secondary school students in Grades 9 and 10 over a period of 3 weeks after each mathematics lesson. Multilevel confirmatory factor analyses confirmed a two-factor between-level and within-level structure of the state measures. We used dynamic structural equation modeling to specify a multilevel first-order vector autoregressive model to examine the dynamic relations between self-concept and perceived achievement. We found significant reciprocal effects between academic self-concept and perceived achievement on a lesson-to-lesson basis. Further, we found that these relations were independent of students' gender, reasoning ability, or mathematics grades. We discuss implications for methodology, theory, and practice in self-concept research and educational psychology more generally.

Introduction

Academic self-concept (ASC)—the mental representation of one's own ability (Brunner et al., 2010; Marsh & Shavelson, 1985; Shavelson et al., 1976)—is key for students' academic success and well-being (e.g., Eccles, 2009; Marsh, 2006; Marsh et al., 2019). Researchers have repeatedly found that ASC is related to a broad range of outcomes, such as academic interests (Marsh et al., 2005; Schurtz et al., 2014), academic emotions (Arens et al., 2017; Pekrun et al., 2019), achievement goals (Dörendahl et al., 2021; Niepel et al., 2014a), and career aspirations (Guo et al., 2017). However, most research has focused on relations between ASC and students' academic achievement (e.g., for an overview, see Trautwein & Möller, 2016). At the core of this research is the finding that ASC and achievement are reciprocally related across time (Marsh & Craven, 2006; Marsh & Martin, 2011; Valentine et al., 2004; Wu et al., 2021). This finding implies that students with higher levels of achievement tend to develop higher ASC levels over time, whereas, at the same time, students higher in ASC tend to deliver higher levels of achievement in the long run.

However, despite the vast body of research on the relation between ASC and achievement, it appears that no studies on the longitudinal reciprocal relations between ASC and achievement have applied intensive longitudinal methods on the situational level to capture intraindividual dynamics in ASC and achievement in real time in actual learning situations (through ambulatory assessment, ecological momentary assessments, or experience sampling; for an overview of these methods, see, e.g., Bolger & Laurenceau, 2013; Hamaker & Wichers, 2017; Trull & Ebner-Priemer, 2014; Zirkel et al., 2015). We know from previous research that students' competence perceptions can be subject to everyday variation (Malmberg & Martin, 2019; Tsai et al., 2008). Nevertheless, due to the lack of intensive longitudinal studies on the reciprocal relations between ASC and achievement, the momentary (state) intraindividual (within-person) dynamics between ASC and achievement remain a black box. The existing longitudinal research on their reciprocal relations does not allow inferences to be made about within-person dynamics (see Murayama et al., 2017). By implication (as will be described in more detail below), the idea that ASC and achievement are mutually reinforcing has yet to be verified when shifting toward an intraindividual, real-time, and real-life perspective. To start filling this gap, we employed experience sampling (i.e., e-diaries via smartphones) in a sample of 372 German secondary school students. We tested the reciprocal relations and temporal dynamics of lesson-specific (state) ASCs and perceived achievement (i.e., self-reported comprehension) in the domain of mathematics in students' everyday life at school across a period of 3 weeks.

Reciprocal Relations Between Academic Self-Concept and Achievement

ASC is typically regarded as a highly domain-specific construct (e.g., Brunner et al., 2010) with, for instance, mathematics self-concept (MSC) representing a person's mental representation of their mathematics ability. There are a plethora of scientific articles on the relation between students' ASC and achievement. Thereby, achievement has been measured in various ways, such as standardized achievement test scores, record-cards grades, teacher ratings, or self-reports of achievement (see Valentine et al., 2004). Historically, three different main theoretical models that describe the two constructs' causal ordering can be distinguished (Calsyn & Kenny, 1977; Marsh & Martin, 2011). First, the skill-development model claims that students' previous achievement causes ASC (i.e., skill-development effect), whereas students' ASC has no impact on their later achievement. Second, the self-enhancement model claims that students' ASC influences their achievement (i.e., self-enhancement effect), whereas the latter is supposed to have no impact on their later ASC. Third, the reciprocal effects model claims that previous achievement affects ASC, and previous ASC affects achievement (Marsh, 1990). According to the reciprocal effects model, the relation between the constructs is characterized by long-lasting, mutually reinforcing skill-development and self-enhancement effects. Self-enhancement effects (ASC causes achievement), which are claimed by both the self-enhancement model and the reciprocal effects model, play a central role in ASC theory and research. Their support implies that interventions that are aimed at fostering ASC would also impact students' academic achievement (Ehm et al., 2019).

The vast majority of empirical findings have supported the reciprocal effects model (e.g., Arens et al., 2017; Guay et al., 2003; Marsh & O'Mara, 2008; Marsh et al., 2018; Retelsdorf et al., 2014; but see Ehm et al., 2019). These findings include meta-analytical evidence (Valentine & DuBois, 2005; Valentine et al., 2004; Wu et al., 2021; cf. Huang, 2011). For example, Wu et al. (2021) reported average effect sizes of .08 for self-enhancement and somewhat stronger effect sizes of .16 for skill-development effects. Moreover, the reciprocal effects model pattern has been found to generalize across gender as well as different cultures and ability levels (e.g., Gorges et al., 2018; Marsh & Martin, 2011; Marsh et al., 2005; Seaton et al., 2015; Valentine et al., 2004; cf. Wu et al., 2021).

Lesson-to-Lesson Dynamics Between Academic Self-Concept and Perceived Achievement: Shifting Toward a Within-Person and Short-Term Perspective

Despite the extensive empirical support for reciprocal relations between ASC and achievement, there appears to be no research that has used (a) a within-person, intraindividual approach combined with (b) a short-term, state-based approach in drawing on intensive longitudinal data obtained through experience sampling. Instead, previous research has typically taken a between-person approach and focused solely on long-term relations in drawing on only a few assessments of ASC and achievement, bridging time spans of several months to years (Ehm et al., 2019; Huang, 2011; Valentine et al., 2004; Wu et al., 2021).

Most research has employed cross-lagged panel models to examine the long-term relations between ASC and achievement longitudinally. Therefore, such research has focused on students' relative rank-order position in their self-concept and their relation to their relative rank-order position in their future achievement (and vice versa). Such cross-lagged panel models do not allow researchers to properly distinguish intraindividual (within-person) processes from stable interindividual (between-person) differences (Ehm et al., 2019;

Hamaker et al., 2015). This is problematic because major theoretical models on the longitudinal relations between ASC and achievement (e.g., the reciprocal effects model) clearly focus on self-enhancement and skill-development effects and thus on motivational within-person processes.

Indeed, in their recent attempt to disentangle within-and between-person processes in drawing on four measurement occasions bridging 3.5 years in German primary school children using a random intercept cross-lagged panel model, Ehm et al. (2019) found no evidence of self-enhancement effects, contradicting assumptions of the reciprocal effects model. However, the authors acknowledged the need for future research. Notably, the lack of self-enhancement effects might be due to the relatively long lags between their measurement occasions. In order to detect cross-lagged effects, optimal time lags should be rather short so that within-person processes can be examined while controlling for interindividual differences (Dormann & Griffin, 2015; see Ehm et al., 2019).

To this end, researchers need to shift toward a short-term perspective. Studying longitudinal short-term relations is typically associated with intensive longitudinal (or microlongitudinal) data such as those obtained with experience sampling to study students' experiences in their everyday life at school (e.g., Hamaker & Wichers, 2017; Trull & Ebner-Priemer, 2014; Zirkel et al., 2015). Intensive longitudinal data is characterized by its focus on within-person regulatory mechanisms and associated dynamics as well as its short-term perspective in typically drawing on multiple measurement time points that are close together (McNeish & Hamaker, 2020).

Shifting toward within-person and short-term perspectives in studying dynamics between ASC and achievement with experience sampling is crucial for several reasons. Most importantly, between-person and within-person relations between variables are most likely to be statistically independent unless the (unlikely) assumption of ergodicity holds (Molenaar, 2004; Murayama et al., 2017). Importantly, this implies that research findings on reciprocal effects between ASC and achievement may actually not hold when shifting toward an everyday, within-person perspective (see Ehm et al., 2019). Also, in collecting intensive longitudinal data, the researcher obtains a relatively large number of repeated measurements of interest variables in real time and real life. As such, they shift their perspective from merely a trait-based perspective toward a state-based perspective.

Specifically, in the present study, we applied an experience-sampling, e-diary methodology to obtain lesson-specific—or state— measures of MSC and perceived achievement (i.e., self-reported comprehension) for every single lesson in mathematics. We suggest that this represents repeated snapshots (Hamaker & Wichers, 2017) of students' ASC and perceived achievement over time (described in the Method section). Notably, state assessments capture a broader range of momentarily perceived situations and may be less biased than conventional trait-based self-reports (Trull & Ebner-Priemer, 2014). When research on the reciprocal relations between ASC and achievement exclusively relies on only a few assessment occasions that bridge longer time spans, such an approach is arguably insufficient for capturing the bandwidth of the dynamics that students actually perceive in their everyday life at school. For instance, students who generally perform well in mathematics may experience difficulty understanding the material in particular lessons or may fail to complete some tasks from time to time. Conversely, students who generally struggle in mathematics may feel that they were able to follow a particular lesson well or may sometimes feel that they understood the material the teacher went over. In addition, feedback from teachers and peers, which students can use to infer their current performance, can also vary across lessons. Such dynamics arguably remain undetected in long-term studies in which longitudinal assessments are separated by months or years. We know from previous research that even allegedly well-established notions had to be revised in the field after a state-based perspective was adopted (see, e.g., Goetz et al., 2013; research on state vs. trait mathematics anxiety).

Notably, the intraindividual, within-person approaches used in the present study also capture differences *between* persons. As such, they enable researchers to potentially reveal interindividual differences across the observed within-person relations, or stated differently, the heterogeneity of the functional within-person relations between the variables of interest (see, e.g., Pekrun et al., 2002).

The Present Study

In the present experience-sampling study, we drew on e-diary data collected in German secondary schools across a time-period of 3 weeks to uncover real-life and real-time dynamics between students' ASC and their lesson-specific perceived achievement in the domain of mathematics. This study is the first to revisit the reciprocal relations between ASC and (perceived) achievement in everyday life in shifting toward a state-based, within-person perspective. The overarching aim of this study was to examine the existence and significance of self-enhancement and skill-development effects when studying students in every single lesson in a given domain at school. In focusing on the mathematics domain, we studied everyday relations between ASC and perceived achievement within the arguably most frequently analyzed domain in ASC research (Marsh, 2006), thus enabling us to better compare and embed our results into the existing between-person research.

As mentioned earlier, previous between-student research on the reciprocal relations between ASC and achievement has deployed different indicators and proxies for measuring achievement (Huang, 2011; Valentine et al., 2004; Wu et al., 2021). Standardized achievement tests or report-card grades may be the indicators of choice when drawing on panel designs spanning longer time intervals (Marsh & Martin, 2011). However, when researchers examine intraindividual skill-development and self-enhancement effects in real time and real life, students' achievement should be assessed in an ecologically valid way on a lesson-to-lesson basis. In comparison with conventional panel designs, e-diary designs allow researchers to collect longitudinal data in a natural, spontaneous context (Reis, 2014) in a far less intrusive manner with fewer barriers (McNeish & Hamaker, 2020). To measure students' cognitive learning outcomes, previous e-diary studies have widely employed self-reports of learning or perceived achievement (see, e.g., Giannakos et al., 2020; Peterson & Miller, 2004; Shernoff, Sannella, et al., 2017) instead of more objective (but arguably also more intrusive) daily measures of achievement, such as standardized tests. The use of self-reports of learning or perceived achievement has a long tradition (e.g., Richmond et al., 1987). Previous research has shown that students are able to accurately assess their own learning (e.g., Brown et al., 2015; Chesebro & McCroskey, 2000; Ross, 2006), and measures of students' perceived achievement have been widely used as a valid way to measure students' cognitive learning outcomes (e.g., Kurucay & Inan, 2017; Rovai et al., 2009; Shin, 2003; Yoon et al., 2020). In the present study, we asked students to indicate their comprehension and their learning progress for every single mathematics lesson. Students thus reported in real time how well they understood the material that they had just gone through in class. This understanding indicates students' lesson-specific perceived achievement with respect to what they were supposed to learn (see the Method section below).

However, it is important to note that although we built on existing research on the reciprocal effects model (Marsh, 1990; Marsh & Craven, 2006; Wu et al., 2021), the present study should not be understood as a direct replication of the classic reciprocal effects model at the within-person and short-term levels. Standardized achievement test scores or report card grades have been recommended to test the reciprocal effects model (e.g., Marsh & Martin, 2011). In contrast, perceived achievement measures were used in the present study.

Specifically, our overarching aim resulted in two focal research questions, which are both located on the within-person, intraindividual level:

RQ 1: Is there a positive and significant path from previous perceived achievement to subsequent MSC (i.e., the skill-development effect) on a lesson-to-lesson basis?

RQ 2: Is there a positive and significant path from previous MSC to subsequent perceived achievement (i.e., the self-enhancement effect) on a lesson-to-lesson basis?

Both research questions are critically important for ASC theory and research because of the lack of within-person research that has applied experience sampling rather than the between-person approach used in most research (see Marsh & Craven, 2006). In addition to our two focal research questions, we aimed to examine interindividual (between-person) differences in the observed intraindividual relations between ASC and perceived achievement. Specifically, we explored whether interindividual differences in the observed within-person associations between MSC and perceived achievement could be explained by students' gender, reasoning ability, or mathematics grades. This resulted in our third research question, which focused on the between-person, interindividual level:

RQ 3: Are everyday skill-development and self-enhancement dynamics generalizable across or moderated by students' gender, reasoning ability, and mathematics grades?

Gender, reasoning ability, and mathematics grades have all been shown to play predominant roles in the formation of MSC: Students with higher reasoning ability and those obtaining better report-card grades in mathematics typically report higher levels of MSC (Möller et al., 2020), whereas gender disparities in MSC to the disadvantage of girls and young women have repeatedly been found in previous research regardless of students' actual mathematics performance (e.g., Frenzel et al., 2007; Niepel et al., 2019). Overall, previous between-person research on longitudinal relations between ASC and achievement (e.g., Valentine et al., 2004; Wu et al., 2021) led us to expect to find state-based skill-development (RQ 1) and self-enhancement effects (RQ 2) between MSC and perceived achievement in students' everyday life. Further, previous between-person findings on the generalizability of skill-development and self-enhancement effects across gender and ability levels (e.g., Marsh et al., 2005; Seaton et al., 2015; Valentine et al., 2004) led us to expect that both effects would be largely generalizable across gender, reasoning ability, and obtained mathematics grades (RQ 3). We applied multilevel confirmatory factor analyses (MCFA) and dynamic structural equation modeling (DSEM; Asparouhov et al., 2018; see the Method section below) to address our research questions.

Method

Procedure and Participants

In the present study, we drew a sample of $N = 372$ students (34.1% young men, based on $n = 301$) whose data were collected as part of the larger "Dynamics of Academic Self-Concept in Everyday Life" (DynASCEL) project,[1] which focused on the everyday dynamics of ASC. Students attended one of 18 different classes at six different academic-track schools (Gymnasium) in Grade 9 ($n = 308$) or Grade 10 ($n = 64$) in four different federal states of Germany (Baden-Württemberg, Mecklenburg-Vorpommern, Nordrhein-Westfalen, Rheinland-Pfalz). The average self-reported age was 15.3 years ($SD = .68$; range: 13.3 to 17.4; $n = 298$). Students' participation was voluntary, and written parental consent was obtained for all participating students. Students could skip prompts or single questions. All procedures were approved by the ethics review panel of the University of Luxembourg and by all involved education authorities.

The data collection took part over a 5-week period at the respective schools. In Week 1, students completed a background inventory (paper-pencil format). In Weeks 2 to 4, students were given a smartphone as a hub for experience sampling over 3 weeks (e-diary approach). In Week 5, students completed a shorter postquestionnaire (paper-pencil format). In the present study, we focused on the e-diary data on state MSC and perceived achievement in every single mathematics lesson that we collected across the 3-week period from Weeks 2 to 4. In addition, we drew on data collected in Week 1 (i.e., background inventory) to obtain information about students' gender, reasoning ability, and mathematics grades (see the Measures section below).

To obtain data on state MSC and perceived achievement in every single mathematics lesson, students received prompts through an auditory signal that asked them to complete a brief electronic questionnaire on the smartphone at the end of each mathematics lesson. Prompts were programmed using the movisensXS software (Movisens GmbH, 2017) following the timetable of each respective class with the number of mathematics lessons (i.e., measurement points) thus varying from class to class (M = 10.11 mathematics lessons; SD = 3.39; range: 3 to 16). As can be expected from intensive longitudinal designs, there are many reasons for missing values. These included, for example, students' or mathematics teachers' sick leave, exams, excursions, or other obligations, as well as technical issues (e.g., dead battery, students left smartphone at home). Per design, students were instructed not to respond to prompts when they had not had classroom instruction or were absent from school. We obtained responses to 2,702 prompts (students * mathematics lessons), representing 71.4% of the previously programmed prompts. Of these 2,702 accepted prompts (i.e., with at least one item answered), 97.5% provided complete data (i.e., six items per prompt; see Measures section below).

Measures

E-Diary

State Mathematics Self-Concept. We assessed students' state MSC after every single mathematics lesson across a time-period of 3 weeks. State MSC was assessed with three items based on the Self-Description Questionnaire (SDQ; Marsh et al., 1983), which is considered to be one of the best self-concept instruments available (e.g., Byrne, 2002). Three-item short-form (trait) MSC instruments based on the SDQ have been shown to be psychometrically sound for educational research purposes (Gogol et al., 2014) and are commonly used in longitudinal MSC research (e.g., Marsh et al., 2015, 2018; Möller et al., 2011; Niepel et al., 2014b). Specifically, we adapted the MSC items to the specific demands of experience sampling by beginning every item with the passage "Currently, I think that   " Students responded on a 6-point Likert scale ranging from 0 (*false*) to 5 (*true*). The item wordings were "Currently, I think that I am good at mathematics," "[ ]work in mathematics is easy for me," and "[ ] I learn quickly in mathematics," such that higher item scores indicated higher state MSC. The original German-language item wordings are listed in Table S1 in the online supplemental material.

Lesson-Specific Perceived Achievement in Mathematics. As we did for state MSC (see the previous section), we assessed students' perceived achievement in terms of their lesson-specific comprehension and learning progress after every single mathematics lesson across a time-period of 3 weeks. Three items were used to assess perceived achievement. Students responded on a 6-point Likert scale, ranging from 0 (*false*) to 5 (*true*). The item wordings were "I was able to follow the last lesson well," "I understood a lot in the last lesson," and "I learned a lot in the last lesson," such that higher item scores indicated better lesson-specific perceived achievement in mathematics. Similar items have been commonly used in e-diary studies (e.g., Peterson & Miller, 2004; Shernof et al., 2017; Shernoff, Ruzek, & Sinha, 2017; Shernoff, Sannella, et al., 2017) as well as in previous between-person research (e.g., Yoon et al., 2020) to measure perceived (learning) achievement (see Richmond et al., 1987). The original German-language item wordings are listed in Table S1 in the online supplemental material.

Background Inventory

Reasoning Ability. We applied the Intelligenz-Struktur-Test-Screening (IST-Screening; Liepmann et al., 2012) to assess students' reasoning ability. The IST-Screening is an economic (less than 30 min) reasoning ability measure that includes three groups of tasks consisting of verbal analogies, number sequences, and figural matrices (each consisting of 20 items). It is based on the Intelligenz-Struktur-Test (IST; Amthauer, 1970; Liepmann et al., 2007), an intelligence test that is widely used in Germany (Schmidt-Atzert & Amelang, 2012). The IST-Screening exists in two parallel versions, A and B; in the present study, we used Version A. It was presented in a paper-pencil format in the week before the e-diary assessment began. Liepmann et al. (2012) reported an internal consistency of $a = .87$ for the full-scale reasoning ability composite score. In our subsequent analyses, we used the full-scale reasoning ability composite raw score; the observed reliability in the present study was $x = .77$ ($a = .76$).

Mathematics Grades. Students reported their grades in mathematics as obtained from their most recent report card. Research on the validity of self-reported grades in Germany suggests that self-reported school grades are not subject to systematic reporting biases (Dickhäuser & Plenter, 2005; Sparfeldt et al., 2008). In Germany, a 6-point grading system is used; we reverse-scored the grades in the present study such that higher values indicated better school grades in mathematics (i.e., ranging from 1 = *unsatisfactory* to 6 = *very good*).

Data Analysis

Multilevel Confirmatory Factor Analysis (MCFA)

Prior to our main analyses, we conducted a MCFA of the intra-individual e-diary data using the statistical software Mplus 8.3 (Muthén & Muthén, 1998–2019) to inspect the psychometric properties of the e-diary measures (Kim et al., 2016). Utilizing MCFA, we accounted for timepoints nested within students by explicitly modeling the factor structure on the within-person level (Level 1; i.e., state-like factors) as well as on the between-person level (Level 2; trait-like factors). To control for class-level effects, we included 17 dummy variables (based on 18 classrooms) at Level

2.2 We used the Mplus MLR estimator, which is robust against mild violations of normality and allowed us to deal with missing data (Kaplan, 2009).

In a first step, we calculated intraclass correlations (ICCs) to estimate the amount of variation in MSC and perceived achievement across Levels 1 and 2. Second, we tested a series of different models to analyze the measures' factor structure across both levels. To evaluate the model fit, we considered the comparative fit index (CFI), the Tucker-Lewis index (TLI), the root mean square error of approximation (RMSEA), and the standardized root-mean-square residual (SRMR). We used the recommended cut-off values (CFI $.95; TLI $ .95; RMSEA #.06; SRMR # .08; Kline, 2005).

To examine the model fit for each level separately, we employed partially saturated models (Janis et al., 2016). In doing so, we specified (a) the hypothesized two-factor structure as well as (b) a more parsimonious one-factor structure at Level 1, while specifying a saturated model (i.e., item variances and covariances only) at Level 2 and vice versa (c and d). In a third step, to ensure a meaningful interpretation of the constructs across levels (Stapleton et al., 2016), we tested for cross-level invariance by restricting the factor loadings of the corresponding items to be equal across Levels 1 and 2, and we freely estimated the factor variances at Level 2 (Jak & Jorgensen, 2017). Competing models were compared based on decreases in model fit and differences in the Akaike information criterion (AIC) and Bayes information criterion (BIC) with a preference given to the model with the lower value. Finally, we calculated level-specific reliabilities in terms of the Level 1 and Level 2 omega coefficients in freely estimating all factor loadings and fixing the factor variances to 1 at both levels (Geldhof et al., 2014).

Dynamic Structural Equation Modeling

To address our research questions, we conducted DSEM (Asparouhov et al., 2018) in Mplus 8.3 (Muthén & Muthén, 1998–2019). Before we began, we ensured that there were no mean trends in state MSC and perceived achievement (i.e., neither variable consistently increased or decreased over the 3-week period) that needed to be incorporated into our model (McNeish & Hamaker, 2020). To this end, we calculated a linear regression with time as the only predictor at Level 1 (McNeish & Hamaker, 2020) in Mplus. The effects were close to zero and not statistically significant (for state MSC: b = .03, p = .222; for perceived achievement: b = -.04, p = .091), suggesting no linear trend over time.

We specified a multilevel first-order vector autoregressive (VAR(1)) model (Hamaker et al., 2018), a multilevel extension of a time series model. The VAR(1) model can also be thought of as a multilevel extension of a cross-lagged panel model that allows for interindividual differences in means and lagged effects (Hamaker et al., 2018, p. 826). Data were decomposed into within-person (Level 1) and between-person (Level 2) components. At Level 1, we specified the cross-lagged relations between state MSC and lesson-specific perceived achievement (i.e., by using manifest mean scores of the three indicators at each time point). We regressed state MSC (MSC_t) on lesson-specific perceived achievement at the previous time point ($Achievement_{t-1}$) to test the skill-development effects on a lesson-to-lesson basis (RQ1). We regressed lesson-specific perceived achievement ($Achievement_t$) on state MSC at the previous time point (MSC_{t-1}) to test for self-enhancement effects on a lesson-to-lesson basis (RQ2). For the autoregressive paths from MSC_{t-1} ($Achievement_{t-1}$) to MSC_t ($Achievement_t$) at Level 1, it is important to note that these paths indicate the amount of carryover (or inertia) from one lesson to the next for each student (Hamaker et al., 2018). Therewith, these two autoregressive paths indicate how quickly students return to their habitual, trait-like MSC (or habitual level of perceived achievement) after experiencing situation-specific ups and downs in their state MSCs (perceived achievements). Put differently, the larger the carryover across mathematics lessons, the more the current state depends on the previous lesson's state, and the longer it takes to return to the trait level. Students' mean levels for MSC and perceived achievement, which can be interpreted as their trait scores, are modeled at Level 2 (see next paragraph).

At Level 2, we estimated the interindividual variances, fixed effects, and intercorrelations of six variables. More specifically, we estimated the two mean values for MSC and perceived achievement (i.e., trait scores for MSC and perceived achievement), the two autoregressive parameters for MSC and perceived achievement (indicating carryover), the skill-development effect from perceived achievement to MSC, and the self-enhancement effect from MSC to perceived achievement. DSEM is based on Bayesian estimation, and missing data are sampled from their conditional posterior in this kind of analysis. For our analyses, we used Mplus' default priors. As the time intervals between consecutive measurement points (i.e., mathematics lessons) varied in accordance with each class's timetable, we controlled for the time intervals by using the TINTERVAL option implemented in Mplus. To this end, we specified a time interval variable, which indicated the time difference for every measurement point in hours from the very first prompt per class. The Mplus code for our specified VAR(1) model can be found in the online supplemental material.

In the last step, we explored whether interindividual differences in the observed within-person associations between MSC and perceived achievement were generalizable across persons in relation to gender, reasoning ability, and grades (RQ3). To this end, we saved the factor scores of the Level 2 (skill-development and self-enhancement) effects by using a multiple imputation approach (Graham et al., 2003) to impute 50 data sets in Mplus containing the imputed values for the factor scores. The imputed data sets were then used to calculated bivariate correlations between the factor scores and gender, mathematics grades, and reasoning ability using maximum likelihood estimation. We chose this approach to decrease model complexity and to ensure model convergence. To handle missing values in the student background variables, most of which occurred due to technical problems (percentages of missing values: 19.1% for gender; 17.7% for reasoning; 20.7% for mathematics grades), we used full-information maximum likelihood (FIML) implemented in Mplus.

Results

Preliminary Analyses

Before we computed the main analyses, we examined the ICCs, the latent factor structure, the cross-level invariance, and the reliabilities of the intraindividual e-diary measures (state MSC and lesson-specific perceived achievement) by means of an MCFA. The ICCs were $ICC_{ACH1} = .384$, $ICC_{ACH2} = .373$, and $ICC_{ACH3} =.370$ for the perceived achievement items, and $ICC_{MSC1} = .744$, $ICC_{MSC2} = .737$, and $ICC_{MSC3} = .739$ for the MSC items. The ICCs indicated that students' perceived achievement showed stronger intraindividual variation across the observed 3-week period than MSC did. Most of the variance in perceived achievement originated from Level 1 (variation within students). In contrast, most of the variance in MSC originated from Level 2 (variation between students).

Table 1 presents the results of testing alternative factor structures across the two levels. Inspections of the fits of the partially saturated models suggested a very good approximation to the data for the two-factor solutions on both Levels 1 and 2 (see Table 1, Models a and c). By contrast, the more parsimonious one-factor solutions exhibited poor fits at both levels (Models b and d). In addition to the partially saturated models, we ran further MCFAs to inspect the latent factor structure. As such, we specified Models e and f, which assumed two latent factors on one level but only one general latent factor at the other level (see Table 1). Finally, we specified Model g, which assumed a two-factor structure at both levels. The fit indices, as well as differences in the AIC and BIC, indicated that model g provided a better approximation to the data than Models e or f (see Table 1), in line with the results we got from the partially saturated model test approach (Models a to d). Together, our tests pointed to a two-factor structure at both levels, indicating that MSC and perceived achievement were empirically distinguishable constructs at the within-and between-person levels.

After establishing the two-factor structure at both levels, we tested for cross-level invariance in a next step. To this end, we built upon Model g (see Table 1), which consisted of two intercorrelated latent factors (MSC and perceived achievement) with three indicators per construct. The six items were constrained to load only on their respective factor, and the latent constructs were not allowed to correlate across levels. In specifying Model h, we restricted the factor loadings of the corresponding items to be equal across Levels 1 and 2. The overall fit suggested an excellent overall approximation to the data (see Table 1). Compared with Model g, no meaningful decreases in the CFI, TLI, RMSEA, or SRMR could be detected; we observed higher AIC but lower BIC values. Overall, the results suggested that the assumption of cross-level invariance held for both constructs. The factor loadings ranged from $k = .742$ to $k = .934$ for perceived achievement and from $k =.795$ to $k = .845$ for MSC at Level 1. They ranged from $k = .708$ to $k = .917$ for perceived achievement and from k = .959 to k =.972 for MSC at Level 2 (all *p*s , .001). The latent correlations between MSC and perceived achievement were q = .646 (*p* ,.001) at Level 1 and q = .863 (*p* , .001) at Level 2 (the coefficients came from Model h).

Reliability was tested by calculating McDonald's x on Levels 1 and 2, indicating good reliabilities of x = .863 for MSC and x =.894 for perceived achievement at Level 1 and of x = .996 for MSC and x = .950 for perceived achievement at Level 2.

Main Analyses

We examined our three research questions using a multilevel VAR

(1) model for MSC and perceived achievement within the DSEM framework. We used 100,000 Markov chain Monte Carlo (MCMC) iterations with two Markov chains. The results here were based on 10,000 iterations because the first half of each chain was discarded as burn-in, and a thinning of 10 iterations was used (i.e., only one in 10 iterations was saved; Gelman et al., 2014; see Hamaker et al., 2018). Model convergence was evaluated by applying the potential scale reduction criterion (PSR; Asparouhov & Muthén, 2010). PSR is the ratio of the total variance across chains and the pooled variance within a chain. We used PSR , 1.05 as an appropriate convergence criterion (Gelman & Rubin, 1992). The DSEM resulted in good convergence. PSR values were below 1.05 for each parameter. Moreover, the trace plots for each parameter did not indicate any signs of nonconvergence.

Figure 1 represents the path diagram for the within-person part of the model and depicts the observed results (i.e., standardized parameters and their 95% credible intervals). To address our two focal research questions, RQ 1 and RQ 2, we looked at the cross-lagged relations. We found intraindividual effects from perceived achievement to MSC at the next mathematics lesson (ACH_{t-1} to MSC_t) and from MSC to perceived achievement at the next mathematics lesson (MSC_{t-1} to ACH_t). None of the parameters' CIs contained zero. These results point to the existence and significance of skill development effects (RQ 1) and self-enhancement effects (RQ 2) in students' everyday life at school. Specifically, we found average individually standardized effects at the within-person level of .186 (95% CI [.117, .257]) for skill-development effects and of .052 (95% CI [.003, .095]) for self-enhancement effects.

Concerning the autoregressive paths, the average individually standardized autoregressive effects were stronger for perceived achievement than for MSC (see Figure 1). This suggests that students have more carryover in their lesson-specific perceived achievement from one lesson to the next than is the case for their state MSC (i.e., carryover effects). Furthermore, the average correlation between the within-person residuals of perceived achievement and MSC was .484 (95% CI [.438, .534]; see Figure 1). The averaged proportion of explained variance on the within-person level in our model was .222 (95% CI [.198, .250]) for lesson-specific perceived achievement and .326 (95% CI [.288, .359]) for state MSC.

Table 2 depicts interindividual, Level 2 means, variances, and cor relations. Looking at the variance estimators, we found that students differed not only in their (trait-like) MSC and (trait-like) perceived achievement values but also in the strengths of their (trait-like) individual skill-development and self-enhancement effects (i.e., between-person variances of the within-person cross-lagged relations across mathematics lessons). Seven out of the 15 correlations (see Table 2) were statistically significant (their 95% credible intervals did not contain zero). The two mean values were positively intercorrelated, indicating that students higher in MSC also tended to report higher levels of perceived achievement. Further, we found that the skill-development and self-enhancement effects were negatively related. This indicates that students who experienced stronger self-enhancement effects tended to experience weaker skill-development effects. The remaining five correlations involved the autoregressive (carryover) effects for perceived achievement (ACH_{t-1} ! ACH_t) or MSC (MSC_{t-1} ! MSC_t). Specifically, mean MSC and mean perceived achievement were both negatively correlated with the autoregressive effects for perceived achievement. This indicates that students with a higher trait level for MSC (or for perceived achievement) also tended to have less carryover in their lesson-specific perceived achievements from one lesson to the next. Further, skill-development effects (ACH_{t-1} ! MSC_t) were positively correlated with autoregressive effects for perceived achievement (ACH_{t-1} ! ACH_t) and negatively correlated with autoregressive effects for MSC (MSC_{t-1} ! MSC_t). This indicates that students who had stronger skill-development effects also tended to have more carryover in their lesson-specific perceived achievement from lesson to lesson. At the same time, students who had stronger skill-development effects tended to have less carryover in their state MSCs. Finally, we found that the self-enhancement effects (MSC_{t-1} ! ACH_t) were positively related to the autoregressive effects for MSC (MSC_{t-1} ! MSC_t). This indicates that students who had stronger self-enhancement effects also tended to have more carryover in their state MSCs from one lesson to the next. Students who were generally high versus low in trait MSC (and perceived achievement) did not seem to differ in how often they experienced skill-development or self-enhancement effects: We observed no relation between students' mean values and the cross-lagged parameters.

To address RQ 3, in the last step we calculated correlations between the factor scores of the six Level 2 variables (obtained from the DSEM we conducted as described in the previous paragraph) and the student background variables (gender, mathematics grades, and reasoning ability). Gender was not significantly related to reasoning ability (r = –.109, p = .057; 0 = young men; 1 = young women) or to grades (r = –.084, p = .152); reasoning ability and grades showed a significant positive relation (r =.389, p , .001). Table 3 depicts the observed correlations between factor scores and background variables. Importantly, we did not find any significant correlations between either the self-enhancement or the skill-development effects with the inter-individual student characteristics. This suggests that both effects operated independently from gender, reasoning ability, and obtained mathematics grade (RQ 3). However, we did find significant relations for students' mean MSC and mean perceived achievement with gender, reasoning, and grades. These results indicate that students who had higher reasoning test scores and obtained better mathematics grades tended to have higher trait levels on both MSC and perceived achievement. Male students tended to have higher state MSCs and lesson-specific perceived achievements than female students. However, we did not find any significant association between students' carryover effect for MSC and any of the interindividual characteristics we examined. Students' carryover effects for perceived achievement were significantly negatively correlated with their mathematics grade but not with gender or their reasoning test scores. This indicates that students who obtained better mathematics grades tended to have less carryover in their lesson-specific perceived achievement from one lesson to the next.

Discussion

Previous research on longitudinal reciprocal relations between ASC and achievement has emphasized interindividual differences between students averaged across relatively long time-periods. Our study is apparently the first to evaluate this issue by exploring intraindividual relations between ASC and perceived achievement over short time-periods with an experience-sampling approach. Our overarching aim was to examine the self-enhancement and skill-development effects on students' concrete learning situations in school. Further, we examined whether these effects were moderated by or generalized across students' gender, reasoning ability, or mathematics grades.

Dynamics Between Mathematics Self-Concept and Perceived Achievement From Lesson to Lesson

We found that both self-enhancement and skill-development effects operated to form students' self-concept and perceived achievement on a lesson-to-lesson basis. Our results are thereby in line with predictions derived from existing interindividual research on the reciprocal effects model (e.g., Marsh & Martin, 2011; Wu et al., 2021). However, our study differs considerably from previous research. As highlighted earlier, self-development and self-enhancement effects are formation processes at the intraindividual level, but they have been studied almost exclusively at the between-person level. The present study is apparently the first to show the within-person effects of self-concept on perceived achievement (and vice versa) in real time in actual learning situations. As such, we provide evidence that a higher ASC leads a student to better follow and understand the learning material in concrete learning situations in real time. Also, a student uses their concrete learning experience in real time to shape their self-concept.

The sizes of the skill-development and self-enhancement effects observed in the present study were comparable to the effect sizes reported in the meta-analysis on the reciprocal effects model by Wu et al. (2021) for trait-like self-concept and achievement, with descriptively stronger skill-development effects than self-enhancement effects. According to typical guidelines for the evaluation of effect sizes (Cohen, 1988; Kline, 2005), the observed effects could be interpreted as small and possibly even negligible (but see also Gignac & Szodorai, 2016). However, we would like to emphasize that both effects are incremental (i.e., in controlling for lesson-specific carryover effects and interindividual differences in MSC and perceived achievement), mutually reinforcing across time, and refer to relatively short time intervals from one school lesson to another.

Another observation from our applied within-person perspective

was that the residuals of the two within-person measures (i.e., state MSC and lesson-specific perceived achievement) showed a substantial contemporaneous correlation. This correlation might be due to both potential (unobserved) third variables in the within-part of our model (e.g., lesson-specific demands, social interactions in the classroom, students' mood) and potential reciprocal effects between the two variables, which may have occurred within school lessons (Hamaker et al., 2018).

Interindividual Differences in the Dynamics Between Mathematics Self-Concept and Perceived Achievement

The present study examined not only intraindividual relations between state MSC and perceived achievement but also whether there were interindividual differences in these intraindividual relations. Our results suggest that students differ in the extent to which they experience self-enhancement effects (current situation-specific perceived achievement is influenced by state MSC from the previous lesson) and skill-development effects (current state MSC is influenced by the student's level of perceived achievement in the previous lesson) in their daily school life. As such, we provided evidence that students differ in the extents to which (a) their ASC is influenced by their daily learning performance (interindividual differences in skill-development effects) and (b) their ASC influences their daily learning performance (interindividual differences in self-enhancement effects).

Further, the self-enhancement and skill-development effects were negatively correlated with each other. Thus, students who experience more of one tend to experience less of the other. Importantly, this is a substantive new finding with significant implications that cannot be readily examined with the traditional (between-person) approaches that are typically used to study reciprocal relationships between ASC and achievement. Thus, implicit in the traditional approach is the assumption that the relative strengths of the cross-lagged paths are the same across different students (this holds true not only for cross-lagged panel models but also for more recent approaches using random intercept cross-lagged panel models; see, e.g., Ehm et al., 2019). However, our within-person evaluation suggests that this assumption is wrong. Not only were there differences in the relative strengths of the self-enhancement and skill-development effects, but within each student, the two effects tended to counterbalance each other. These findings have important implications for the design of studies that are intended to enhance ASC, achievement, or both constructs. Our study suggests that such interventions must account for interindividual differences in students' individual propensity for self-enhancement or skill-development effects to be more effective.

The interindividual student characteristics that we examined—gender, reasoning ability, and mathematics grades—seem to be unrelated to differences in skill-development and self-enhancement effects: We found these differences to be independent from gender, reasoning ability, and grades. These findings were in line with our expectations, which were based on previous between-person research. Specifically, previous research has shown that gender does not moderate relations between achievement and ASC (Valentine et al., 2004). The present study offers some first tentative support for this finding with intensive longitudinal data. In a similar vein, previous between-person research has provided evidence for the generalizability of skill-development and self-enhancement effects across ability levels (e.g., Gorges et al., 2018; Seaton et al., 2015). The present study supported these findings in providing initial evidence that lesson-to-lesson skill-development and self-enhancement effects generalize across different levels of reasoning ability and mathematics grades. Notably, in their recent meta-analysis on the longitudinal relations between trait-like ASC and achievement, Wu et al. (2021) found that skill-development effects generalized across different achievement levels. In at-risk and poor-performing samples, however, they found evidence that self-enhancement effect sizes were weaker than in unselected samples. Future research that is aimed at replicating our results in lower performing samples is therefore warranted.

Carryover Effects From Lesson to Lesson

We found evidence for within-person carryover effects for state MSC and perceived achievement from one lesson to the next. These findings suggest that situation-specific ups and downs in students' state MSC continue to affect the next lesson before students return to their habitual levels of MSC. Similarly, it suggests that situation-specific ups and downs in students' lesson-specific perceived achievement continue to affect the next lesson before students return to their typical levels of perceived achievement. Further, our results suggest that students differ in the extent to which they experience carryover effects. Again, this is a finding that could not be tested with typical between-person approaches. As such, we provided evidence that students differ in how strongly their current state self-concept (lesson-specific perceived achievement) depends on their previous lesson's state and how long it takes to return to their habitual trait level.

Here, we found some relations with specific student characteristics. Concerning carryover effects for perceived achievement, students with higher trait MSC levels, higher mean perceived achievement levels, and better mathematics grades tended to have less carryover in their perceived achievements from one lesson to another. Such students thus seemed to return to their typical levels of perceived achievement more quickly than students with less confidence in their own abilities and lower achievement levels. Conversely, students with stronger skill-development effects seemingly tended to have more carryover in their perceived achievement levels from lesson to lesson. Thus, students who do not return to their typical levels of perceived achievement as quickly as their peers after experiencing lesson-specific ups and downs tend to simultaneously experience stronger daily influences of their perceived achievements on their self-concepts.

Similarly, with regard to carryover effects for state MSC, we found that students with stronger self-enhancement effects seemingly tended to have a stronger carryover for state MSC from lesson to lesson. Thus, students who do not return to their typical levels of self-concept as quickly after experiencing lesson-specific ups and downs tend to simultaneously experience stronger daily influences of their self-concepts on their perceived achievements. Conversely, students with weaker skill-development effects seemingly tend to have a stronger carryover for state MSC from lesson to lesson. Thus, students who do not return to their typical levels of self-concept as quickly after experiencing lesson-specific ups and downs tend to simultaneously experience weaker daily influences of their perceived achievements on their self-concepts. Overall, our study points to the need for further research on carry-over effects in ASC research in general, their role in ASC formation, and their relations to students' characteristics.

State Academic Self-Concept

ASC is commonly considered a construct that is characterized by high stability, even over long periods of time (Jansen et al., 2020; see also Orth et al., 2021). However, our results provide strong evidence for substantial situation-specific fluctuations in ASC as experienced by students in their everyday life at school. These results are consistent with hallmark conceptual work on self-concept that already described situation-specific aspects of self-concept (James, 1890, p. 307; Shavelson et al., 1976) as well as with existing empirical work related to our study (e.g., Malm-berg & Martin, 2019; Tsai et al., 2008). The positive correlations we found between students' averaged state MSC and their reasoning ability and mathematics grades, as well as the observed gender differences in averaged state MSC in favor of male students, corresponded to previous findings on the nomological network of (trait) MSC, thus providing further evidence for the construct validity of state MSC.

Shavelson et al.'s (1976) multidimensional, hierarchical model of self-concept is nowadays often considered to be the starting point of modern empirical research on ASC (Marsh, 2006; Marsh et al., 2019). In the original Shavelson et al. (1976) model, there was an implicit hierarchy of stability, from general school ASC (e.g., "I am good at most school subjects") at the apex, to domain-specific ASCs (e.g., "I am good at mathematics") at the next level, with even more task-specific ASCs at lower levels of the hierarchy. At the base of this hierarchy, however, the authors claimed a situation-specific level, where "self-concept varies greatly with variation in situations" (Shavelson et al., 1976, p. 414). Apparently, this level was thought to reflect state ASCs.

e propose a need to conceptually disentangle the dimension of domain level or task specificity in ASC from the situation-specificity dimension in ASC. Whereas the first dimension can be thought of as falling on a continuum that ranges from more general school to more task-specific (e.g., general school ASC vs. MSC), the latter can be thought of as falling on a trait-to-state continuum (e.g., trait ASC vs. state ASC). Notably, many person variables typically regarded as between-person phenomena have been shown to exhibit substantial within-person variability (Podsakoff et al., 2019)—including, for instance, supposedly stable traits such as the Big Five (Fleeson, 2001). The notion that, in principle, any person variable can be thought of on a trait-to-state continuum and could potentially even be operationalized as both a trait and a state has been advocated for and has garnered ample empirical support (Rauthmann, 2021). With this in mind, it would make sense to assume that even the general school ASC can in principle also be operationalized as both a trait and a state. In future ASC theory and research, we thus suggest that the two dimensions—domain-level (or task) specificity and situation (or temporal) specificity—should be thought of as conceptually distinct from each other. Further research is needed to explicitly integrate the dimension of situation specificity into a future, expanded taxonomy of ASC.

Limitations

As is characteristic of intensive longitudinal approaches, our experience-sampling data can be characterized as highly ecologically valid. Our data were temporally ordered by design, thus allowing for a fine-grained assessment of students' real-time experiences in their everyday life. Nevertheless, the omitted variable problem persisted for the within-person part, according to which a time-varying variable not included in the analyses could have caused the observed relations (Hamaker et al., 2018). As such, support for implicit causal interpretations must be made with appropriate caution (Grosz et al., 2020).

Furthermore, we asked the students about their perceptions of their everyday learning progress for each individual mathematics lesson to operationalize students' lesson-specific perceived achievements. We acknowledge that this particular operationalization represents only an approximation of students' actual achievement in that lesson. In ASC research, report card grades or standardized achievement tests are often used to represent the achievement component. In an experience-sampling design, both measures are arguably not or only partially applicable as daily assessments while maintaining ecological validity. Our empirical examination of the factorial validity demonstrated that situation-specific, state ASC can clearly be empirically distinguished from students' perceptions of their lesson-specific achievement. The latter is subject to much stronger lesson-to-lesson fluctuations. This suggests that the students did indeed distinguish accurately between their own abilities (self-concept) and the learning progress they achieved (achievement) in that particular lesson. Nevertheless, it would be useful to replicate the study by using alternative criteria to measure students' achievement. For example, teachers' ratings might be a viable alternative. However, student gains from lesson to lesson for each student might not be readily visible to teachers, and teacher ratings are subject to other judgment biases (Loibl et al., 2020). Hence, we leave this concern as a direction for further research.

Our study does not provide a direct replication of the classic between-person reciprocal effects model (Marsh, 1990; Marsh & Craven, 2006; Wu et al., 2021) at the within-person level. A large body of research has established the reciprocal effects model using between-person, long-term approaches and more objective achievement indicators such as grades or test scores. The present study diverges from this research tradition not only in the level of analysis (within-person and short-term vs. between-person and long-term), but also in the nature of the achievement indicator (perceived achievement vs. more objective achievement indicators). Further research is indicated here, linking the findings of the present study to the ongoing debate on the juxtaposition of within-person and between-person perspectives on the reciprocal effects model (see, e.g., Ehm et al., 2019).

Finally, in the current study, we focused on German secondary school students attending the academic track (i.e., the Gymnasium, which 44% of German students attend after elementary education; Autorengruppe Bildungsberichterstattung, 2018). Future research drawing on data from other educational systems, age groups, or ability tracks, and other cultural contexts is needed to test the results' generalizability.

Conclusion

ASC plays a prominent role in students' everyday life for their personal academic development. In the comparatively long tradition of research on ASC in educational psychology, there is clearly a lack of experience-sampling studies that have examined the psychosocial intraindividual processes postulated by ASC theory over time in an ecologically valid setting in real time. We are off to a good start with this study. Our results indicate that a student's ASC does indeed influence their achievement formation in everyday school life and that this, in turn, is influenced by previous achievement. Our results should thereby also be seen as a call and starting point for further research, not only to use within-person, experience-sampling approaches to replicate previous results in ASC research, but also to better understand situation-specific fluctuations in ASC and student motivation more broadly in concrete learning situations, which, as our results suggest, students actually experience. We strongly believe that this will significantly help to advance ASC theory and research toward providing a better understanding of students' learning experiences.

References

Amthauer, R. (1970). Intelligenz-Struktur-Test [Intelligence structure test]. Hogrefe.

Arens, A. K., Becker, M., & Möller, J. (2017). Social and dimensional comparisons in math and verbal test anxiety: Within-and cross-domain relations with achievement and the mediating role of academic self-concept. Contemporary Educational Psychology, 51, 240–252. https://doi.org/10.1016/j.cedpsych.2017.08.005

Arens, A. K., Marsh, H. W., Pekrun, R., Lichtenfeld, S., Murayama, K., & vom Hofe, R. (2017). Math self-concept, grades, and achievement test scores: Long-term reciprocal effects across five waves and three achievement tracks. Journal of Educational Psychology, 109(5), 621–634. https://doi.org/10.1037/edu0000163

Asparouhov, T., & Muthén, B. (2010). Bayesian analysis using Mplus: Technical implementation. Mplus Technical Report. http://statmodel.com/download/Bayes3.pdf

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation models. Structural Equation Modeling, 25(3), 359–388. https://doi.org/10.1080/10705511.2017.1406803

Autorengruppe Bildungsberichterstattung. (2018). Bildung in Deutschland 2018: Ein indikatorengestützter Bericht mit einer Analyse zu Wirkungen und Erträgen von Bildung [Education in Germany 2018: An indicator-based report with an analysis of the effects and returns of education]. https://www.bildungsbericht.de/de/bildungsberichte-seit 2006/bildungsbericht-2018/pdf-bildungsbericht-2018/bildungsbericht-2018.pdf

Bolger, N., & Laurenceau, J.-P. (2013). Intensive longitudinal methods: An introduction to diary and experience sampling research. Guilford Press. Brown, G. T. L., Andrade, H.

L., & Chen, F. (2015). Accuracy in student self-assessment: Directions and cautions for research. Assessment in Education: Principles, Policy & Practice, 22(4), 444–457. https://doi.org/10.1080/0969594X.2014.996523

Brunner, M., Keller, U., Dierendonck, C., Reichert, M., Ugen, S., Fischbach, A., & Martin, R. (2010). The structure of academic self-concepts revisited: The nested Marsh/Shavelson model. Journal of Educational Psychology, 102(4), 964–981. https://doi.org/10.1037/a0019644

Byrne, B. M. (2002). Validating the measurement and structure of self-concept: Snapshots of past, present, and future research. American Psychologist, 57(11), 897–909. https://doi.org/10.1037/0003-066X.57.11.897

Calsyn, R. J., & Kenny, D. A. (1977). Self-concept of ability and perceived evaluation of others: Cause or effect of academic achievement? Journal of Educational Psychology, 69(2), 136–145. https://doi.org/10.1037/0022-0663.69.2.136

Chesebro, J. L., & McCroskey, J. C. (2000). The relationship between students' reports of learning and their actual recall of lecture material: A validity test. Communication Education, 49(3), 297–301. https://doi.org/ 10.1080/03634520009379217

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Academic Press.

Dickhäuser, O., & Plenter, I. (2005). "Letztes Halbjahr stand ich zwei". Zur Akkuratheit selbst berichteter Noten [On the accuracy of self-reported school marks]. Zeitschrift Für Pädagogische Psychologie, 19(4), 219–224. https://doi.org/10.1024/1010-0652.19.4.219

Dörendahl, J., Scherer, R., Greiff, S., Martin, R., & Niepel, C. (2021). Dimensional comparisons in the formation of domain-specific achievement goals. Motivation Science, 7(3), 306–318. https://doi.org/10.1037/ mot0000203

Dormann, C., & Griffin, M. A. (2015). Optimal time lags in panel studies. Psychological Methods, 20(4), 489–505. https://doi.org/10.1037/met0000041

Eccles, J. S. (2009). Who am I and what am I going to do with my life? Personal and collective identities as motivators of action. Educational Psychologist, 44(2), 78–89. https://doi.org/10.1080/00461520902832368

Ehm, J. H., Hasselhorn, M., & Schmiedek, F. (2019). Analyzing the developmental relation of academic self-concept and achievement in elementary school children: Alternative models point to different results. Developmental Psychology, 55(11), 2336–2351. https://doi.org/10.1037/ dev0000796

Fleeson, W. (2001). Toward a structure-and process-integrated view of personality: Traits as density distribution of states. Journal of Personality and Social Psychology, 80(6), 1011–1027. https://doi.org/10.1037/ 0022-3514.80.6.1011

Frenzel, A. C., Pekrun, R., & Goetz, T. (2007). Girls and mathematics—A “hopeless” issue? A control-value approach to gender differences in emotions towards mathematics. European Journal of Psychology of Education, 22(4), 497–514. https://doi.org/10.1007/BF03173468

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. Psychological Methods, 19(1), 72–91. https://doi.org/10.1037/a0032138

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. Statistical Science, 7(4), 457–472. https://doi.org/10.1214/ss/1177011136

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. (2014). Bayesian data analysis (3ed.). Taylor & Francis.

Giannakos, M. N., Sharma, K., Papavlasopoulou, S., Pappas, I. O., & Kostakos, V. (2020). Fitbit for learning: Towards capturing the learning experience using wearable sensing. International Journal of Human Computer Studies, 136, 102384. https://doi.org/10.1016/j.ijhcs.2019.102384

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. Personality and Individual Differences, 102, 74–78. https://doi.org/10.1016/j.paid.2016.06.069

Goetz, T., Bieg, M., Lüdtke, O., Pekrun, R., & Hall, N. C. (2013). Do girls really experience more anxiety in mathematics? Psychological Science, 24(10), 2079–2087. https://doi.org/10.1177/0956797613486989

Gogol, K., Brunner, M., Goetz, T., Martin, R., Ugen, S., Keller, U., Fischbach, A., & Preckel, F. (2014). "My questionnaire is too long!" The assessments of motivational-affective constructs with three-item and single-item measures. Contemporary Educational Psychology, 39(3), 188–205. https://doi.org/10.1016/j.cedpsych.2014.04.002

Gorges, J., Neumann, P., Wild, E., Stranghöner, D., & Lütje-Klose, B. (2018). Reciprocal effects between self-concept of ability and performance: A longitudinal study of children with learning disabilities in inclusive versus exclusive elementary education. Learning and Individual Differences, 61, 11–20. https://doi.org/10.1016/j.lindif.2017.11.005

Graham, J. W., Cumsille, P. E., & Elek-Fisk, E. (2003). Methods for handling missing data. In A. S. Velicer & W. F. Velicer (Eds.), Research methods in psychology (pp. 87–114). Wiley.

Grosz, M. P., Rohrer, J. M., & Thoemmes, F. (2020). The taboo against explicit causal inference in nonexperimental psychology. Perspectives on Psychological Science, 15(5), 1243–1255. https://doi.org/10.1177/1745691620921521

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. https:// doi.org/10.1037/0022-0663.95.1.124

Guo, J., Marsh, H. W., Parker, P. D., Morin, A. J. S., & Dicke, T. (2017). Extending expectancy-value theory predictions of achievement and aspirations in science: Dimensional comparison processes and expectancy-by-value interactions. Learning and Instruction, 49, 81–91. https://doi.org/10.1016/j.learninstruc.2016.12.007

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(1), 10–15. https://doi.org/10.1177/ 0963721416666518

Hamaker, E. L., Asparouhov, T., Brose, A., Schmiedek, F., & Muthén, B. (2018). At the frontiers of modeling intensive longitudinal data: Dynamic structural equation models for the affective measurements from the COGITO study. Multivariate Behavioral Research, 53(6), 820–841. https://doi.org/10.1080/00273171.2018.1446819

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. Psychological Methods, 20(1), 102–116. https://doi.org/10.1037/a0038889

Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. Journal of School Psychology, 49(5), 505–528. https://doi.org/10.1016/j.jsp.2011.07.001

Jak, S., & Jorgensen, T. D. (2017). Relating measurement invariance, cross-level invariance, and multilevel reliability. Frontiers in Psychology, 8, 1640. https://doi.org/10.3389/fpsyg.2017.01640

James, W. (1890). The principles of psychology (Vol. I). Henry Holt and Co. https://doi.org/10.1037/10538-000

Janis, R. A., Burlingame, G. M., & Olsen, J. A. (2016). Evaluating factor structures of measures in group research: Looking between and within. Group Dynamics, 20(3), 165–180. https://doi.org/10.1037/gdn0000043

Jansen, M., Lüdtke, O., & Robitzsch, A. (2020). Disentangling different sources of stability and change in students' academic self-concepts: An integrative data analysis using the STARTS model. Journal of Educational Psychology, 112(8), 1614–1631. https://doi.org/10.1037/edu0000448

Kaplan, D. (2009). Structural equation modeling: Foundations and extensions. Sage.

Kim, E. S., Dedrick, R. F., Cao, C., & Ferron, J. M. (2016). Multilevel factor analysis: Reporting guidelines and a review of reporting practices. Multivariate Behavioral Research, 51(6), 881–898. https://doi.org/10.1080/00273171.2016.1228042

Kline, R. B. (2005). Principles and practice of structural equation modeling. Guilford Press.

Kurucay, M., & Inan, F. A. (2017). Examining the effects of learner-learner interactions on satisfaction and learning in an online undergraduate course. Computers & Education, 115, 20–37. https://doi.org/10.1016/j.compedu.2017.06.010

Liepmann, D., Beauducel, A., Brocke, B., & Amthauer, R. (2007). Intelligenz-Struktur-Test 2000R [Intelligence structure test 2000R]. Hogrefe. Liepmann, D., Beauducel, A., Brocke, B., & Nettelnstroth, W. (2012).

Intelligenz-Struktur-Test-Screening [Intelligence structure test-screening]. Hogrefe.

Loibl, K., Leuders, T., & Dörfler, T. (2020). A framework for explaining teachers' diagnostic judgements by cognitive modeling (DiaCoM). Teaching and Teacher Education, 91, 103059. https://doi.org/10.1016/j.tate.2020.103059

Malmberg, L. E., & Martin, A. J. (2019)., April Processes of students' effort exertion, competence beliefs and motivation: Cyclic and dynamic effects of learning experiences within school days and school subjects. Contemporary Educational Psychology, 58, 299–309. https://doi.org/10.1016/j.cedpsych.2019.03.013

Marsh, H. W. (1990). Causal ordering of academic self-concept and academic achievement: A multiwave, longitudinal panel analysis. Journal of Educational Psychology, 82(4), 646–656. https://doi.org/10.1037/0022-0663.82.4.646

Marsh, H. W. (2006). Self-concept theory, measurement and research into practice: The role of self-concept in educational psychology. British Psychological Society.

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. Perspectives on Psychological Science, 1(2), 133–163. https://doi.org/10.1111/j.1745-6916.2006.00010.x

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. The British Journal of Educational Psychology, 81(Pt. 1), 59–77. https://doi.org/10.1348/ 000709910X503501

Marsh, H. W., & O’Mara, A. (2008). Reciprocal effects between academic self-concept, self-esteem, achievement, and attainment over seven adolescent years: Unidimensional and multidimensional perspectives of self-concept. Personality and Social Psychology Bulletin, 34(4), 542–552. https://doi.org/10.1177/0146167207312313

Marsh, H. W., & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. Educational Psychologist, 20(3), 107–123. https://doi.org/10.1207/s15326985ep2003_1

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Abduljabbar, A. S., Abdelfattah, F., & Jansen, M. (2015). Dimensional comparison theory: Paradoxical relations between self-beliefs and achievements in multiple domains. Learning and Instruction, 35, 16–32. https://doi.org/ 10.1016/j.learninstruc.2014.08.005

Marsh, H. W., Pekrun, R., Murayama, K., Arens, A. K., Parker, P. D., Guo, J., & Dicke, T. (2018). An integrated model of academic self-concept development: Academic self-concept, grades, test scores, and tracking over 6 years. Developmental Psychology, 54(2), 263–280. https://doi.org/10.1037/dev0000393

Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Arens, A. K. (2019). The murky distinction between self-concept and self-efficacy: Beware of lurking jingle-jangle fallacies. Journal of Educational Psychology, 111(2), 331–353. https://doi.org/10.1037/edu0000281

Marsh, H. W., Relich, J. D., & Smith, I. D. (1983). Self-concept: The construct validity of interpretations based upon the SDQ. Journal of Personality and Social Psychology, 45(1), 173–187. https://doi.org/10.1037/ 0022-3514.45.1.173

Marsh, H. W., Seaton, M., Dicke, T., Parker, P. D., & Horwood, M. S. (2019). The centrality of academic self-concept to motivation and learning. In K. A. Renninger & S. E. Hidi

(Eds.), The Cambridge handbook of motivation and learning (pp. 36–62). Cambridge University Press. https://doi.org/10.1017/9781316823279.004

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. Child Development, 76(2), 397–416. https://doi.org/10.1111/j.14678624.2005.00853.x

McNeish, D., & Hamaker, E. L. (2020). A primer on two-level dynamic structural equation models for intensive longitudinal data in Mplus. Psychological Methods, 25(5), 610–635. https://doi.org/10.1037/met0000250

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. Measurement: Interdisciplinary Research and Perspectives, 2(4), 201–218. https://doi.org/10.1207/s15366359mea0204_1

Möller, J., Retelsdorf, J., Köller, O., & Marsh, H. W. (2011). The reciprocal internal/external frame of reference model: An integration of models of relations between academic achievement and self-concept. American Educational Research Journal, 48(6), 1315–1346. https://doi.org/10.3102/0002831211419649

Möller, J., Zitzmann, S., Helm, F., Machts, N., & Wolff, F. (2020). A meta-analysis of relations between achievement and self-concept. Review of Educational Research, 90(3), 376–419. https://doi.org/10.3102/0034654320919354

Movisens GmbH. (2017). movisensXS (Version 1.3.0-1.3.4) [Mobile App]. https://www.movisens.com

Murayama, K., Goetz, T., Malmberg, L.-E., Pekrun, R., Tanaka, A., & Martin, A. J. (2017). Within-person analysis in educational psychology: Importance and illustrations. British Journal of Educational Psychology Monograph Series II, 12, 71–87.

Muthén, L. K., & Muthén, B. O. (1998–2019). Mplus (Version 8.3) [Computer software]. https://www.statmodel.com

Niepel, C., Brunner, M., & Preckel, F. (2014a). Achievement goals, academic self-concept, and school grades in mathematics: Longitudinal reciprocal relations in above average ability secondary school students. Contemporary Educational Psychology, 39(4), 301–313. https://doi.org/ 10.1016/j.cedpsych.2014.07.002

Niepel, C., Brunner, M., & Preckel, F. (2014b). The longitudinal interplay of students' academic self-concepts and achievements within and across domains: Replicating and extending the reciprocal internal/external frame of reference model. Journal of Educational Psychology, 106(4), 1170–1191. https://doi.org/10.1037/a0036307

Niepel, C., Stadler, M., & Greiff, S. (2019). Seeing is believing: Gender diversity in STEM is related to mathematics self-concept. Journal of Educational Psychology, 111(6), 1119–1130. https://doi.org/10.1037/edu0000340

Orth, U., Dapp, L. C., Erol, R. Y., Krauss, S., & Luciano, E. C. (2021). Development of domain-specific self-evaluations: A meta-analysis of longitudinal studies. Journal of Personality and Social Psychology, 120(1), 145–172. https://doi.org/10.1037/pspp0000378

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. Educational Psychologist, 37(2), 91–105. https://doi.org/10.1207/S15326985EP3702_4

Pekrun, R., Murayama, K., Marsh, H. W., Goetz, T., & Frenzel, A. C. (2019). Happy fish in little ponds: Testing a reference group model of achievement and emotion. Journal of Personality and Social Psychology, 117(1), 166–185. https://doi.org/10.1037/pspp0000230

Peterson, S. E., & Miller, J. A. (2004). Quality of college students' experiences during cooperative learning. Social Psychology of Education, 7(2), 161–183. https://doi.org/10.1023/B:SPOE.0000018522.39515.19

Podsakoff, N. P., Spoelma, T. M., Chawla, N., & Gabriel, A. S. (2019). What predicts within-person variance in applied psychology constructs? An empirical examination. *Journal of Applied Psychology*, *104*(6), 727–754. https://doi.org/10.1037/apl0000374

Rauthmann, J. F. (2021). Capturing interactions, correlations, fits, and transactions: A person-environment relations model. In J. F. Rauthmann (Ed.), *The handbook of personality dynamics and processes* (1st ed., pp. 427–522). Elsevier. https://doi.org/10.1016/B978-0-12-813995-0.00018-2

Reis, H. T. (2014). Why researchers should think "real-world": A conceptual rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook for research methods for studying daily life* (pp. 3–21). Guilford Press.

Retelsdorf, J., Köller, O., & Möller, J. (2014). Reading achievement and reading self-concept—Testing the reciprocal effects model. *Learning and Instruction*, *29*, 21–30. https://doi.org/10.1016/j.learninstruc.2013.07.004

Richmond, V. P., McCroskey, J. C., Kearney, P., & Plax, T. G. (1987). Power in the classroom VII: Linking behavior alteration techniques to cognitive learning. *Communication Education*, *36*(1), 1–12. https://doi.org/10.1080/03634528709378636

Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research & Evaluation*, *11*, Article 10. https://doi.org/10.7275/9wph-vv65

Rovai, A. P., Wighting, M. J., Baker, J. D., & Grooms, L. D. (2009). Development of an instrument to measure perceived cognitive, affective, and psychomotor learning in traditional and virtual classroom higher education settings. *Internet and Higher Education*, *12*(1), 7–13. https:// doi.org/10.1016/j.iheduc.2008.10.002

Schmidt-Atzert, L., & Amelang, M. (2012). *Psychologische Diagnostik* [Psychological assessment]. Springer. https://doi.org/10.1007/978-3-642-17001-0

Schurtz, I. M., Pfost, M., Nagengast, B., & Artelt, C. (2014). Impact of social and dimensional comparisons on student's mathematical and English subject-interest at the beginning of secondary school. *Learning and Instruction*, *34*, 32–41. https://doi.org/10.1016/j.learninstruc.2014.08.001

Seaton, M., Marsh, H. W., Parker, P. D., Craven, R. G., & Yeung, A. S. (2015). The reciprocal effects model revisited: Extending its reach to gifted students attending academically selective schools. *Gifted Child Quarterly*, *59*(3), 143–156. https://doi.org/10.1177/0016986215583870

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, *46*(3), 407–441. https://doi.org/10.3102/00346543046003407

Shernof, D. J., Ruzek, E. A., Sannella, A. J., Schorr, R. Y., Sanchez-Wall, L., & Bressler, D. M. (2017). Student engagement as a general factor of classroom experience: Associations with student practices and educational outcomes in a university gateway course. *Frontiers in Psychology*, *8*, 994. https://doi.org/10.3389/fpsyg.2017.00994

Shernoff, D. J., Ruzek, E. A., & Sinha, S. (2017). The influence of the high school classroom environment on learning as mediated by student engagement. *School Psychology International*, *38*(2), 201–218. https:// doi.org/10.1177/0143034316666413

Shernoff, D. J., Sannella, A. J., Schorr, R. Y., Sanchez-Wall, L., Ruzek, E. A., Sinha, S., & Bressler, D. M. (2017). Separate worlds: Theinfluence of seating location on student engagement, classroom experience, and performance in the large university lecture hall. *Journal of Environmental Psychology*, *49*, 55–64. https://doi.org/10.1016/j.jenvp.2016.12.002

Shin, N. (2003). Transactional presence as a critical predictor of success in distance learning. *Distance Education*, *24*(1), 69–86. https://doi.org/10.1080/01587910303048

Sparfeldt, J. R., Buch, S. R., Rost, D. H., & Lehmann, G. (2008). Akkuratesse selbstberichteter Zensuren [The accuracy of self-reported grades in school]. *Psychologie in Erziehung Und Unterricht*, *55*(1), 68–75.

Stapleton, L. M., Yang, J. S., & Hancock, G. R. (2016). Construct meaning in multilevel settings. *Journal of Educational and Behavioral Statistics*, *41*(5), 481–520. https://doi.org/10.3102/1076998616646200

Trautwein, U., & Möller, J. (2016). Self-concept: Determinants and consequences of academic self-concept in school contexts. In A. A. Lipnevich, F. Preckel, & R. D. Roberts (Eds.), *Psychosocial skills and school systems in the 21st century: Theory, research, and practice* (pp. 187–214). Springer. https://doi.org/10.1007/978-3-319-28606-8_8

Trull, T. J., & Ebner-Priemer, U. (2014). The role of ambulatory assessment in psychological science. *Current Directions in Psychological Science*, *23*(6), 466–470. https://doi.org/10.1177/0963721414550706

Tsai, Y.-M., Kunter, M., Lüdtke, O., & Trautwein, U. (2008). Day-to-day variation in competence beliefs. How autonomy support predicts young adolescents' felt competence. In H. W. Marsh, R. G. Craven, & D. M. McInerney (Eds.), *Self-processes, learning, and enabling human potential* (pp. 119–143). Information Age Publishing.

Valentine, J. C., & DuBois, D. L. (2005). Effects of self-beliefs on academic achievement and vice-versa: Separating the chicken from the egg. In H. W. Marsh, R. G. Craven, & D. M. McInerney (Eds.), *International advances in self research* (Vol. 2, pp. 53–78). Information Age Publishing.

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist*, *39*(2), 111–133. https://doi.org/10.1207/s15326985ep3902_3

Wu, H., Guo, Y., Yang, Y., Zhao, L., & Guo, C. (2021). A meta-analysis of the longitudinal relationship between academic self-concept and academic achievement. *Educational Psychology Review*. Advance online publication. https://doi.org/10.1007/s10648-021-09600-1

Yoon, S., Kim, S., & Kang, M. (2020). Predictive power of grit, professor support for autonomy and learning engagement on perceived achievement within the context of a flipped classroom. *Active Learning in Higher Education*, *21*(3), 233–247. https://doi.org/10.1177/1469787418762463

Zirkel, S., Garcia, J. A., & Murphy, M. C. (2015). Experience-sampling research methods and their potential for education research. *Educational Researcher*, *44*(1), 7–16. https://doi.org/10.3102/0013189X14566879

Tables and Figures

Table 1

Testing Alternative Factor Structures and Cross-Level Invariance

Model	Within	Between	MLR V2 (*df*)	CFI	TLI	RMSEA	$SRMR_W$	$SRMR_B$	AIC	BIC
				Alternative factor structures						
(a)	Saturated	2 factors	12.499 (8)	1.000	.993	.014	.002	.003	31,590.651	32,428.699
(b)	Saturated	1 factor	158.181 (9)	.985	.786	.078	.044	.040	31,734.823	32,566.969
(c)	2 factors	Saturated	11.243 (8)	1.000	.995	.012	.009	.001	31,598.737	32,436.785
(d)	1 factor	Saturated	848.415 (9)	.918	.204	.186	.122	.004	33,438.489	34,270.635
(e)	1 factor	2 factors	1,135.797 (17)	.891	.150	.156	.122	.004	33,435.386	34,220.318
(f)	2 factors	1 factor	362.261 (17)	.966	.738	.087	.044	.040	31,745.567	32,530.499
(g)	2 factors	2 factors	26.695 (16)	.999	.991	.016	.009	.003	31,600.308	32,391.142
				Cross-level invariance						
(h)	2 factors	2 factors	41.082 (20)	.998	.986	.020	.015	.012	31,614.708	32,381.935

Note. MLR = Maximum likelihood estimation with robust standard errors; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; $SRMR_w$ = standardized root mean square residual value for within; $SRMR_b$ = standardized root mean square residual value for between; AIC = Akaike's information criterion; BIC = Bayesian information criterion.

Table 2

Level 2 Means, Variances (Unstandardized Estimates), and Correlations (Standardized Estimates) With Their 95% Credible Intervals (in Brackets)

Variable	Fixed effects (means)	Random effects (variances)	Correlations				
			1	2	3	4	5
1. MSC	3.247* [3.128, 3.368]	1.300* [1.117, 1.522]	—				
2. ACH	3.477* [3.390, 3.562]	0.558* [0.461, 0.675]	.860* [.814, .896]	—			
3. $ACH_{t-1} \rightarrow MSC_t$	0.125* [0.069, 0.181]	0.086* [0.057, 0.120]	−.130 [−.290, .027]	−.070 [−.263, .110]	—		
4. $MSC_{t-1} \rightarrow ACH_t$	0.091* [0.005, 0.169]	0.046* [0.019, 0.115]	−.030 [−.460, .398]	−.146 [−.581, .277]	−.560* [−.806, −.199]	—	
5. $MSC_{t-1} \rightarrow MSC_t$	0.171* [0.106, 0.242]	0.119* [0.083, 0.168]	−.142 [−.330, .047]	−.206 [−.405, .019]	−.529* [−.693, −.308]	.823* [.576, .954]	—
6. $ACH_{t-1} \rightarrow ACH_t$	0.320* [0.251, 0.386]	0.092* [0.062, 0.133]	−.413* [−.588, −.225]	−.453* [−.693, −.229]	.457* [.210, .631]	−.143 [−.541, .381]	−.057 [−.300, .232]

Note. MSC = mathematics self-concept (indicating students' trait score); ACH = perceived achievement (indicating students' trait score); ACH_{t-1} ! MSC_t = cross-lagged relation indicating students' skill-development effect; MSC_{t-1} ! ACH_t = cross-lagged relation indicating students' self-enhancement effect; MSC_{t-1} ! MSC_t = autoregressive effect for mathematics self-concept indicating students' carryover from lesson to lesson; ACH_{t-1} ! ACH_t = autoregressive effect for perceived achievement indicating students' carryover from lesson to lesson.

*Parameter's credible interval does not contain zero.

Table 3

Correlations Between the Factor Scores and Students' Gender, Reasoning Ability Scores, and Mathematics Grades

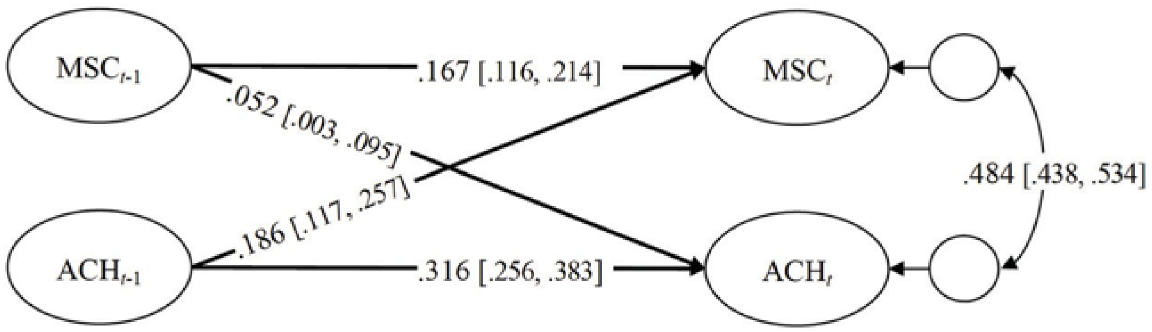
Variable	Gender	Reasoning ability	Mathematics grade
1. MSC	−.272**	.303**	.509**
2. ACH	−.217**	.244**	.414**
3. $ACH_{t-1} \rightarrow MSC_t$	.102	−.030	−.095
4. $MSC_{t-1} \rightarrow ACH_t$	−.034	−.043	−.013
5. $MSC_{t-1} \rightarrow MSC_t$	−.001	−.072	−.065
6. $ACH_{t-1} \rightarrow ACH_t$	.138	−.152	−.203*

Note. Gender: 0 = young men; 1 = young women; MSC = mathematics self-concept (within-person mean indicating students' trait score); ACH = perceived achievement (within-person mean indicating students' trait score); ACH_{t-1} ! MSC_t = cross-lagged relation indicating students' skill-development effect; MSC_{t-1} ! ACH_t = cross-lagged relation indicating students' self-enhancement effect; MSC_{t-1} ! MSC_t = autoregressive effect for mathematics self-concept indicating students' carryover from lesson to lesson; ACH_{t-1} ! ACH_t = autoregressive effect for perceived achievement indicating students' carryover from lesson to lesson.

* *p* , .05. ** *p* , .001.

Figure 1

Graphical Representation of and Results for the Within-Person Part of the Multilevel VAR(1) Model



Note. Paths indicating lesson-to-lesson self-enhancement effects (MSC_{t-1} to ACH_t), skill-development effects (ACH_{t-1} to MSC_t), as well as students' carryover from lesson to lesson (MSC_{t-1} to MSC_t; ACH_{t-1} to ACH_t). Standardized model parameters are shown; credible intervals are depicted in brackets. MSC = mathematics self-concept; ACH = perceived achievement.