Received: 28 July 2022 | Revised: 12 March 2023 | Accepted: 3 April 2023

DOI: 10.1002/pits.22929

RESEARCH ARTICLE

WILEY

The measurement of motivation and self-concept within the Students' approaches to learning framework

Ivan Ropovik[1,2] | **David Greger**[1]

[1]Institute for Research and Development of Education, Faculty of Education, Charles University, Prague, Czech Republic

[2]Faculty of Education, University of Presov, Prešov, Slovakia

Correspondence
Ivan Ropovik, Faculty of Education, Institute for Research and Development of Education, Charles University, Myslikova 7, 110 00 Prague, Czech Republic.
Email: ivan.ropovik@pedf.cuni.cz

Funding information
Grantová Agentura, Univerzita Karlova; NPO Systemic Risk Institute, Grant/Award Number: LX22NPO5101

Abstract

Motivation and self-concept count among the educationally most relevant factors and the evaluation of many educational interventions requires their valid measurement. The present study examined the psychometric properties of a shortened version of the Students' Approaches to Learning questionnaire measuring 10 distinct motivational and self-concept constructs. This large-scale study drew a nationally representative sample ($N = 6209$; 49% female) from a population of 11–12-year-old Czech students. The assessment of construct validity indicated (1) the structural relations between the constructs, (2) the predictive relations to scores in standardized achievement tests or teacher-assigned grades, and (3) the distinct differences in constructs' latent means between genders and students of the academic and mainstream track were mostly consistent with theory-derived predictions. The nomological network of the measure maps relatively well onto the observed relations. Although not intended for individual assessment, the measure allows for psychometrically sound group inferences in relatively diverse student populations.

KEYWORDS

academic track, motivation, self-concept, Students' Approaches to Learning, validity

1 | INTRODUCTION

Among the many theoretical constructs that likely contribute to educational outcomes, previous research frequently focused on self-concept and motivation. From a theoretical standpoint, self-concept and motivation are closely linked. Academic self-concept (SCACAD) represents self-related beliefs and expectancies about the ability to succeed in academic tasks and is considered a prerequisite of achievement motivation (Shavelson et al., 1976; Wigfield & Eccles, 2000). Broadly, it reflects both beliefs about general ability based on past performance and interactions with the environment and also expectations about the ability to perform academic tasks yet to come, called self-efficacy (SELFEF) (Bandura, 2001; Bong & Skaalvik, 2003; Shavelson et al., 1976). Apart from these two distinct temporal facets, SCACAD has a hierarchical structure, with global self-concept being driven by relatively unrelated domain-specific self-concepts like verbal (SCVERB) or math self-concept (SCMATH) (Marsh & Shavelson, 1985). These become gradually more differentiated and complex during development (Byrne & Shavelson, 1996; Marsh & Ayotte, 2003).

Previous research has shown that self-concept is associated with various educational outcomes, well-being, and future occupational prospects in general (Craven & Marsh, 2008; Marsh & Hau, 2003; Nagengast & Marsh, 2012). On the other hand, motivation, as a process of initiating and maintaining goal-directed behavior, has also been studied extensively from various perspectives within several theoretical frameworks. The most frequent theoretical standpoints refer to the distinction between intrinsic and extrinsic motivation (Deci & Ryan, 1985), how implicit theories orient individuals toward specific goals, triggering goal-specific behavioral patterns (Deci & Ryan, 1985), or how expectations and self-directed beliefs guide behavior (Wigfield & Eccles, 2000). In young learners, the typology of motivation based on external regulation (e.g., trying to fulfill the desires of parents) is still more frequent than intrinsic motivation, in which the regulation is based on the pleasure connected to the study activity (Alivernini et al., 2008). Past research has shown that motivation factors show moderate-to-strong interrelations with achievement as well as self-concept (Skaalvik & Rankin, 1990), more specifically, that academic motivation predicts later academic achievement (Guay et al., 2010). Since both motivation and self-concept are tightly associated with student achievement, both concepts have been studied in the context of gender and social achievement gaps. Using various contemporary theories of achievement motivation Meece et al. (2006) argue that girls' and boys' motivation beliefs and behaviors follow gender role stereotypes. Gender differences in motivation among adolescents have become a widely studied topic, yet, the application of different theories of motivation for scale construction represents a challenge that hampers cumulative knowledge.

In fact, girls generally report higher levels of academic motivation, on average (Bugler et al., 2015). Studies also show that girls tend to exhibit a more adaptive profile across a wide spectrum of academic motivation domains which is then moderately linked to higher academic achievement (King, 2016; Lam et al., 2012). Apart from academic motivation, girls also tend to report higher intrinsic motivation (Green & Foster, 1986), more autonomous motives for studying (Alivernini et al., 2018), and higher levels of emotional and behavioral engagement, while boys report higher levels of behavioral dissatisfaction (Skinner et al. 2008).

When broken down by gender, the intrinsic achievement motivation is domain-specific, with boys exhibiting higher motivation for learning math and girls being more motivated in the domain of language, even after accounting for prior achievement in these subjects (Skaalvik & Skaalvik, 2004). Recently, research into gender differences in domain-specific motivation started to differentiate between the effects of sex and gender identity corroborating findings of previous research (e.g., girls have higher levels of intrinsic motivation), but arguing, that differences in reading motivation are better explained by gender identity (feminine vs. masculine identity) then sex (McGeown et al. 2012; McGeown & Warhurst, 2020).

Similar gender stereotypic pattern can be seen in self-concept (Nosek & Smyth, 2011), however, the differences are asymmetric—the advantage favoring girls in the SCVERB is substantial while the advantage favoring boys in SCMATH is much smaller (Marsh et al., 2006). A meta-analysis of the relationship between self-concept and gender showed that significant gender-based differences in self-concept dimensions remain relatively stable across primary and secondary grades (Wilgenbusch & Merrell, 1999).

Motivation and self-concept are further modulated by socioeconomic status (SES) (Fan, 2011). A large-scale study on students in Italy examined the differences in the levels and profile of motivation between students with different SES and particularly among first- and second-generation immigrant students. Studies showed more favorable profiles toward intrinsic motivation in favor of students with high SES (Manganelli et al., 2021). With respect to immigrant status, first-generation immigrant students reported higher levels of intrinsic motivation, identified regulation, and external regulation than native students. Also, second-generation immigrants had higher academic motivation than natives, though the size of the differences was smaller (Alivernini et al., 2018). Apart from the various proximal effects of disadvantaging socioeconomic background, students stemming from these backgrounds also tend to be underrepresented in high-ability tracks or higher-tier schools (e.g., Ireson et al., 2002; Greger, 2012) and have less access to various outcome-relevant extracurricular activities. Age-related decline in self-concept also tends to be sharper in students from economically less well-off families (Kuscuoglu & Hartas, 2022).

There are numerous critical stages in which the child is more sensitive to changes in the affective domains. One such stage is the transition to lower secondary education, occurring by the end of the primary school age (around 11 years in many education systems). This is a critical stage, coinciding with early adolescence, where the students are at various risks of psychological disorders (see Evans et al., 2018). Evaluating the prevalence and age of onset for various mental disorders in the adult American population, Kessler et al. (2005) found that at age of 11, anxiety disorders and impulse-control disorders typically start to develop.

In a review of more than 100 studies, Symonds and Galton (2014) found that changes in school environment related to the transition from primary to lower-secondary education (e.g., from one classroom teacher to more subject-specific specialists) influence the psychological development of early adolescents, including their school engagement (their feelings about school and school subjects), and also found some contradictory trends in relation to the development of students' self-esteem and mental health. Other studies highlighted the effects of transition on student achievement. Transition to middle school was also found to be associated with dips in progress in attaining educational goals compared to typical progress in primary school (Galton et al., 1999), with most students' progress at this age starting to fall behind the expected progress. There is also longitudinal evidence (from third to eighth grade) on the decline in academic achievement evidencing that the momentary reduction of academic growth may be related to the changes occurring during the transition period (Akos et al., 2015). In general, higher educational demands tend to cause a drop in motivation (but see Alivernini et al., 2008) and self-concept (irrespective of gender) during the transition to secondary education (Arens et al., 2013; Coelho et al., 2017; Wigfield et al., 1991). While there are several well-validated measures of these largely affective factors intended for older age groups, it is, however, crucial to be able to assess motivation and self-concept factors reliably and validly already by the end of primary school age.

1.1 | Present study

The present psychometric study sought to evaluate the construct validity of a shortened version of the Students' Approaches to Learning questionnaire (SAL) in a nationally representative sample of Czech sixth graders. Although SAL is referred to as a theoretical framework, it is largely a carefully assembled, theoretically nonoverlapping set of measures (see Marsh et al., 2006), that aim to validly screen some of the most important affective constructs in the field of educational psychology. The 10 theoretical constructs mainly fall under the rubric of motivational preferences and self-related beliefs and cognitions. These include the following constructs: instrumental motivation (INSMOT), effort/perseverance (EFFPER), SELFEF, control expectations (CEXP), interest in reading (INTREA), interest in math (INTMAT), competitive learning (COMLRN), SCVERB, SCMATH, and SCACAD. For a detailed theoretical treatise of these constructs, see Marsh et al. (2006). To date, SAL is one of the most rigorously validated instruments measuring the given constructs, developed for the Organization for Economic Co-Operation and

Development's (OECD) Program for International Student Assessment (PISA) 2020 data collection. During its extensive pilot (carried out across 22 countries on a population of 15-year-old students), it has been narrowed down from an initial set of subscales measuring 29 constructs, preserving only scales that passed a stringent psychometric evaluation involving the study of the reliability, internal structure, item characteristics using item response theory methods, cross-cultural invariance, interrelations with external criteria, and usefulness for educational psychologists (Marsh et al., 2006).

Rather than examining the validity of SAL constructs in isolation, we aimed to test these measures simultaneously within a single structural model. Such an approach provides a much more stringent test with respect to concurrent and discriminant facets of construct validity. Although most of these constructs originate in distinct theoretical frameworks, it is an empirical question of whether the scales actually measure different constructs. Moreover, to provide evidence for the identity of these latent constructs, the tested nomological network (i.e., a system of hypotheses; see Cronbach & Meehl, 1955) has to show a specific pattern of relations between SAL constructs and academic achievement. Namely, (1) motivation and self-concept constructs should show stronger links within domains; (2) subject-specific self-concept and interest constructs should correlate much stronger within the subject domains compared to correlations between domains; (3) according to the internal/external frame of reference model (Marsh, 1986), subject-specific self-concept constructs should be dissociated despite strongly linked achievement in those subjects; (4) both motivational and self-concept measures should be positively associated with teacher-assigned grades and performance in achievement tests; (5) teacher-assigned grades should be more closely linked to self-concept in both subject domains than the actual academic achievement, as measured by standardized achievement tests; and (6) girls should exhibit higher interest and self-concept in language and reading and boys the same but in math.

Marsh et al. (2006) tested all these predictions in a large cross-cultural study utilizing the PISA 2000 data. While PISA samples 15-year-old students, the current psychometric study tested the construct validity and reliability of the SAL questionnaire in a markedly younger, nationally representative sample of 11- to 12-year-old Czech students. The reason for focusing on a rather narrow age range is that we wanted to assess the validity evidence in a population where (1) the measured constructs are expected to be already differentiated (Byrne & Shavelson, 1996; Marsh & Ayotte, 2003) and (2) where the measurement of those constructs arguably has peak utility. Namely, (1) it is a critical stage of elevated sensitivity to changes in the affective domains (Evans et al., 2018) and (2) students undergo an important transition from primary to lower-secondary education which is marked by a decline in motivational and self-concept factors (Arens et al., 2013; Coelho et al., 2017; Wigfield et al., 1991), with most students starting to lag behind the expected progress (Galton et al., 1999).

Although the sampling frame was the population of Czech students, Marsh et al. (2006) have shown that the inferred measurement properties of SAL are very similar across 25 OECD countries (including the Czech Republic). It may, therefore, be rational to assume that the results can be generalized beyond the population of Czech or Central European sixth graders. This assumption is, however, not testable against present data.

Extrapolating validity claims to the measurement of populations for which the measure was not originally intended requires either an inferential leap or data. That is because the relations between the items and the measured constructs may easily change over the levels of populations' characteristics. That even includes the possibility that some of the hypothesized constructs are the realization of different processes or are not yet differentiated at all. In this psychometric study, we aimed to provide novel empirical evidence for or against the use of SAL scales in the given younger age. Given the high potential utility of SAL for the comprehensive measurement of motivational and self-concept factors in educational practice, data are needed to back up the inferences drawn from such measurement.

Specifically, we set out to test whether the sample of younger students shows an identical factor structure and whether the same (above-described) pattern of theoretical predictions involving SAL factors and external criteria (performance in achievement tests and teacher-assigned grades) also holds for this age. We also tested whether the SAL scales show adequate reliability for group comparisons, whether they have the same measurement properties in both genders, across students attending different types of school (mainstream "basic schools" vs. 8-year gymnasia[1]), and across levels of SES. The inferences based on group comparisons are, however, only valid if the same attribute relates to the same set of observations in the same way in each group; that is, observed group differences in raw scores need to map well onto the group differences in theoretical attributes (Borsboom, 2006).

In the present study, we chose to examine whether the observed scores comply with specific invariance restrictions related to three, likely causally relevant, key educational population characteristics (see Sammons, 1995), gender, SES, and type of school (tracking by ability). Given that these three factors tend to be prognostic with respect to probably most learning outcomes and students' motivation and self-concept tend to vary significantly along these dimensions (e.g., Chmielewski et al., 2013; Meece et al., 2006), we regarded it important to test (and possibly invalidate) the notion of SAL's invariant functioning across these important strata of the target population. Only after establishing these measurement invariances, it is justified to interpret the measurement of the given constructs across different levels of these student characteristics without additional layers of (potentially violated) assumptions.

In contrast to Marsh et al.'s study (2006), we accounted for the dependencies in observations caused by the inherently hierarchical structure of the data, where students are nested within classes, and those are nested within schools. Together with additional screening procedures for careless responders or careful model fit inspection, we aimed to provide a stringent test of the validity implications drawn from the results of SAL when administered to younger students.

Last but not least, we tried to adhere to rigorous and accountable research practices. By openly sharing our code and SAL data, any colleague can reproduce our analyses and freely implement our code into their own analytic workflow.

2 | METHODS

2.1 | Participants and procedure

The total sample of this study[2] comprised 141 basic schools (sixth-graders; 4798 students) and 43 multiyear gymnasia ("first-graders"; 1745 students) sampled across the entire Czech Republic, totaling 6543 students (the sampling frame was represented by 3423 basic schools and 209 multiyear gymnasia). Reflecting the ethnic homogeneity of the underlying Czech population, a large majority of the students (95.3%) indicated coming from the Czech or Slovak ethnic majority (judged by the language in which their parents and grandparents talk to each other),[3] and only up to 3.5% students self-reported a non-Caucasian origin.

The selected schools were drawn from the original representative pool of 177 schools that took part in the nationwide TIMSS and PIRLS 2011 international assessments. A stratified two-stage cluster random sampling design was used. First, schools were randomly sampled from the population of all Czech schools attended by

[1]The Czech education system is characterized by an early tracking of students by ability. Even though many methods of nationwide and school-specific tracking exist in the system, 8-year gymnasia represent the most selective type of institution, accounting for approximately 11% of the students' population in grade 6. The selection for gymnasia is highly biased by SES, even after controlling for student achievement (Straková & Greger, 2018; Straková et al., 2017). Eight-year gymnasia are similar to academic school types in England known as secondary grammar schools and even more to academic gymnasia in the old German tripartite school system, which continue to dominate the academic track despite many other types of schools having evolved in Germany over time (Becker et al., 2017).

[2]The present study is part of the large-scale national longitudinal project Czech Longitudinal Study of Education, which tracks and studies the fourth graders who participated in the TIMSS and PIRLS 2011 international assessments.

[3]When estimated based on the language in which the students' parents talk to each other or the language that the students use to talk to their parents, respectively, the proportions were practically the same, 95.7% and 95.8%, respectively.

eligible students, using sampling probabilities proportional to the size of the school. Second, one or more intact classes of students were selected from each of the sampled schools. Above that, we supplemented the sample of basic schools with 8-year gymnasia using the same stratified two-stage cluster random sampling strategy. To account for oversampling (academic track students made up 28% of the total sample), we weighted the data using sampling weights (described in Section 2.3). The detailed sampling strategy for the present study is described in Greger et al. (2020); the sampling for the 2011 TIMSS and PIRLS international assessments is outlined in Jonas and Foy (2012). The data collection was carried out by a specialized research agency employing trained research administrators.

Absence in the regular and follow-up testing session or refusal to participate ($N = 333$) brought the final sample size of students participating in the study down to $N = 6209$. Another 13 participants were excluded for failing to answer at least two-thirds of the SAL items. Finally, 29 subjects were excluded for having an uninterrupted string of more than 20 consecutive identical responses for the 34 SAL items, likely indicating careless responding.[4] After these exclusions, the effective sample size for the analyses was $N = 6167$, with girls representing 49.02% of the sample. With respect to age, the mean of the present sample was 11.82 years ($SD = 0.41$). A total of 4572 students in the final sample attended mainstream schools, and 1624 students attended 8-year gymnasia.

We declare that the subjects were treated in accordance with established ethical standards, as stipulated in the Declaration of Helsinki.

2.2 | Measures

2.2.1 | SAL

The original SAL is a set of 52 self-report items that measure 14 of the most utilized theoretical constructs in educational psychology (Marsh et al., 2006). In this study, we used a shortened version of SAL composed of 34 self-report items that measure 10 factors related to motivation and self-concept, namely, INSMOT, EFFPER, SELFEF, CEXP, INTREA, INTMAT, COMLRN, SCVERB, SCMATH, and SCACAD. The original 14-factor SAL instrument was developed for OECD PISA 2000 based on literature reviews leading to the selection of 29 constructs, which were piloted among 15-year-old students in 22 countries (see Peschar et al., 1999). Our shortened 10-factor version of SAL did not include three constructs that measure learning strategies, namely, control strategies, memorization, and elaboration, plus one construct of cooperative learning. As a matter of fact, learning strategies have been shown to have an effect on learning outcomes (e.g., De Beni & Moè, 2003), either directly or via the mutual link between the choice of the learning strategy and motivation (Boekaerts, 1999). In the present study, however, we chose to focus on motivational and self-concept constructs, which, empirically, show markedly stronger relations to academic achievement (Marsh et al., 2006). Each of the 10 SAL scales was measured by 3–4 rating scale items with four response categories. As shown by Gogol et al. (2014), the measurement of motivational-affective constructs using scales with a small number of items may still have adequate psychometric properties. The measure was adapted (no age adjustments were carried out) in accordance with the PISA translation and adaptation standards (cApStAn & Béatrice Halleux, 2016). Table 1 shows examples of scale items (ones with the highest factor loading in the present study).

[4]While the conservative threshold of 20 consecutive identical responses resulted in the exclusion of 29 participants, with the threshold at ≥17, we excluded 93 participants, 212 participants exhibited ≥15 consecutive identical responses, and 547 participants had ≥10 such identical responses. Different exclusion rules had the most notable effect on latent correlations and model fit (with a more strict exclusion criterion logically lowering the χ^2 statistic). That said, the numeric differences did not have a substantive effect on the resulting inferences.

TABLE 1 Examples of scale items.

Scale	Item examples
Instrumental motivation	I study to get a good job.
Effort and perseverance	When studying, I try to do my best to acquire the knowledge and skills taught.
Perceived self-efficacy	I'm certain I can master the skills being taught.
Control expectation	If I want to learn something well, I can.
Interest in reading	I read in my spare time.
Interest in mathematics	Because doing mathematics is fun, I wouldn't want to give it up.
Competitive learning	Trying to be better than others makes me work well.
Self-concept in reading	I learn things quickly in the Czech language class.
Self-concept in mathematics	I have always done well in mathematics.
Academic self-concept	I'm good at most school subjects.

2.2.2 | Language aptitude test

A 26-item test with two variants was used to measure language aptitude. Multiple item formats were utilized. There were 13 multiple-choice items, 9 dichotomous items, and 4 open tasks. The tasks required the participant to identify and correct various linguistic errors (morphological, lexical, syntactic, and semantic) in authentic texts.

2.2.3 | Reading literacy test

The reading aptitude test comprised 19 tasks, half of them closed, the other half open-ended requiring a brief answer. The test utilized tasks from the PISA and PIRLS international assessments of reading literacy. Each of the two equivalent variants included three texts, each being associated with 3–12 tasks. The tasks focused on three domains of reading literacy: (1) information search, (2) information processing—in-text relations, and (3) text evaluation—linking with information not contained in the text.

2.2.4 | Math aptitude test

The test included 24 tasks, covering three content domains: (1) numbers and operations involving natural numbers, (2) dependencies, relations, and data processing, and (3) geometry of two- and three-dimensional shapes. Most of the tasks were motivated by the TIMSS 2007 items for fourth grade (directly adapted, newly created as paired equivalent variants, or completely new).

2.2.5 | SES

SES was modeled as a factor (principal component) score, computed using a formative factor model via principal component analysis. The indicators were the sum of the mother's and father's reported highest level of education achieved, the mother's and father's occupation, and the number of books at home.

All the materials are in Czech and are available upon request.

2.3 | Analysis

Latent models were estimated in the lavaan R package (Rosseel, 2012). SAL items were modeled as ordinal endogenous variables. To account for the dependent observations due to a two-stage clustered sampling (students nested within classes nested within schools) and unequal selection probabilities (requiring the use of sampling weights[5]), the models were refitted using the lavaan.survey R package (Oberski, 2014) and the robust maximum likelihood means- and variance-adjusted estimator, adjusting the standard errors of the estimated parameters (and also the model test indices) using a design-based approach. The given type of estimator (1) is robust with respect to the assumption of normal distribution of errors (especially kurtosis; not likely in Likert scales), (2) induces less bias in parameter estimation and model fit test of misspecified models, and (3) the proportion of type I errors in assessing correctly specified models with the given data is way more similar to the apriori defined nominal α value, as compared to, for example, the method of maximum likelihood (Beauducel & Herzberg, 2006).

Models were regarded as falsified based on a significant value of the χ^2 test statistics (see Ropovik, 2015). Following the rejection of the exact model-data fit hypothesis, we carried out a set of careful diagnostic procedures to identify the possible local sources of causal misfit and assess whether the model can be seen as a reasonable approximation to a model that would fit the observed data exactly. The fit was regarded as adequate if either (1) the exact fit test (χ^2 test) did not signal significant discrepancies between the data and the model or (2) if there was no larger pattern of substantial residuals (>0.1) indicating systematic local misfit. No model respecifications were carried out unless there were indications of severe local misfit, and the given respecification was also based on clear substantive grounds.

Apart from the model test, we provide a different perspective on the global fit using the following approximate fit indices: comparative fit index (CFI), Tucker–Lewis index (TLI), root mean square error of approximation (RMSEA), and standardized root mean squared error (SRMR). Given the model complexity and sample size, the statistical power for detecting a generally poor fitting model (RMSEA > 0.08) converged to 1. The Matrix of residuals and modification indices were examined to look for local sources of model misspecification.

For the final model, we computed bivariate latent correlations between the 10 factors. As these are intercorrelations between latent variables (reflecting the true scores), the relationships are corrected for (1) unreliability by partialling out the random errors, and for (2) scale-specific method factors. Together with the matrix of partial correlations reported in the supplementary analysis output, these matrices may help guide the conceptualization of alternative factor models.

To establish whether the factor structure and psychometric properties of the final factor model hold across several subpopulations, we carried out tests of measurement invariance, using a mean- and variance-adjusted weighted least squares estimator and modeling the indicators as ordered categorical, providing for rather conservative estimates. Measurement invariance was tested in a series of hierarchically nested models, gradually constraining (1) the configural structure of the latent-item relations, (2) factor loadings, (3) intercepts, and (4) error variances. Step 3, also called "strong invariance," allows for the comparison of factor means. That is because participants with the same factor level have identical expected values on the indicators. Any group differences for the factors thus reflect real underlying differences rather than the way the latents were measured. Step 4 is not required for comparisons of latent means but reflects the invariance in construct reliabilities (Beaujean, 2014). A nonsignificant Satorra–Bentler scaled χ^2 difference test was used as an indication of invariance between two respective, differently constrained models.

To assess the degree of comparative evidence for the structural coefficients, we also calculated approximate Bayes factors (BFs). BFs show whether there is comparative evidence either for *Ha* (effect present) or *H0* (effect absent), that is, whether the data are more consistent with *Ha*, *H0*, or inconclusive. The present study estimated BF based on a model-selection/information-criteria approach as proposed by Wagenmakers (2007), that is, by employing a Bayesian information criterion approximation that implicitly assumes a unit of information prior. Apart

[5]The applied sampling weights factor in to retain the population distribution with respect to gender proportion and type of school.

from BF, we also estimated the respective posterior probability of each of these parameters. Posterior probability refers to the probability of the parameter not being zero (as opposed to the probability of the data under a null). The estimation of posterior probability assumed a 1:1 prior odds of *H0* and *Ha* being true, respectively.

The analysis was preceded by data screening to filter out careless responders (by looking for uninterrupted long strings of identical responses) and mistyped values, as well as by the visualization of the variables (with a focus on kurtosis). None of the subjects were considered an outlier, and no data transformations were applied. There was a negligible amount of missing data (0.6%), which warranted treating missing data effectively as missing completely at random. To impute the missing data, we used the bootstrapped expected maximization procedure.

All the analyses reported in this paper are intended to be fully reproducible. Data for the analyzed SAL items, R code, and analytic outputs (Supporting Information: Material) are freely available at the Open Science Framework:https://osf.io/fkjdz/.

3 | RESULTS

Almost all the items of the SAL questionnaire were right-skewed, with lower item values denoting positive appreciation. Descriptive statistics and the matrix of polychoric correlations for individual scale items can be seen in Table 2 and Figure 1, respectively.

3.1 | The internal structure of SAL

3.1.1 | A priori model

First, we tested the a priori structure where each of the SAL items loaded on one of 10 intercorrelated factors. This model did not converge. Model diagnostics showed that the defined latent variables failed to explain away all the systematic covariance between the items "Mathematics is one of my best subjects" and "Because doing mathematics is fun, I wouldn't want to give it up." Superficially, the items' content validity is very similar, but they load on different factors (INTMAT, SCMATH). This single misspecification (an unmodeled residual of 0.88) caused the convergence problem, so we allowed that error covariance to take values different from zero. Doing so reflects ignorance with respect to the underlying source of this misspecification.

3.1.2 | Respecified model

The respecified CFA model converged normally but failed the model test; $\chi^2(80) = 615.15$, $p < .001$. The comparative fit indices showed a rather mediocre fit to the data; CFI = 0.94, TLI = 0.93, falling short of the conventional criteria proposed by Hu and Bentler (1999) (see Ropovik, 2015). The entire confidence interval (CI) width of RMSEA, on the other hand, fell within the region indicating a good approximate fit; RMSEA = 0.033, 95% CI (0.032, 0.034). SRMR of 0.038 did not indicate much global absolute misfit either. However, given the beyond-chance deviations of the data from the theorized structure, further detailed model diagnostics were needed.

3.1.3 | Local fit assessment

The inspection of residual matrix and modification indices pointed to the poor psychometric properties of one of the SCVERB items, "I'm hopeless in Czech language classes." This (the only reverse-coded) item exhibited high

TABLE 2 Descriptive statistics for scale items.

Scale item	*M*	SD	Skewness	Kurtosis
INSMOT 1	1.66	0.80	0.94	0.03
INSMOT 2	1.83	0.92	0.81	−0.39
INSMOT 3	1.51	0.74	1.33	0.98
EFFPER 1	2.05	0.82	0.29	−0.69
EFFPER 2	1.94	0.81	0.40	−0.66
EFFPER 3	1.79	0.78	0.60	−0.47
EFFPER 4	1.71	0.76	0.72	−0.34
SELFEF 1	2.46	0.73	−0.24	−0.34
SELFEF 2	2.40	0.78	−0.10	−0.48
SELFEF 3	2.10	0.81	0.22	−0.66
SELFEF 4	1.86	0.75	0.40	−0.67
CEXP 1	2.04	0.83	0.24	−0.84
CEXP 2	2.09	0.85	0.34	−0.62
CEXP 3	2.30	0.80	0.04	−0.56
CEXP 4	1.80	0.76	0.55	−0.45
INTREA 1	2.16	1.03	0.41	−1.02
INTREA 2	2.34	1.01	0.16	−1.09
INTREA 3	2.09	1.07	0.51	−1.05
INTMAT 1	2.21	0.92	0.30	−0.77
INTMAT 2	2.09	1.01	0.50	−0.87
INTMAT 3	1.83	0.85	0.83	−0.01
COMLRN 1	2.20	0.92	0.32	−0.75
COMLRN 2	2.01	0.84	0.53	−0.31
COMLRN 3	1.51	0.75	1.43	1.45
COMLRN 4	2.26	0.86	0.23	−0.62
SCVERB 1	2.23	0.80	0.30	−0.31
SCVERB 2	2.20	0.90	0.31	−0.70
SCVERB 3	2.18	0.82	0.33	−0.38
SCMATH 1	1.91	0.85	0.65	−0.25
SCMATH 2	2.24	1.08	0.30	−1.21
SCMATH 3	2.04	0.91	0.49	−0.63
SCACAD 1	1.96	0.71	0.41	0.08

TABLE 2 (Continued)

Scale item	*M*	SD	Skewness	Kurtosis
SCACAD 2	1.84	0.67	0.51	0.36
SCACAD 3	2.08	0.74	0.36	−0.05

Abbreviations: CEXP, control expectations; COMLRN, competitive learning; EFFPER, effort/perseverance; INSMOT, instrumental motivation; INTMAT, interest in math; INTREA, interest in reading; SCACAD, academic self-concept; SCMATH, math self-concept; SCVERB, verbal self-concept; SELFEF, self-efficacy.

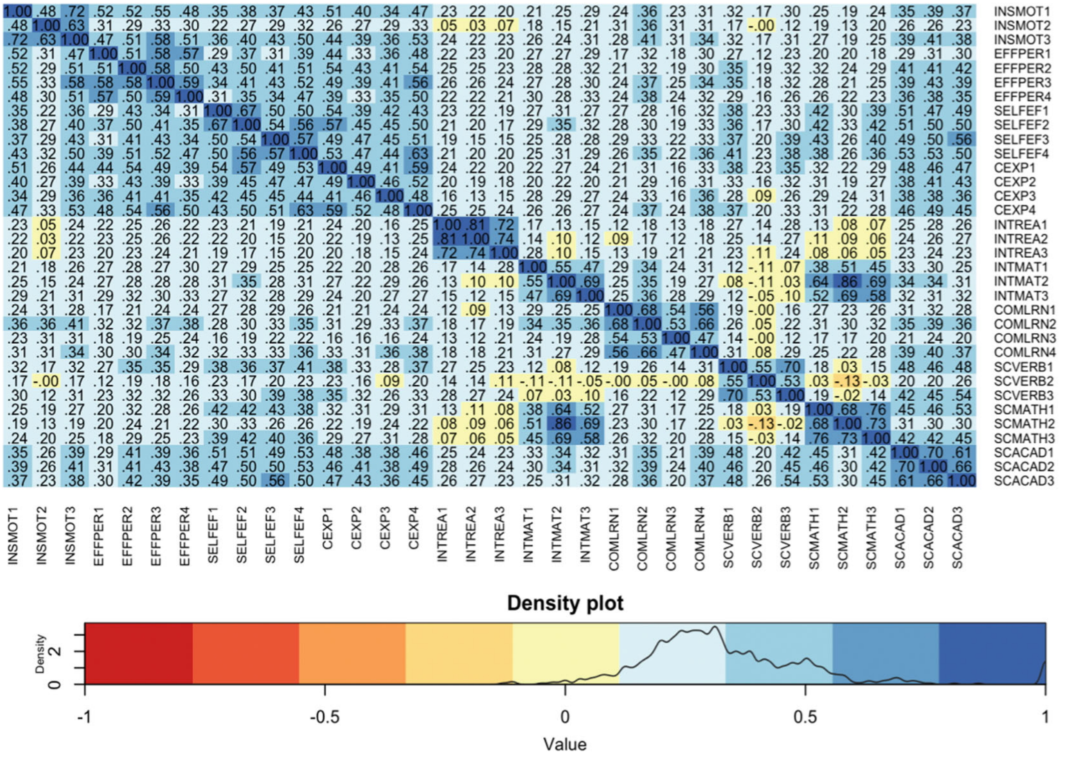

FIGURE 1 Polychoric correlation heatmap for individual scale items.

cross-loadings on other latent variables. Relatively strongest residuals were observed among the subject-specific factor items. Apart from these more pronounced residuals, there was a pattern of weaker but significant residuals among the motivation-related and self-concept-related factors (but not so much between).

Due to a large *N*, the estimation procedure was well-powered to pick up even quite small misspecifications. Overall, 35% (*N* = 198) of the residuals were significant at 0.05, while only 28 could be expected solely due to chance variation. Most (*N* = 10) of the high raw residuals (>0.1, thus allowing a product of two rather salient loadings ~0.3) were associated with the item “I'm hopeless in Czech language classes.” There were only three other residuals (0.05%) with an absolute magnitude slightly larger than 0.1.

TABLE 3 The fit of alternative models.

Model	χ^2	df	*p*	CFI	TLI	RMSEA [95% CI]	BIC
Primary 10-factors model	615	80/481	<.001	0.94	0.93	0.33 [0.32, 0.34]	434,159
Unitary factor	3291	82/526	<.001	0.61	0.61	0.80 [0.80, 0.81]	463,616
One second-order factor	1121	81/516	<.001	0.87	0.87	0.46 [0.45, 0.46]	439,718
Two second-order factors	1109	82/515	<.001	0.88	0.87	0.45 [0.44, 0.46]	439,418
Bifactor	1067	81/492	<.001	0.88	0.88	0.44 [0.44, 0.45]	438,767

Abbreviations: BIC, Bayesian information criterion; CFI, comparative fit index; CI, confidence interval; df, robust/standard degrees of freedom; *p*, *p* value; RMSEA, root mean square error of approximation; TLI, Tucker–Lewis index; χ^2, chi-square statistic.

3.1.4 | Alternative models

As the examination of alternative latent structures may lead to novel insights regarding how the given instrument works, we also fitted the following models: (1) a unitary factor model with all the SAL items loading onto a single dimension, reflecting a personal tendency for weighting learning and its outcomes positively; (2) a second-order model, where the 10 specific factors loaded on a unitary higher-order general factor; (3) a second-order factor model, where the 10 individual factors loaded onto two higher-order broad concepts—motivation and self-concept; (4) and a bifactor model, where each of the SAL items loaded on one of the 10 specific factors while also directly loading on a general factor, with covariances between the 10 specific factors and the general factor fixed to zero.

As can be seen in Table 3, all the alternative models (rows 2–5) markedly departed from the observed data. Because the models were all nested, special cases of the most complex bivariate model, we compared their fit to the primary model (row 1). In all instances, the χ^2 difference test significantly favored the primary model, with $p < .001$. Across the alternative models, the bifactor model fared the best, even when compared to the second best, the two second-order factors model, with $\Delta\chi^2(23) = 420$, $p < .001$.[6] This is pretty much an expected result while being tricky to interpret. The issue with the bifactor model is that its better fit (even after accounting for the difference in model parsimony) may be due to its superior ability to accommodate unmodeled complexity in the structure of the measure (e.g., nontrivial cross-loadings), which may well be the case here (see Muray & Johnson, 2013).

We also tested the most parsimonious, unitary-factor structure in more homogenous subgroups of students, that is, separately for basic schools and 8-year gymnasia, and for each of the quintiles of students' SES. For none of the subgroups did the model test yield a χ^2 value smaller than 3286 (df = 82)—a degree of fit that is very similar to the test of the full-sample single-factor model.

Overall, the results of testing alternative, more parsimonious factor structures show that the theory-based 10-factor model is a superior explanatory structure across the entire sample or in more homogenous subgroups.[7] The 10-factor model showed a relatively adequate, global approximate fit, with only 1 out of 34 items exhibiting severely unsatisfactory construct validity. Further details of model testing and diagnostics can be found in Supporting Information: Material.

[6]Please note that the χ^2 difference test is based on the standard test statistics, not the robust test that is reported per model. A robust difference test is then a function of two standard statistics (Rosseel, 2012).

[7]It has to be noted that the four alternative explanatory structures are not supported by the psychometric theory of the SAL measure (see Marsh et al., 2006) in which the individual factors conceptually tied to either academic motivation or self-concept are expected to be domain-specific, which would make the scientific interpretation and practical use of any unitary or higher-order factor inherently challenging.

TABLE 4 Latent correlations and reliability estimates.

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
[1] Instrumental motivation	0.72									
[2] Effort/persistence	0.77	0.80								
[3] self-efficacy	0.61	0.76	0.86							
[4] control expectation	0.68	0.86	0.93	0.78						
[5] Interest in reading	0.28	0.36	0.31	0.34	0.85					
[6] Interest in math	0.39	0.50	0.56	0.50	0.22	0.73				
[7] Competitive learning	0.50	0.51	0.50	0.53	0.25	0.50	0.85			
[8] Verbal self-concept	0.38	0.47	0.57	0.53	0.34	0.17	0.32	0.74		
[9] Math self-concept	0.31	0.39	0.57	0.46	0.12	0.91	0.41	0.16	0.82	
[10] Academic self-concept	0.52	0.63	0.83	0.75	0.36	0.56	0.52	0.66	0.58	0.76

Note: Estimates of scale reliabilities on the diagonal. All latent correlations were significant at $\alpha < .05$.

3.1.5 | Model interpretation

With some reservations, the relatively adequate model fit allowed us to interpret the model coefficients. The factor loadings are reported in Supporting Information: Material The mean of the factor loadings was at 0.72. Seventy-one percent of items had a factor loading higher than 0.7, and 82% of loadings were higher than 0.6. None of the items misbehaved in terms of their loading. In line with Gogol et al. (2014), who showed that motivational-affective constructs may have adequate psychometric properties even with a small number of items per scale, the SAL factors in our study also showed relatively good overall reliability[8] (in terms of internal consistency, see Table 4, diagonal values), ranging from McDonald's $\Omega = 0.72$ to $\Omega = 0.86$, with mean reliability of $\Omega = 0.79$ ($SD_{\Omega} = 0.05$).

Almost all the below-stated results are described not because they represent novel findings, but because they address the evidence on the construct validity of the 10 measured factors. Namely, the nomological network underlying the construct validation involves not only the assumptions about the existence of structural relations but also a set of assumptions about their strength (as linear relations or mean differences).

With respect to the relations between the latent factors, the data showed a high subject-specificity of self-concept and interest constructs. This pattern was especially apparent in math-related factors. For instance, INTMAT and SCMATH turned out to be practically collinear (0.91, with a shared variance of 83%), while the links between SCMATH and verbal constructs were very weak (<0.16, with a shared variance of 3%). Thus, the range of estimated correlation magnitudes was quite wide in the given population. There were also strong associations between domain-general self-concept and motivation constructs ranging from 0.61 to 0.93. Some of the latent constructs thus appear to be rather yet undifferentiated in the given age (see Table 4).

Lastly, SELFEF turned out to be the best predictor of subject-specific interest and self-concept constructs. That holds even after controlling for the other latent variables. The full matrix of correlations between the latent variables is shown in Table 4, the partial correlation matrix can be seen in Supporting Information: Material.

[8]Based on standards suggested by Nunnally (1978), the internal consistency estimates above 0.70 can be regarded as acceptable for scale development and values >0.80 for basic research. On one hand, such reliability values make comparisons of small differences in individual scores rather shaky, on the other hand, the items are not overly redundant, providing a measure of a construct that is not too narrow.

To examine the expected sampling error due to the hierarchical nature of the data, we also computed intraclass correlations (ICCs) for all measured variables. The ICCs (students nested within classes) ranged from r_{ICC} 0.023 to 0.098, with a median equal to $r_{ICC} = 0.048$. Relatedly, the mean value of the design effect (i.e., the increase in sampling error in the present clustered design compared with simple random sampling of individual students) was $d_{eff} = 1.128$, meaning a rather small effect of clustering on the sampling errors. The table of intra-class correlations for the model variables can be found in Supporting Information: Material.

3.2 | Measurement invariance

We tested the measurement invariance with respect to gender, SES (divided into quintiles), and type of school (mainstream schools vs. 8-year gymnasia). For levels of gender and SES, the nonsignificant Satorra–Bentler scaled χ^2 difference test indicated that the full measurement model was consistent with strong invariance but not strict invariance. That warrants the unbiased interpretability of structural relations and comparison of latent means. However, the absence of strict invariance indicates that the groups (boys vs. girls; SES quintiles) differ with respect to the reliability of the measured constructs. With respect to the type of school, the data only failed to rule out weak metric invariance, that is, equal factor loadings, which provides a firm ground for the interpretation of structural relations. This form of invariance refers to the absence of predictive bias due to differential item functioning. The lack of a strong metric invariance for the type of school, however, indicates that either there is a true difference in latent means or, if not, that it is reasonable to expect some bias in the comparison of latent means. On the other hand, the ΔRMSEA for the sequentially more restricted models did not exceed 0.003 for any of the three sets of invariance tests, which can be regarded as a relatively acceptable degree of absolute difference in fit (Meade et al., 2008). Detailed results of measurement invariance tests can be found in Supporting Information: Material.

3.3 | Comparisons of latent means

To compare the latent means across genders and type of school (which represents part of the nomological network for the studied constructs), we constrained loadings and latent intercepts to be equal. We set girls and mainstream schools, respectively, as the reference group, thus fixing their factor means to zero while estimating the factor means for the boys and 8-year gymnasia, respectively. These magnitudes equaled the difference between the groups. For presentation, the differences were converted to standardized latent mean difference units (Cohen's *d*) and a common language effect size, the probability of superiority, that is, the chance that a random boy/student from a mainstream school will be higher on the given variable than a random girl/student from an 8-year gymnasium.

Table 5 shows that all the differences in latent means apart from INSMOT and CEXP are significant. Negative values denote higher mean values for girls, and positive values have higher mean values for boys (factors are scaled inversely). Girls reported a markedly higher INTREA, and boys had a higher INTMAT. Girls also reported spending more effort and being more perseverant, while boys had higher values of SELFEF and COMLRN. Self-concept in language was found to be higher in girls, while the opposite was true for math. Students at 8-year gymnasia showed higher means compared to mainstream schools (factors are scaled inversely) for all factors with standardized effect sizes ranging from 0.07 to 0.21.

To set the effects in context by relating to some easily comparable and empirical benchmark, Hill et al. (2008) report the mean effect of educational interventions on achievement in upper elementary mainstream students to be at 0.22 ($\tau = 0.24$). That implies that the observed differences fell mostly into the small-to-medium part of the range usually reported in the educational literature. From an interpretational standpoint, we considered all the

TABLE 5 Comparison of latent means.

	Gender			Type of school		
	d (CLES)	*z*	*p*	*d* (CLES)	*z*	*p*
Instrumental motivation	0.01 (50%)	0.67	.51	0.12 (54%)	6.96	<.001
Effort/persistence	−0.04 (49%)	−2.53	.01	0.11 (53%)	6.43	<.001
self-efficacy	0.11 (53%)	6.30	<.001	0.21 (56%)	11.87	<.001
Control expectation	−0.01 (50%)	−0.28	.78	0.14 (54%)	7.91	<.001
Interest in reading	−0.26 (43%)	−14.71	<.001	0.21 (56%)	11.96	<.001
Interest in math	0.17 (55%)	9.63	<.001	0.15 (54%)	8.31	<.001
Competitive learning	0.09 (52%)	4.76	<.001	0.07 (52%)	3.78	<.001
Verbal self-concept	−0.12 (47%)	−6.69	<.001	0.20 (56%)	11.34	<.001
Math self-concept	0.23 (56%)	12.78	<.001	0.20 (56%)	11.22	<.001
Academic self-concept	0.05 (51%)	2.67	.01	0.19 (55%)	10.62	<.001

Note: For gender, positive mean differences reflect more positive construct values for boys. For the type of school, positive estimates reflect more positive values on a construct in 8-year gymnasium students. Cohen's *d* was estimated assuming equal variances of latent scores in both groups.

Abbreviation: CLES, common language effect size (probability of superiority).

comparisons to be independent, without the intention of generalizing inference. Should the reader wish to regard these as a family of tests, they should interpret the given *p* values against a stricter threshold of $\alpha = .005$ (Bonferroni correction; α/k with $\alpha = .05$ and $k = 10$ tests).

3.4 | Predictive validity

Lastly, we tested the predictive validity of each of the 10 SAL factors with respect to standardized achievement measures and teacher-assigned grades in math and language. To model these predictive relationships, we included in the primary model the math and language achievement test scores and let them freely correlate with each other and with the 10 SAL factors. The same was done in a separate model for teacher-assigned grades in math and language.

The fit of the model did not deteriorate markedly due to the inclusion of language and math achievement measures; $\chi^2(82) = 649$, $p < .001$; CFI = 0.93, TLI = 0.93, RMSEA = 0.033, 95% CI (0.033, 0.034), SRMR = 0.038. The achievement scores correlated rather strongly and showed a positive relation toward each of the SAL factors. There was a correlation of 0.60 between math and language achievement measures. Most of the relations were of small-to-medium magnitude, by Cohen's (1988) standards. SELFEF and SCACAD showed relatively strongest and most balanced links to math and language achievement. The Bayesian analysis indicated very strong (BF_{10} for the second weakest predictive SAL factor was equal to 277) comparative evidence in favor of the existence of all but one of the effects with posterior probabilities close to 1.

The pattern is almost the same for the teacher-assigned grades,[9] although the relations are generally stronger (mean correlations of 0.29 for math grades and 0.28 for language grades). Detailed results are reported in Supporting Information: Material.

[9]Teacher-assigned grades/assessments for Language (averaged over Czech and English language grades) and Math.

4 | DISCUSSION

The aim of the present psychometric study was to examine the validity of the shortened version of the SAL in a population of students transitioning to secondary education. The theoretical framework of this measure (Marsh et al., 2006) involves 10 of the most educationally relevant constructs falling under the rubric of motivational preferences and self-concept. The nomological network (see Cronbach & Meehl, 1955) of this measure involved specific structural links between its constructs, specific predictive relations to achievement and teacher-assigned grades, as well as distinct differences in the constructs' latent means.

The results generally suggest that the nomological network of this multidimensional measure maps relatively well onto the observed relations. Although the hypothesis of exact fit between the data and the latent measurement model of SAL had to be rejected, further examination of local fit did not show any pattern of severe causal misspecifications. However, there was mainly one item that showed markedly inadequate construct validity and may introduce a considerable amount of systematic, construct-irrelevant variance into any sum score of the SCVERB subscale. Moreover, one of the INTMAT items had a very large cross-loading on SCMATH.

Apart from that, all the items behaved in accordance with the SAL's modeled theoretical structure. Despite the small number of items measuring each of the constructs (3–4), the scales also showed relatively good reliability (in terms of internal consistency).

With regard to the associations between the substantive latent dimensions, the data indicated high subject-specificity of self-concept and interest constructs, especially in the domain of math, as was expected based on past research (Byrne & Shavelson, 1996; Marsh & Ayotte, 2003; Marsh & Shavelson, 1985). We also found strong links between domain-general self-concept and motivation constructs. Excessive collinearity between some of the constructs like SCMATH and INTMAT, or between SELFEF and CEXP, however, suggests that these constructs may still be undifferentiated in the given age. Given that previous research (Marsh et al., 2006) showed similarly strong interrelations between exactly these pairs of constructs even in substantially older age groups (15 years), this raises the question of whether regarding the indicators of these constructs as measures of substantively distinct constructs is defensible. Doing so might be an example of a jingle-jangle fallacy (Kelley, 1927). While SELFEF and CEXP are best regarded to be both self-concept constructs (Marsh et al., 2019), SCMATH and INTMAT reflect self-concept and motivational aspects. Although it has been argued that self-related beliefs about the ability to succeed in school may represent the precondition for achievement motivation (Shavelson et al., 1976; Wigfield & Eccles, 2000), available data indicate that the affinity to the specific subject domain of math outweighs any discriminant validity for measuring self-concept and motivation, respectively, that the indicators of these two constructs may possess. No such collinearity between self-concept and motivational constructs was observed in the verbal domain, in line with previous research (Marsh et al., 2006).

The measure was shown to have invariant measurement properties across levels of gender, SES, and type of school (students from mainstream schools vs. 8-year gymnasia). For the former two, strong measurement invariance was concluded based on a nonsignificant χ^2 test (e.g., cross-gender invariance of motivational constructs has also been previously established by Grouzet et al., 2006). As these two student characteristics are known to be strongly related to long-term outcomes (Sammons, 1995), it is important that the observed scores map in the same way onto the here examined theoretical attributes (see Borsboom, 2006). For the type of school, any inferences stemming from the comparison of latent means must, however, be made with caution, since the model test indicated that the data were consistent with beyond-chance differences in intercepts between the populations. Although interpreting structural relations seems warranted, the lack of strong metric invariance means that one has to expect some degree of bias (but probably not of substantial magnitude) when interpreting the comparison of group means as the difference between the latent means.

The data showed expected differences in latent means—a higher INTREA and SCVERB among girls, boys showing a higher INTMAT and SCMATH; girls being more persistent and boys more self-efficacious and competitive; students attending 8-year gymnasia showing more positive values for all the studied factors in the given sample. Likewise, we found small-to-medium cross-sectional correlations between the motivational and self-concept factors and achievement measures and generally stronger links with grades. Although this was also found in past studies (Marsh & Hau, 2003; Nagengast & Marsh, 2012), there seems to be a temporal effect of motivation on later achievement, but not self-concept on later achievement (Guay et al., 2010).

4.1 | Limitations

All the findings of the present study need to be interpreted in light of the following limitations. The character of the present sample implies some constraints on generality (see Simons et al., 2017) regarding the possible inferences drawn from the current study. This study was conducted in Czechia, which is ethnically a very homogenous country with a low proportion (in the low units) of non-Caucasians or immigrants from culturally different countries. It was, therefore, not possible to quantitatively address whether the modeled latent structure is invariant across various ethnic groups. Measurement properties in the examined age range in other cultures (primarily those more or less western, educated, industrialized, rich, and democratic) may be expected to be different (Rad et al., 2018). Furthermore, the present study sampled students attending mainstream schools and 8-year gymnasia to test measurement invariance with respect to different tracking of students by ability. Although low-achievers are an integral part of the population attending mainstream schools, the current study does not provide a formal comparison of students selected in the lower-tier academic track. Including this specific subsample or samples from other cultures would provide a test to the boundary conditions of the findings presented in this study. It may well be possible that the measurement properties and structural relations between the studied constructs would be markedly different.

This study was primarily psychometrically oriented, with the goal to assess the measurement properties of self-report measures of 10 educationally relevant motivational and self-concept factors. Interpretation of any substantive findings needs to consider the self-report nature of the measurement used to assess the target constructs and the biases and response patterns inherent in such measurement (Paulhus & Vazire, 2007). Lastly, the theoretical model required respecification. Any data-driven changes to the a priori model may, however, capitalize on chance and thus compromise the prospects of model cross-validation (MacCallum et al., 1992). Although we tried to keep that at a minimum by doing only one model respecification, there were still a bit too many beyond-chance model residuals (especially among the subject-specific factor items), suggesting the presence of probably more than one unmodeled causes/factor. Their absolute magnitude was not substantial, but still, it muddies the psychometric meaning of the modeled constructs.

The present paper aimed to provide various types of valid evidence, embedded within a larger theory-based nomological network of the SAL measure. This is in line with the realist (and thus causal) notion of validity first laid out by Kelley (1927) and further elaborated by Cronbach and Meehl (1955) and Borsboom (2006) as a narrower, more technical concept—a property of a test to measure what it is supposed to measure, where the aim is to provide evidence consistent with the claim that the variations in an ontologically grounded attribute causally produce variation in the measurement outcomes. There are, however, other important aspects and qualities of measurement (like social utility and consequences of testing) that are an inherent part of another prominent concept of validity by Messick (1989) but were not investigated in this study.

From a substantive standpoint, it also has to be noted that noncognitive factors like motivation and subsequently also the achievement are directionally modulated by several other factors, like the perceived instrumentality (depending on the proximity in utility and whether the motivation is internal or external) (Simons et al., 2004). The personal preconceptions about the purpose of learning, beliefs about own learning process and strategies, or self-concept are then further externally shaped by the specific institutional contexts, performance requirements, and social environments (rather than explained by cultural stereotypes) (Gan, 2009). These and other further layers of relevant moderating effects may then represent a viable target for future studies.

In conclusion, the present study shows that even for a younger population of 11- to 12-year-old students, SAL maintains a relatively solid construct validity (relative to its complexity). Assessing motivation and self-concept at the age marked by the transition to secondary education is important, since it is a critical and challenging period in the life of the student (see Evans et al., 2018), frequently associated with falling behind the expected progress in meeting educational goals (Galton et al., 1999) and drop in these affective factors (Coelho et al., 2017). Although the reliability of SAL scales does not warrant interpretation at the level of an individual student, it can be considered a psychometrically valid and comprehensive measure of 10 motivational and self-concept factors, which can be employed for research in relatively diverse student populations, allowing for sound inferences concerning the scores it yields.

ACKNOWLEDGMENTS

The authors are grateful to colleagues from the Institute for Research and Development of Education, Faculty of Education, Charles University, who cooperated on data collection and processing, particularly on scaling and scoring of subject-matter tests within the Czech Longitudinal Study in Education project. This work was supported by project PRIMUS/20/HUM/009 and by the NPO "Systemic Risk Institute" number LX22NPO5101, funded by European Union—Next Generation EU (Ministry of Education, Youth, and Sports, NPO: EXCELES).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Data, R code documenting the analytic workflow, and analytic outputs (the Supporting Information: Material) are freely available at the Open Science Framework: https://osf.io/fkjdz/.

ORCID

Ivan Ropovik http://orcid.org/0000-0001-5222-1233

REFERENCES

Akos, P., Rose, R. A., & Orthner, D. (2015). Sociodemographic moderators of middle school transition effects on academic achievement. *The Journal of Early Adolescence*, *35*(2), 170–198.

Alivernini, F., Lucidi, F., & Manganelli, S. (2008). Assessment of academic motivation: A mixed methods study. *International Journal of Multiple Research Approaches*, *2*(1), 71–82.

Alivernini, F., Manganelli, S., Cavicchiolo, E., Girelli, L., Biasi, V., & Lucidi, F. (2018). Immigrant background and gender differences in primary students' motivations toward studying. *The Journal of Educational Research*, *111*(5), 603–611.

Arens, A. K., Yeung, A. S., Craven, R. G., Watermann, R., & Hasselhorn, M. (2013). Does the timing of transition matter? Comparison of German students' self-perceptions before and after transition to secondary school. *International Journal of Educational Research*, *57*, 1–11.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, *52*(1), 1–26.

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(2), 186–203.

Beaujean, A. A. (2014). *Latent variable modeling using R: A step-by-step guide*. Routledge.

Becker, M., Neumann, M., & Dumont, H. (2017). Recent developments in school tracking practices in Germany: an overview and outlook on future trends. *Orbis Scholae*, *10*(3), 9–25.

De Beni, R., & Moè, A. (2003). Imagery and rehearsal as study strategies for written or orally presented passages. *Psychonomic Bulletin & Review*, *10*(4), 975–980.

Boekaerts, M. (1999). Self-regulated learning: Where we are today. *International Journal of Educational Research*, *31*, 445–457.

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: how different are they really? *Educational Psychology Review*, *15*(1), 1–40.

Borsboom, D. (2006). When does measurement invariance matter? *Medical Care*, *44*(11 Suppl 3), S176–S181.

Bugler, M., McGeown, S. P., & St Clair-Thompson, H. (2015). Gender differences in adolescents' academic motivation and classroom behaviour. *Educational Psychology*, *35*(5), 541–556.

Byrne, B. M., & Shavelson, R. J. (1996). On the structure of social self-concept for pre-, early, and late adolescents: A test of the Shavelson, Hubner, and Stanton (1976) model. *Journal of Personality and Social Psychology*, *70*(3), 599–613.

cApStAn & Béatrice Halleux. (2016). *PISA 2018 translation and adaptation guidelines*. OECD. https://www.oecd.org/pisa/pisaproducts/PISA-2018-TRANSLATION-AND-ADAPTATION-GUIDELINES.pdf

Chmielewski, A. K., Dumont, H., & Trautwein, U. (2013). Tracking effects depend on tracking type: An international comparison of students' mathematics self-concept. *American educational research journal*, *50*(5), 925–957.

Coelho, V. A., Marchante, M., & Jimerson, S. R. (2017). Promoting a positive middle school transition: A randomized-controlled treatment study examining self-concept and self-esteem. *Journal of Youth and Adolescence*, *46*(3), 558–569.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.

Craven, R. G., & Marsh, H. W. (2008). The centrality of the self-concept construct for psychological wellbeing and unlocking human potential: Implications for child and educational psychologists. *Educational and Child Psychology*, *25*, 104–118.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior* (1985th ed.). Kluwer Academic/Plenum.

Evans, D., Borriello, G. A., & Field, A. P. (2018). A review of the academic and psychological impact of the transition to secondary education. *Frontiers in Psychology*, *9*, 1482.

Fan, W. (2011). Social influences, school motivation and gender differences: An application of the expectancy-value theory. *Educational Psychology*, *31*(2), 157–175.

Galton, M., Gray, J., & Ruddock, J. (1999). *The Impact of School Transitions and Transfers on Pupil Progress and Attainment*. Department for Education and Employment.

Gan, Z. (2009). 'Asian learners' re-examined: An empirical study of language learning attitudes, strategies and motivation among mainland Chinese and Hong Kong students. *Journal of Multilingual and Multicultural Development*, *30*(1), 41–58.

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.

Green, L., & Foster, D. (1986). Classroom intrinsic motivation: Effects of scholastic level, teacher orientation, and gender. *The Journal of Educational Research*, *80*(1), 34–39.

Greger, D. (2012). When PISA does not matter? The case of the Czech Republic and Germany. *Human Affairs*, *22*(1), 31–42.

Greger, D., Straková, J., & Martinková, P. (2020). Extending the ILSA study design to a longitudinal design: TIMSS and PIRLS extension in the Czech Republic—CLoSE study. In T. Nilsen, A. Stancel-Piątak, & J.-E. Gustafsson (Eds.), *International handbook of comparative large-scale assessment in education. An extensive review of perspectives and findings*. Springer.

Grouzet, F. M. E., Otis, N., & Pelletier, L. G. (2006). Longitudinal cross-gender factorial invariance of the academic motivation scale. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(1), 73–98.

Guay, F., Ratelle, C. F., Roy, A., & Litalien, D. (2010). Academic self-concept, autonomous academic motivation, and academic achievement: Mediating and additive effects. *Learning and Individual Differences*, *20*(6), 644–653.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, *2*(3), 172–177.

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), 1–55.

Ireson, J., Clark, H., & Hallam, S. (2002). Constructing ability groups in the secondary school: Issues in practice. *School Leadership & Management*, *22*(2), 163–176.

Jonas, M., & Foy, P. (2012). Sample design in TIMSS and PIRLS. In M. O. Martin, & I. V. S. Mullis (Eds.), *Methods and procedures in TIMSS and PIRLS 2011*. TIMSS & PIRLS International Study Center, Boston College.

Kelley, T. L. (1927). *Interpretation of educational measurements*. World Book Co.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National comorbidity survey replication. *Archives of General Psychiatry*, *62*(6), 593–602.

King, R. B. (2016). Gender differences in motivation, engagement and achievement are related to students' perceptions of peer—but not of parent or teacher—attitudes toward school. *Learning and Individual Differences*, *52*, 60–71.

Kuscuoglu, A., & Hartas, D. (2022). Academic self- concept, self-esteem and school attitudes in pre and mid adolescents: Gender, SES and parenting. *Research Papers in Education*, 1–24.

Lam, S., Jimerson, S., Kikas, E., Cefai, C., Veiga, F. H., Veiga, F. H., Nelson, B., Hatzichristou, C., Polychroni, F., Basnett, J., Duck, R., Farrell, P., Liu, Y., Negovan, V., Shin, H., Stanculescu, E., Wong, B. P. H., Yang, H., Zollneritsch, J., & (2012). Do girls and boys perceive themselves as equally engaged in school? The results of an international study from 12 countries. *Journal of School Psychology*, *50*(1), 77–94.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, *111*(3), 490–504.

Manganelli, S., Cavicchiolo, E., Lucidi, F., Galli, F., Cozzolino, M., Chirico, A., & Alivernini, F. (2021). Differences and similarities in adolescents' academic motivation across socioeconomic and immigrant backgrounds. *Personality and Individual Differences*, *182*, 111077.

Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal*, *23*(1), 129–149.

Marsh, H. W., & Ayotte, V. (2003). Do multiple dimensions of self-concept become more differentiated with age? The differential distinctiveness hypothesis. *Journal of Educational Psychology*, *95*(4), 687–706.

Marsh, H. W., & Hau, K.-T. (2003). Big-fish-little-pond effect on academic self-concept. A cross-cultural (26-country) test of the negative effects of academically selective schools. *American Psychologist*, *58*(5), 364–376.

Marsh, H. W., Hau, K.-T., Artelt, C., Baumert, J., & Peschar, J. L. (2006). OECD's brief self-report measure of educational psychology's most useful affective constructs: Cross-cultural, psychometric comparisons across 25 countries. *International Journal of Testing*, *6*(4), 311–360.

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.

Marsh, H. W., & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist*, *20*(3), 107–123.

McGeown, S., Goodwin, H., Henderson, N., & Wright, P. (2012). Gender differences in Reading motivation: does sex or gender identity provide a better account?: Gender differences in reading motivation. *Journal of Research in Reading*, *35*(3), 328–336.

McGeown, S. P., & Warhurst, A. (2020). Sex differences in education: Exploring children's gender identity. *Educational Psychology*, *40*(1), 103–119.

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*, *93*(3), 568–592.

Meece, J. L., Glienke, B. B., & Burg, S. (2006). Gender and motivation. *Journal of School Psychology*, *44*(5), 351–373.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13–103). Macmillan Publishing Co, Inc, American Council on Education.

Nagengast, B., & Marsh, H. W. (2012). Big fish in little ponds aspire more: Mediation and cross-cultural generalizability of school-average ability effects on self-concept and career aspirations in science. *Journal of Educational Psychology*, *104*(4), 1033–1053.

Nosek, B. A., & Smyth, F. L. (2011). Implicit social cognitions predict sex differences in math engagement and achievement. *American Educational Research Journal*, *48*(5), 1125–1156.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.

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

Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology*. Guilford Press.

Peschar, J. L., Veenstra, R., & Molenaar, I. W. (1999). *Self-regulated learning as a cross-curricular competency. The construction of instruments in 22 countries for the PISA main study 2000*. Organisation for Economic Co-Operation and Development (OECD).

Rad, M. S., Martingano, A. J., & Ginges, J. (2018). Toward a psychology of *Homo sapiens*: Making psychological science more representative of the human population. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(45), 11401–11405.

Ropovik, I. (2015). A cautionary note on testing latent variable models. *Frontiers in Psychology*, *6*, 1715.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36.

Sammons, P. (1995). Gender, ethnic and socio-economic differences in attainment and progress: A longitudinal analysis of student achievement over 9 years. *British Educational Research Journal*, *21*(4), 465–485.

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

Simons, J., Dewitte, S., & Lens, W. (2004). The role of different types of instrumentality in motivation, study strategies, and performance: Know why you learn, so you'll know what you learn! *British Journal of Educational Psychology*, *74*(3), 343–360.

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, *12*(6), 1123–1128.

Skaalvik, E. M., & Rankin, R. J. (1990). Math, verbal, and general academic self-concept: The internal/external frame of reference model and gender differences in self-concept structure. *Journal of Educational Psychology*, *82*(3), 546–554.
Skaalvik, S., & Skaalvik, E. M. (2004). Gender differences in math and verbal self-concept, performance expectations, and motivation. *Sex Roles*, *50*(3/4), 241–252.
Skinner, E., Furrer, C., Marchand, G., & Kindermann, T. (2008). Engagement and disaffection in the classroom: Part of a larger motivational dynamic? *Journal of Educational Psychology*, *100*(4), 765–781.
Straková, J., & Greger, D. (2018). Faktory ovlivňující přechod žáků 5. ročníků na osmileté gymnázium. *Orbis Scholae*, *7*(3), 73–85.
Straková, J., Greger, D., & Soukup, P. (2017). Factors affecting the transition of fifth graders to the academic track in the Czech Republic. *International Studies in Sociology of Education*, *26*(3), 288–309.
Symonds, J. E., & Galton, M. (2014). Moving to the next school at age 10-14 years: An international review of psychological development at school transition. *Review of Education*, *2*(1), 1–27.
Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review*, *14*(5), 779–804.
Wigfield, A., Eccles, J. S., Mac Iver, D., Reuman, D. A., & Midgley, C. (1991). Transitions during early adolescence: Changes in children's domain-specific self-perceptions and general self-esteem across the transition to junior high school. *Developmental Psychology*, *27*(4), 552–565.
Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 68–81.
Wilgenbusch, T., & Merrell, K. W. (1999). Gender differences in Self-Concept among children and adolescents: A meta-analysis of multidimensional studies. *School Psychology Quarterly*, *14*(2), 101–120.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Ropovik, I., & Greger, D. (2023). The measurement of motivation and self-concept within the Students' approaches to learning framework. *Psychology in the Schools*, 1–21. https://doi.org/10.1002/pits.22929