ZfE

SCHWERPUNKT

Teachers' and students' perceptions of students' ability and importance value in math and reading: A latent difference score analysis of intra-individual cross-domain differences

Inga ten Hagen (iD) · **Daria K. Benden** (iD) · **Fani Lauermann** (iD) · **Jacquelynne S. Eccles** (iD)

Abstract Informed by Eccles and colleagues' expectancy-value theory and Möller and Marsh's dimensional comparison theory, we examined cross-domain intra-individual differences in elementary teachers' ($N=57$) and their students' ($N=469$) ratings of students' ability and subjective importance of math and reading. Latent difference score analyses revealed that students perceived greater intra-individual differences in their own math versus reading ability than did their teachers. Analogous results emerged for students' and teachers' ratings of students' valuing (i.e., perceived importance) of math versus reading, suggesting differing dimensional comparison processes for students' self-judgments vs. their teachers' judgments. Cross-domain differences in teachers' and students' perceptions were positively associated for ratings of students' ability but not for ratings of students' perceived importance. Moreover, intra-individual differences varied substantially across students, in both students' and teachers' ratings. Students' gender and prior achievement in math and reading contributed to this variation.

Keywords Dimensional comparisons · Expectancy-value theory · Intra-individual differences · Student motivation · Teacher judgment

Inga ten Hagen (✉) · Daria K. Benden · Prof. Dr. Fani Lauermann
Center for Research on Education and School Development (IFS), TU Dortmund University,
Vogelpothsweg 78, 44227 Dortmund, Germany
E-Mail: inga.tenhagen@tu-dortmund.de

Daria K. Benden
E-Mail: daria.benden@tu-dortmund.de

Prof. Dr. Fani Lauermann
E-Mail: fani.lauermann@tu-dortmund.de

Prof. Dr. Jacquelynne S. Eccles
Irvine, 2068 Education, University of California, Irvine, CA 92697, USA
E-Mail: jseccles@uci.edu

✎ Springer

Wahrgenommene Kompetenz und Wichtigkeit von Mathematik und Lesen aus Sicht der Lernenden und der Lehrkräfte: Latente Differenzmodelle fachübergreifender Heterogenität innerhalb von Lernenden

Zusammenfassung Auf Basis der Erwartungs-Wert-Theorie von Eccles et al. und der Theorie dimensionaler Vergleiche von Möller und Marsh untersucht dieser Beitrag intra-individuelle Differenzen der Lernenden zwischen Mathe und Lesen bezüglich ihrer fachspezifischen Fähigkeiten und subjektiven Wichtigkeit der Fächer, beurteilt von den Lernenden ($N = 469$) und ihren Grundschullehrkräften ($N = 57$). Latente Differenzmodelle ergaben, dass die Lernenden größere intra-individuelle Unterschiede zwischen Mathe und Lesen sowohl in ihrer Fähigkeit als auch der Wichtigkeit des Fachs wahrnahmen als ihre Lehrkräfte, was auf unterschiedliche dimensionale Vergleichsprozesse bei Selbst- und Lehrkrafturteilen hindeutet. Lernenden- und Lehrkraftbeurteilung der Differenzen zwischen Mathe und Lesen waren positiv assoziiert für die Beurteilung der Fähigkeiten, aber nicht für die Beurteilung der wahrgenommenen Wichtigkeit des Fachs. Darüber hinaus variierten die intra-individuellen Unterschiede zwischen den Lernenden, sowohl in den Beurteilungen der Lernenden als auch der Lehrkräfte. Das Geschlecht der Lernenden und ihre Vorleistungen in beiden Fächern erwiesen sich als Prädiktoren dieser inter-individuellen Unterschiede.

Schlüsselwörter Dimensionale Vergleiche · Erwartungs-Wert-Theorie · Intra-individuelle Unterschiede · Lernendenmotivation · Lehrkrafturteil

1 Introduction

Empirical evidence informed by Eccles and colleagues' expectancy-value theory (EVT) suggests that students' domain-specific motivational beliefs, such as their expected academic success and subjective task values in a given domain (e.g., math or reading), are relevant predictors of students' academic engagement, achievement, and choices (Eccles et al. 1983; Steinmayr et al. 2019). Importantly, not only the absolute levels of students' domain-specific expectancies and values but also the relative placements of these motivational beliefs within students' intra-individual hierarchies of perceived ability and values—e.g., whether students value math more than reading or vice versa—shape their academic choices (Eccles 2009; Eccles and Wigfield 2020). For instance, recent evidence shows that students who perceive math as more valuable compared to the verbal domain are more likely to aspire to a career in the math domain than students who value both domains to a similar extent (Gaspard et al. 2019; Oppermann et al. 2021). Both EVT and the dimensional comparison theory propose that such intra-individual differences in students' motivational beliefs across subjects arise from contrasting comparisons between academic domains (Eccles 2009; Möller and Marsh 2013; Wigfield et al. 2020). Students' comparisons of their ability in one domain (mathematical or verbal) with their ability in another (dissimilar) domain (verbal or mathematical) reduce students' perceived ability in

the worse off domain and increase their perceived ability in the better off domain (Möller and Marsh 2013).

Moreover, according to EVT, messages from key socializers, such as teachers' expressed beliefs about the importance of specific academic domains and the domain-specific performance feedback they provide (e.g., school grades), can influence students' intra-individual hierarchies of motivational beliefs across different domains (Eccles and Wigfield 2020). However, research has focused primarily on cross-domain dimensional comparisons in *students'* self-perceptions (Möller et al. 2020; Wigfield et al. 2020), and comparatively little is known about the degree to which *teachers* might perceive corresponding differences in students' domain-specific abilities and values, particularly in elementary school samples. When teachers teach the same student in more than one subject—which is typical for elementary school teachers—their judgment of the student's ability or values in one subject could also be affected by their perception of this student in another subject. It is therefore important to understand how teachers form domain-specific judgments about their students' abilities and values and the extent to which dimensional comparisons might affect these judgments. However, no study to date has simultaneously examined intra-individual differences in students' and corresponding teachers' perceptions of students' abilities and values across different subjects. Moreover, despite the relevance of these intra-individual differences for students' educational and career choices (e.g., Gaspard et al. 2019), we lack evidence on which factors might contribute to such intra-individual cross-domain discrepancies in students' ability and values, as rated by both students and teachers.

To fill these research gaps, we used a unique dataset from the Childhood and Beyond (CAB) study (Eccles et al. 1993), which includes ratings of elementary school students and their teachers who taught the same students in both math and reading. Using a latent difference score model approach (Geiser et al. 2010), we examined intra-individual differences in students' ratings of their math versus reading ability (i.e., domain-specific ability self-concepts) and valuing (i.e., perceived importance of the domain), corresponding differences in teachers' ratings, associations between students' self-reported and teacher-perceived differences, and potential predictors of these cross-domain differences rated by both students and their teachers (i.e., students' gender, parental education, subject-specific school grades, and cognitive ability).

1.1 Expectancy-value theory and intra-individual cross-domain hierarchies of motivational beliefs

According to EVT (Eccles et al. 1983; Wigfield and Eccles 2000), an individual's choice for and engagement in a given task or domain are most proximally determined by their expectancy for success and their subjective valuing of the task or domain. Expectancies, which are individuals' beliefs about the likelihood of success in a given task or domain, and their self-concepts of ability (SCA), which reflect individuals' perceived domain-specific ability, are empirically highly correlated, and expectancy-value researchers have therefore often combined these two competence beliefs into a single score (e.g., Eccles et al. 1993; but see Wigfield et al. 2020,

for a discussion of conceptual differences). Similar to prior research using the CAB study (e.g., Durik et al. 2006), we are unable to differentiate between these two types of beliefs and reference students' SCA as a component of their expectancy beliefs.

Students' subjective task values (STV) refer to students' attainment value, intrinsic value, utility value, and cost (Eccles and Wigfield 2020). Attainment (or importance) value refers to the subjective importance of engaging in a given task or domain; intrinsic value refers to the experienced enjoyment of doing the task; utility value refers to its perceived usefulness; and the perceived cost represents the negative consequences, such as stress, effort or lost opportunities to engage in other activities, that go along with engaging in the task or domain. In this study, we focus only on the importance value facet because only this facet was assessed from both students' and teachers' perspectives. Importance value is highly correlated with the remaining value components in the expectancy-value framework and contributes to the overall subjective valuing of a given task (e.g., Gaspard et al. 2019; Lauermann et al. 2015). Therefore, we review research on students' STV more broadly to inform our research questions.

Ample evidence suggests that students are most likely to engage in activities in which they have relatively high SCA and to which they attach relatively high STV compared to other valued activities (Eccles and Wigfield 2020). Importantly, as noted previously, not only students' expectancies of success and subjective valuing of a specific task or domain but also the relative placement of these beliefs in the individuals' hierarchy of various tasks and domains determine students' learning engagement and academic career choices (Eccles et al. 1983; Eccles and Wigfield 2020; Gaspard et al. 2019; Jansen et al. 2021; Lauermann et al. 2015). For instance, Lauermann et al. (2015) found a significant negative interaction effect between students' math and verbal SCA in the prediction of their career aspirations in math and science: Students with relatively high SCA in both math and English had a lower probability of pursuing a career in the math/science domain than students with high math SCA but comparatively lower English SCA. In addition, Jansen et al. (2021) found that among students with comparable levels of math value, those who valued German relatively more than other students were less likely to have study intentions in the STEM field (science, technology, engineering, and mathematics). Similarly, profile analyses show that students' in profiles with high motivational beliefs only in math but not the verbal domain are more likely to aspire to a career in the field of STEM (Oppermann et al. 2021), to choose a major in STEM (Gaspard et al. 2019), and to pursue an occupation in STEM (Wang et al. 2013).

Students' intra-individual hierarchies are shaped by internal dimensional comparisons across tasks and domains. There is consistent empirical evidence—primarily from research based on the internal/external frame of reference model (I/E model; Marsh 1986)—suggesting that dimensional comparisons across dissimilar domains (e.g., the math and verbal domains) result in contrast effects, i.e., negative predictive effects of achievement in one domain on students' SCA in a contrasting domain (Möller et al. 2020). Moreover, dimensional comparisons not only affect the formation of students' SCA across domains but also their STV (Gaspard et al. 2018; Jansen et al. 2021; Lauermann et al. 2015; Wigfield et al. 2020). In a sam-

ple of German high school students, Gaspard et al. (2018) investigated dimensional comparison processes for students' SCA and nine STV facets across five academic subjects. Negative correlations between math- and language-related motivational beliefs emerged for students' SCA, whereas the corresponding cross-domain correlations for students' STV varied across STV facets. For instance, the correlation between students' perceived importance of achievement in math versus English was positive, whereas the corresponding cross-domain correlation for students' intrinsic value was not significant. Moreover, stronger contrast effects (i.e., a negative effect of achievement in one domain on students' SCA or value in the contrasting domain) occurred for students' SCA compared to their importance of achievement, indicating greater intra-individual differences in students' SCA than in the perceived importance of the different domains. Overall, the more strongly correlated a given STV facet was with SCA, the more likely it was to observe negative cross-domain associations. Accordingly, dimensional comparisons may affect students' SCA and STV to a different degree so that separate analyses are necessary.

In addition to the strength of the associations between students' motivational beliefs across domains, it is important to consider the size of students' perceived intra-individual cross-domain differences in motivational beliefs (Eccles and Wigfield 2020). For instance, profile analyses of students' intra-individual motivational hierarchies across domains suggest that some students have quite differentiated motivational profiles (e.g., high-math-low-verbal-motivation), whereas others do not (e.g., high overall motivation; Gaspard et al. 2019; Oppermann et al. 2021). However, we know of no prior studies that have investigated the amount of intra-individual cross-domain differences in students' motivational beliefs as well as the variability of these differences between students who might belong to the same motivational profile. Whereas profile analyses describe groups of students with different combinations of motivational beliefs (e.g., high math and verbal STV versus high math and low verbal STV), latent difference scores estimate the discrepancy in students' math versus verbal motivational beliefs for each individual student. Therefore, in the present study, we use latent difference score models to gain a detailed understanding of intra-individual hierarchies in students' SCA and STV across domains.

1.2 Intra-individual cross-domain differences in teachers' perceptions of students' ability and values

Teachers' knowledge about students' domain-specific ability and motivation is essential for the implementation of individualized teaching activities that match students' needs (cf. Hardy et al. 2019; Huber and Seidel 2018). Therefore, a growing body of research has focused on the associations between teachers' and students' judgments of student ability and motivation, and their (presumed) implications for student learning (for an overview, see Urhahne and Wijnia 2021). This evidence suggests that teachers tend to underestimate the amount of intra-individual variability in different student characteristics (e.g., motivation and achievement); however, the vast majority of research has been limited to a single academic domain (cf. Huber and Seidel 2018; Urhahne and Wijnia 2021). So far, only a few studies have examined dimensional comparisons in teachers' perceptions of their students across

domains, either assessing teachers' judgments of students' ability (Dompnier et al. 2006) or teachers' inferences about students' SCA (i.e., teachers' beliefs about students' self-evaluated abilities in different domains; Helm et al. 2018; Marsh et al. 1984; Pohlmann et al. 2004). Dompnier et al. (2006) found a strong positive association between teachers' judgments of a given student's scholastic aptitude in French versus math. Furthermore, students' achievement in math positively predicted their teacher's ratings of scholastic aptitude in French, even when differences in student achievement in French were controlled (and vice versa for math). The strong positive cross-domain association of teachers' judgments of students' ability might be due to a so-called halo effect, i.e., teachers' global impressions about a student (e.g., a smart student) can bias the judgments of this student's characteristics in a manner that is consistent with this overall impression (Nisbett and Wilson 1977).

Similarly, studies on dimensional comparisons in teachers' inferences about students' SCA showed positive correlations between teachers' judgments of students' math and verbal SCA, whereas students' self-reported SCA in the two domains were not significantly (Helm et al. 2018; Marsh et al. 1984) or even negatively associated (Pohlmann et al. 2004). Some authors thus concluded that there might be a halo effect in teachers' judgments of students' SCA across subjects (Marsh et al. 1984). The differing results for dimensional comparisons in teachers' and students' judgments of students' SCA could also be explained by different mechanisms in interpreting the information depending on the evaluator's perspective. For self-judgments, students may try to include as much diverse information as possible to identify their strengths and weaknesses across domains, whereas teachers may be more likely to use all easily available information to form a global judgment about a student to distinguish that student from the group of students (Pohlmann et al. 2004). It is important to note that these few studies on dimensional comparisons across domains in teachers' judgments often had samples of teachers who taught only one of the investigated subjects (i.e., math: Helm et al. 2018; Pohlmann et al. 2004). This is problematic because teachers' ratings of a given student's domain-specific motivational beliefs or ability might differ as a function of whether or not they provide instruction to this student in the evaluated domain. Moreover, because these studies mostly focused on teachers' judgments of students' SCA, it remains unclear whether teachers' perceptions of other motivational beliefs, such as students' STV in different domains, might also be affected by dimensional comparisons.

EVT proposes that students' intra-individual hierarchies of expectancies and values across different tasks or domains are influenced by their teachers' beliefs and behaviors (e.g., through feedback on domain-specific talent, Eccles 2009; Eccles and Wigfield 2020). However, there is a lack of research simultaneously examining both intra-individual cross-domain differences in students' motivational beliefs and corresponding teacher ratings. Despite teachers' possible underestimation of students' intra-individual diversity, teachers' and students' perceptions of intra-individual differences in students' ability and values might be interrelated because teachers' perceptions are formed based on students' characteristics and then communicated to students through their teaching behaviors (Zhu et al. 2018). Since teachers' judgment accuracy is relatively high for students' achievement but comparatively low for students' learning motivation (for a review, see Urhahne and Wijnia 2021), the

association of cross-domain differences in teachers' and students' perceptions might also vary depending on the characteristic to be judged. In order to examine the associations of cross-domain differences in teachers' versus students' perceptions of students' ability and importance, the present study uses a latent difference model and thus captures within-person, between-domain discrepancies in both teachers' and students' judgments.

1.3 Predictors of intra-individual cross-domain differences in teachers' and students' ratings of students' ability and values

There is some evidence that student characteristics such as grade level, socioeconomic status (SES), and gender can shape cross-domain differences in students' self-perceptions of their ability and values (e.g., Gaspard et al. 2020, 2019; Oppermann et al. 2021). For instance, evidence shows that students in higher grade levels distinguish better between their SCA or STV in different subjects, supporting the assumption that both expectancy and value beliefs become more differentiated with age (e.g., Gaspard et al. 2020; Möller et al. 2020). Regarding students' SES, cross-domain profile analyses reported by Gaspard et al. (2019) revealed that secondary school students with high SES are overrepresented in profiles with higher expectancies and values in both math and English, and in English in particular. In addition, females were more likely to have profiles with relatively low (rather than high) expectancies and values in math. Similarly, profile analyses of elementary students' motivational beliefs in Finnish language, math, and science showed an underrepresentation of females in the profile characterized by high intrinsic value and self-concept only in math compared to the other profiles (Oppermann et al. 2021). This evidence suggests that male and female students differ in their prioritization of the math relative to the verbal domain. However, it remains unclear whether these student characteristics also predict the amount of intra-individual cross-domain differences in students' self-perceptions of their ability and values.

Furthermore, not only students' self-perceptions but also teachers' judgments of students' ability and motivation can be affected by student characteristics such as school grades, SES, and gender (e.g., Brandmiller et al. 2020; Jussim et al. 1996; Urhahne and Wijnia 2021). For instance, teachers' judgments of students' performance rely on students' prior achievements (Jussim et al. 1996). Furthermore, Brandmiller et al. (2020) reported predictive effects of gender and SES on teachers' perceptions of students' cognitive abilities and learning motivation (e.g., willingness to perform), after taking into account student achievement and self-reported motivation. The results indicated that teachers tend to overestimate the cognitive skills and motivation of female students and students with high SES. Moreover, studies generally show gender biases in teachers' expectations of students' subject-specific ability (for a review, see Wang et al. 2018), with higher ratings for male compared to female students' aptitude in math (e.g., Heyder et al. 2019) and female compared to male students ability in reading (e.g., Muntoni and Retelsdorf 2018). To our knowledge, however, no study has investigated the association of student characteristics and differences in teacher-perceived student ability and values across different subjects. Therefore, the present study examined the predictive effects of students'

grade level, gender, parental education, cognitive ability, and grades on teacher- and student-rated intra-individual cross-domain differences in students' ability and importance value.

1.4 The present study

The present study examined intra-individual cross-domain differences in students' (self-reported and teacher-rated) ability and importance value across the domains of math and reading. Using latent difference score analyses, we studied three main research questions (RQ) and corresponding hypotheses (H). First, we examined cross-domain differences between individual students' ability and importance in the math versus reading domains (math-reading difference scores). We further contrasted the degree to which students perceive their subject-specific SCA and importance as discrepant across math and reading with the corresponding teachers' perceptions of cross-domain differences in individual students' math versus reading ability and the importance teachers believe their students assign to math versus reading (RQ1). Based on the literature review, which suggests stronger dimensional comparisons for students' self-ratings than for teachers' ratings of their students, we expected greater cross-domain differences in students' compared to teachers' perceptions of students' ability and importance in math versus reading (H1). Second, we examined the strength and nature of the associations between teachers' and students' perceptions of intra-individual cross-domain differences in students' ability and importance value in math versus reading (RQ2). Due to prior evidence showing that teachers' judgment accuracy is higher for students' achievement than students' learning motivation within a given domain (e.g., Urhahne and Wijnia 2021), we expected a stronger association between teachers' and students' perceptions of cross-domain differences in students' ability than importance value (H2). Third, we examined whether students' intra-individual cross-domain differences in ability and importance value in math and reading in students' (RQ3a) and teachers' (RQ3b) ratings might vary as a function of individual student characteristics (i.e., students' grade level, gender, parental education, cognitive ability, and prior math/reading grades). We hypothesized that, if gender differences emerge, male students will be more likely to have cross-domain differences in favor of math whereas female students will be more likely to have cross-domain differences in favor of reading, and that such differences would be reflected in both students' (H3a) and teachers' (H3b) ratings of students' ability and importance. Due to the limited available research on predictors of intra-individual cross-domain differences, we did not pose specific hypotheses for the remaining student characteristics.

2 Method

2.1 Study design and sample

Data come from the Childhood and Beyond study[1] (CAB, Eccles et al. 1993), in which three cohorts of students were followed over their school careers. The CAB study is uniquely suited for our research questions because it was designed to test central assumptions of EVT (e.g., the interrelation of socializers' beliefs and students' expectancy-value beliefs). Participants came from 12 elementary schools in primarily white (91%), middle- and working-class school districts in Southeastern Michigan, and were from middle to middle-upper socioeconomic backgrounds. The participants were contacted through their schools, and families and teachers gave their consent before participation. Data were collected from teachers and students during school time in the spring (i.e., at the end of each academic year) so that teachers had time to get to know their students during the school year. Teachers and students completed questionnaires in three 20-minute sessions in each classroom, supervised by project staff.

Our analyses focused primarily on Wave 4 data because teachers' subject-specific assessments of their students' ability and perceived importance value were available and teachers were explicitly asked to indicate the subject they teach to these students in that school year (1990). Relevant control variables from previous waves were included as well to account for preexisting differences in students' characteristics (e.g., students' grades in the previous school year; at Wave 3). Our analyses included data only from teachers who had indicated that they taught math *and* reading to the same students. The sample thus consisted of 57 teachers (82% female). Teachers' average age was 43 years ($SD = 8.8$), and their average teaching experience was 16 years ($SD = 8.7$). Students were matched with their specific math and reading teacher in all analyses. On average, the teachers had rated nine students per class. Notably, almost all students (93%) had been taught by a different teacher the previous school year, so preexisting differences in students' motivation and achievement were not due to the participating teachers' instruction. The student sample consisted of 469 students (53% female). At the time of the data collection in Wave 4, students across the three CAB cohorts were in third ($n = 148$, $M_{age} = 9$), fourth ($n = 140$, $M_{age} = 10$), and sixth grade ($n = 181$, $M_{age} = 12$).

2.2 Instruments

2.2.1 Student-rated ability self-concept and importance value

In the CAB study, students' SCA in math and reading was assessed with four items each (math: $\alpha = .83$, reading: $\alpha = .85$; e.g. "How good at math [reading] are you?", ranging from $1 = $ *not very good* to $7 = $ *very good*; items are reported in the supple-

---

[1] For further information about the data and its availability, see http://garp.education.uci.edu/cab---for-researchers.html and https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/T1QCQQ

mentary material). Students were asked to rate their subject-specific importance with one item for each domain: "For me, being good in math [reading] is …?" ranging from 1 (*not at all important*) to 7 (*very important*). We included only this item for students' STV because it matched with the assessment of teacher-perceived student importance. We also excluded original items containing comparative statements about (generic) "other subjects" so that our analyses focus specifically on math and reading.

2.2.2 Teacher-rated student ability and importance value

Teachers' perceptions of children's ability in each domain were assessed on a 7-point scale with two items: "Compared to other children, how much innate ability or talent does this child have in math [reading]?" ranging from 1 (*very little*) to 7 (*a lot*) and "How well do you expect this child to do next year in math [reading]?" ranging from 1 (*very poorly*) to 7 (*exceptionally well*; see also prior publications using the CAB that included teachers' ratings of students' ability: Lauermann et al. 2017; Wigfield et al. 1997). The internal consistency for this scale was good (math: $\alpha = .86$; reading: $\alpha = .87$). Teachers' ratings of students' importance value were assessed using
a student- and subject-specific item: "Compared to other children, how important does this child think it is to do well in math [reading]?" on a 7-point scale ranging from 1 (*not at all important*) to 7 (*very important*). Because each teacher was asked to rate all students in the CAB study across different subjects, the assessment was limited to a single item to reduce survey fatigue. Single-item ratings have been shown to be as valid as multi-item measures for teachers' judgments of students' domain-specific motivation (Zhu and Urhahne 2014).

2.2.3 Student characteristics

Students' gender (0 = *female*, 1 = *male*), parental education, and grade level were included as potential predictors of the estimated intra-individual difference scores. Parental education was operationalized as the parent-reported educational level of the parent with the highest education (1 = *grade school*, 2 = *some high school*, 3 = *high school graduate*, 4 = *some college or technical school*, 5 = *associate's degree*, 6 = *college graduate*, 7 = *some graduate school*, 8 = *master's degree*, 9 = *PhD or advanced professional degree*; see Durik et al. 2006). The participants' grade level (Grades 3, 4, and 6), which corresponds to the three cohorts in the CAB study, was represented by two dummy variables using the oldest cohort (Grade 6) as a reference group.

Students' cognitive ability and subject-specific academic achievements were also included as potential predictors of the estimated within-student difference scores. Students' general cognitive ability was assessed with the Slosson Intelligence Test-Revised (Nicholson and Hibpshman 1990; Slosson et al. 1991) when the students first joined the CAB study. Students' prior subject-specific achievement in math and reading (i.e., in Wave 3) was obtained from school records. Students' grades were coded on a scale from 1 (*failing*) to 16 (*A+*).

2.3 Data analysis

Preliminary analyses focused on bivariate correlations, missing data patterns, and tests of measurement invariance. Latent difference score models were estimated in M*plus* 8.1. The latent difference scores (LDS) were modeled such that they assess the difference between math and reading (ΔMath-Reading) for each of the student- and teacher-rated constructs (for a similar analytical approach, see Geiser et al. 2010; Scalas et al. 2014). In total, four LDS captured differences in students' math-reading SCA, students' math-reading importance value, teachers' judgments of students' math-reading ability, and teachers' judgments of students' math-reading importance value (Fig. 1).

For RQ1, we examined the means and variances of the estimated LDS for student- and teacher-perceived math-reading discrepancies in students' ability and importance value. Analyses of plausible values and corresponding confidence intervals for all LDS (Asparouhov and Muthén 2010) allowed us to identify the proportion of students with significantly higher (student- or teacher-reported) ratings in math compared to reading and vice versa. For RQ2, we estimated correlations between corresponding LDS based on teachers' versus students' ratings. Finally, for RQ3, we added individual student and family background characteristics as predictors of the estimated student- and teacher-rated LDS. We tested increasingly complex models
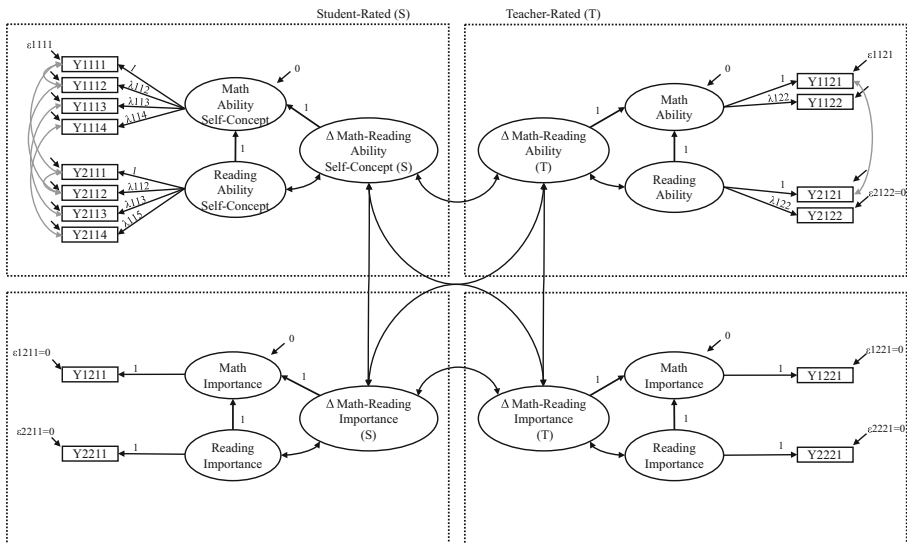

Fig. 1 Structure of the Latent Difference Score Measurement Model. Multi-perspective latent difference model for the two constructs (ability and importance) measured by two perspectives (students' self-rating and teachers' rating of the students' ability and importance) in the two different domains of math and reading. Each dotted square highlights a construct-perspective unit. Latent difference variables are included for each construct-perspective unit. *Ydcpi* denotes the *i*th observed item measuring construct *c* by perspective *p* in domain *d*. λ*cpi* denotes the domain invariant factor loading. The variances of the latent variables are estimated in the model. Intercepts (α*dcp*), latent means, and correlations between the latent variables (all reading and difference scores) are also included in the model but are not shown in the figure for simplification. The scales of the latent variables are identified by fixing the loading of the first indicator to one

that included subsets of predictors (see supplementary materials). Since there were no substantial differences in the estimated coefficients, we present only our final model including all covariates (Fig. 2). We accounted for the hierarchical structure of the data (students nested in classes) by using a robust estimation procedure in M*plus* (type = complex, estimator = MLR; see supplemental material for the Mplus syntax).

Model fit was evaluated based on the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). Good model fit is indicated by a CFI value of .95 or higher and RMSEA and SRMR values of .06 or less, whereas acceptable fit is indicated by a CFI value of about .90 or higher and RMSEA and SRMR values of .08 or less (Marsh et al. 2005). Increasingly stringent models tested for measurement invariance across subject domains, reflecting configural, metric, and scalar invariance (Van de Schoot et al. 2012). For model comparisons, a CFI difference between two models of less than .01 and an RMSEA difference of less than .015 generally indicates a negligible change in overall fit and supports the more parsimonious model (Chen 2007).

Missing data was below 3% for all student-reported or teacher-reported variables, but it was higher for the parent-reported level of education (26%) and school records data (34% in math grade, 32% in reading grade). Missing data were handled with the full information maximum likelihood algorithm (FIML).

3 Results

3.1 Preliminary analyses

Students' and teachers' perceptions of importance were modeled as latent variables with single-item indicators and no error term, whereas multi-item scales were modeled as latent variables with item-specific error terms (Fig. 1). Residual variances of parallel items across subjects were allowed to correlate due to parallel wording (Marsh and Hau 1996). Measurement invariance tests for parallel constructs across the two subjects of math and reading supported partial scalar invariance. The intercept for one item measuring students' perceived ability differed significantly across subjects. The cross-subject equality constraints for all other parameters (intercepts and factor loadings) led to a negligible change in model fit and were thus retained in subsequent models (ΔCFI = −.001, ΔRMSEA = −.001; see supplemental material for further description of the invariance testing). Measurement invariance was also supported across students' gender, as parameter constraints testing scalar invariance led to a negligible change in overall model fit (ΔCFI = −.004, ΔRMSEA = .000).

Descriptive statistics and bivariate correlations are reported in the supplemental materials for all variables. Students' math and reading importance were positively associated for both students' ($r = .40$, $p < .001$) and teachers' ratings ($r = .85$, $p < .001$). The correlation between students' math and reading SCA was not significant ($r = .07$, $p = .160$), whereas teachers' ratings of students' ability in math and
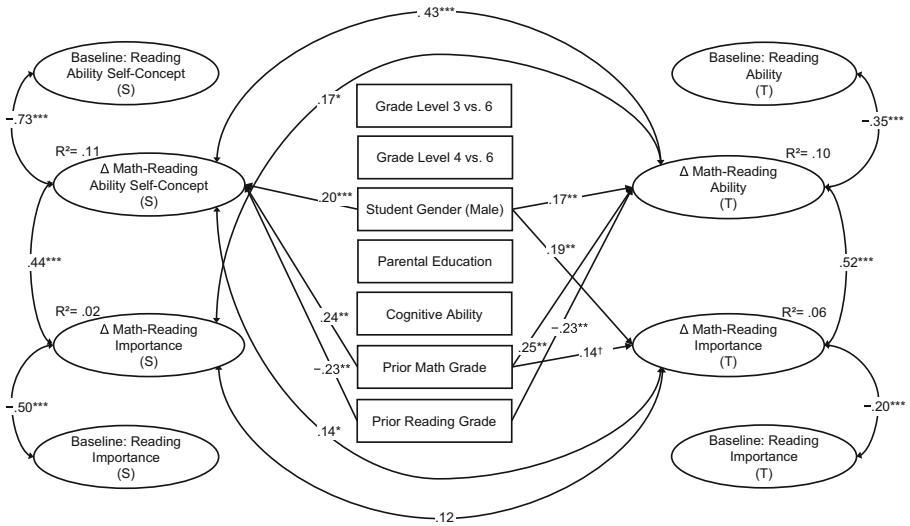
Fig. 2 Latent Difference Score Model Including all Hypothesized Predictors. All possible bivariate associations between the baseline (reading) and difference score variables were estimated but are not shown for the sake of readability and clarity. Only significant predictive effects and the corresponding standardized coefficients on the difference scores are shown in the Figure (see supplementals for non-significant effects and effects on reading scores). *S* student-rated, *T* teacher-rated

reading ($r = .73$, $p < .001$) as well as students' prior achievements in math and reading ($r = .55$, $p < .001$) were strongly positively correlated.

3.2 Intra-individual differences in students' math and reading ability self-concept and importance, and corresponding teacher judgments (RQ1)

A model estimating four LDS (Fig. 1)—i.e., for student-rated math-reading differences in SCA, student-rated math-reading differences in importance value, teacher-rated math-reading differences in student ability, and teacher-rated math-reading differences in students' importance value—fitted the data well (Satorra-Bentler $\chi^2 = 150.376$, $df = 81$, CFI = .977, RMSEA = .043, SRMR = .048). Latent means, variances, and plausible values for the four estimated LDS are shown in Table 1. Reading was defined as a baseline score so that positive LDS indicate that math is rated higher than reading, and negative LDS indicated that reading is rated higher than math. The means of these LDS did not significantly differ from zero. That is, for each construct, positive (math > reading) and negative (reading > math) LDS occurred with equal frequency (see Fig. 3).

However, the variances of the four LDS were significantly different from zero, indicating substantial variability in the size of the estimated (self-reported or teacher-rated) math-reading differences in student ability and importance value. That is, the estimated difference scores were small for some students but quite large for others. Furthermore, supporting H1, the variances of student-rated LDS were significantly
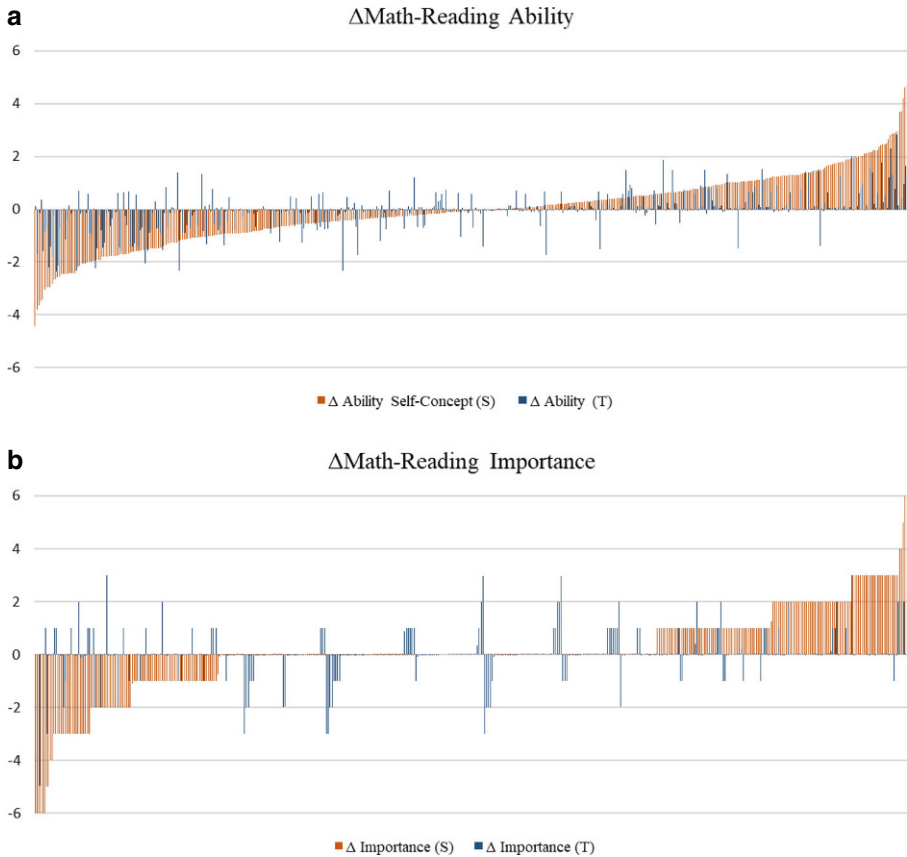
a

ΔMath-Reading Ability

■ Δ Ability Self-Concept (S)   ■ Δ Ability (T)

b

ΔMath-Reading Importance

■ Δ Importance (S)   ■ Δ Importance (T)

Fig. 3 Mean Plausible Values for all Latent Difference Scores per Student. Plausible values were generated for each latent difference score using Bayesian estimation. Plausible values for each student are shown, sorted ascending by math-reading differences in students' ratings. *S* student-rated, *T* teacher-rated

larger than the corresponding variances of teacher-rated LDS, both for students' ability (Wald test: 24.70, $p < .001$) and importance value (Wald test: 24.51, $p < .001$).

The estimated distribution of the plausible values showed that the average amount of difference between math and reading was higher in students' (approx. $M = 1.00$) than teachers' ratings (approx. $M = 0.40$; Table 1). This difference between students' and teachers' ratings regarding the absolute amount of cross-domain difference was significant for both student ability ($t(468) = 15.09$, $p < .001$) and importance ($t(468) = 9.77$, $p < .001$). Notably, the estimated number of students with significant math-reading differences in their ability was equal from both students' and teachers' perspectives (35%, see supplemental materials for frequencies of the plausible values), whereas the number of students with significant math-reading differences in their importance value was higher for students' self-reports than for teachers' ratings (50% vs. 22%). In other words, students were comparatively more likely to perceive one domain as more important than the other, whereas teachers were more likely to believe that students assign similar levels of importance to both domains.

Table 1 Latent Means and Variances for the Estimated Math-Reading Latent Difference Scores and Proportion of Students With Significant Teacher- and Student-Rated Between-Subject Difference Scores

Variable	Reading (baseline)		ΔMath-Reading		Negative Δ: Read > Math			Positive Δ: Math > Read		
	M	σ^2	M	σ^2	M	SD	% sign. (male/ female)	M	SD	% sign. (male/ female)
Ability Self-Concept (S)	5.60	1.11***	–0.05	2.04***	–0.98	0.84	18 (11/24)	1.00	0.86	17 (24/12)
Importance (S)	5.82	1.98***	0.09	2.57***	–1.00	1.41	21 (19/22)	1.01	1.18	29 (33/25)
Ability (T)	5.03	1.18***	–0.04	0.53***	–0.43	0.57	18 (12/23)	0.40	0.51	17 (20/14)
Importance (T)	5.39	1.90***	0.01	0.62***	–0.41	0.81	10 (5/13)	0.36	0.67	12 (19/6)

Plausible values were computed for each latent difference score using Bayesian estimation to examine latent means, standard deviations, and the proportion of students with significant difference scores separately for negative and positive difference scores

S student rated, *T* teacher-rated, *Read* reading, *sign.* significant

3.3 Links between teachers' and students' perceived intra-individual differences in student ability and importance across domains (RQ2)

Fig. 2 shows the final model for math and reading, including all covariances between the estimated math-reading LDS, as well as all predictive effects of student grade level, gender, parental education, cognitive ability, and school grades on the LDS. As shown in Fig. 2, rater-specific effects emerged such that the estimated math-reading LDS for ability and importance were significantly positively correlated within each perspective (student-rated LDS: $r = .44$, $SE = .06$, $p < .001$; teacher-rated LDS: $r = .52$, $SE = .07$, $p < .001$). Student- and teacher-rated math-reading LDS regarding students' ability were also positively correlated ($r = .43$, $SE = .05$, $p < .001$), whereas the corresponding correlation between student- versus teacher-rated LDS of students' importance was not significant ($r = .12$, $SE = .08$, $p = .139$). This pattern is consistent with our expectations and the plausible value analyses reported above, according to which greater discrepancies between teacher- and student-rated LDS emerged for importance value than for student ability (supporting H2). The same correlational patterns emerged with and without control variables (see supplemental materials).

3.4 Predictors of intra-individual differences in student ability and importance across domains (RQ3)

In order to explain the variance in intra-individual differences between students, we examined predictive effects of students' grade level, gender, parental education, cognitive ability, and prior school grades on the estimated LDS. Fig. 2 shows the final model including all predictors, which had satisfactory fit to the data (Satorra-Bentler $\chi^2 = 274.896$, $df = 137$, CFI = .961, RMSEA = .046, SRMR = .042). Of all student characteristics, only students' gender and prior math and reading grades

emerged as significant predictors of the estimated student- and teacher-rated LDS (gender-specific means, variances, and plausible values for all LDS are reported in the online supplemental materials).

Regarding differences in students' ability, the same pattern emerged for students' and teachers' LDS between the two subjects: Gender and math grade were significant positive predictors of the math-reading ability differences, as perceived by both students (male gender: $\beta = .20$, $SE = .05$, $p < .001$; math grade: $\beta = .24$, $SE = .08$, $p = .003$) and teachers (male gender: $\beta = .17$, $SE = .05$, $p = .001$; math grade: $\beta = .25$, $SE = .11$, $p = .017$). Moreover, prior reading grade had a significant negative effect on math-reading differences in students' ability, reported by both students ($\beta = -.23$, $SE = .09$, $p = .014$) and teachers ($\beta = -.23$, $SE = .10$, $p = .022$). All student characteristics jointly explained about 11% of the variance in student-rated LDS for self-concept of ability and 10% of the variance in teacher-rated LDS for students' ability.

Regarding cross-domain differences in student-perceived importance, none of the predictors had a significant effect, whereas differences in teacher-perceived importance between math and reading were significantly predicted by students' gender ($\beta = .19$, $SE = .06$, $p = .001$) and marginally significantly by students' prior math grade ($\beta = .14$, $SE = .07$, $p = .054$). These results suggest that teachers perceived comparatively more positive math-reading differences in students' importance value (i.e., greater differences favoring math and smaller differences favoring reading) for male than female students, as well as for students with comparatively higher prior math grades. The considered student characteristics explained about 2% of the variance in student-rated LDS for importance value and 6% of the variance in teacher-rated LDS for importance value. In summary, the results partially supported H3a and fully supported H3b.

4 Discussion

Drawing on the EVT assumption that students' personal hierarchy of their expectancy-value beliefs across different domains is affected by their teachers' beliefs (Eccles and Wigfield 2020), the present study investigated intra-individual cross-domain differences in both students' self-reports and their teachers' ratings of students' ability and importance across math and reading. Furthermore, we studied predictive effects of potentially influential student characteristics (e.g., gender and family background) that may intensify or reduce these intra-individual differences.

4.1 Intra-individual differences in students' math versus reading ability self-concept and importance, and corresponding teacher judgments

Our analyses revealed different associations for students' SCA and STV across the math and verbal domains, which is consistent with findings by Gaspard et al. (2018). Students' perceived importance of math and reading were positively associated, whereas students' math and verbal SCA were not significantly correlated. However, LDS and plausible value analyses revealed significant math-reading discrepancies in student-rated SCA (35%) and importance value (50%) for a substantial number

of students. These results indicate that even though students' motivational beliefs in different domains may be positively related to each other, individual students can nevertheless perceive considerable cross-domain differences. In addition, the size of the intra-individual differences varied substantially across students. Thus, further studies focusing on cross-domain discrepancies within individual students (e.g., by using latent difference scores) to investigate students' intra-individual hierarchies and their implications for students' academic choices and performances are warranted. In view of profile analyses suggesting that profile membership (i.e., groups of students with different cross-domain combinations of expectancy-value beliefs) can predict students' career aspirations (e.g., Oppermann et al. 2021), the extent of students' intra-individual differences may also play a role for their academic choices and achievement across domains.

Moreover, our results indicated that teachers' ratings of students' ability and importance were more strongly associated across domains than students' self-reports. This finding is in line with prior research reporting positive relations between teachers' ratings of their students' ability in the math and verbal domains (Dompnier et al. 2006). In addition, LDS revealed only about half as much intra-individual math-reading differences in teacher-rated students' ability and importance value compared to students' self-reports of SCA and perceived importance. This finding suggests that teachers' judgments may be affected by a halo effect (cf. Marsh et al. 1984). However, it is also possible that teachers' ratings reflect existing positive associations between students' abilities or motivated behaviors across domains, whereas students might differentiate their relative intra-individual abilities and beliefs (e.g., domain-specific strengths and preferences). Teachers might perceive less domain specificity in their students' ability and motivation in different subjects than students perceive themselves because teachers' judgments mostly rely on external (social) instead of internal dimensional comparison processes (cf. Marsh et al. 2015). For instance, teachers might have more information about other students in a subject, which supports social comparisons, whereas students also see themselves in subjects taught by other teachers, which provides more information for dimensional comparisons. Thus, different available information for students and teachers could lead to discrepancies in their perceptions of intra-individual differences across subjects (Huber and Seidel 2018).

The number of students with significant math-reading differences in their ability was approximately equal for students' own and their teachers' ratings. However, significant math-reading differences in students' importance emerged only about half as frequently in teachers' than in students' ratings. An explanation for this discrepancy might be that teachers have comparatively less information about students' subjective subject-specific importance than ability and may therefore tend to judge students' importance value based on students' displayed behavioral engagement (cf. Urhahne and Wijnia 2021), which might be more consistent across domains than students' subjective ratings of importance. In addition, cross-domain differences in students' importance rated by teachers and students were not significantly correlated, whereas intra-individual differences in teachers' perceptions of students' ability across domains showed a positive link to corresponding math-reading differences in students' SCA. These findings are in line with prior research showing that

teachers tend to assess their students' motivation more inaccurately than students' ability (Urhahne and Wijnia 2021) and imply that teachers also have difficulties to perceive intra-individual differences in their students' importance across math and reading. Nonetheless, if teachers do not accurately perceive the differences in students' valuing of the different domains, they will not be able to account for these differences in domain-specific feedback and may not provide the needed motivational support across domains. It might thus be beneficial to educate elementary teachers about the dimensional comparison processes in the formation of students' SCA and values across domains so that teachers can better adapt their instruction to meet the needs of students in both domains.

4.2 Predictors of intra-individual cross-domain differences in perceived ability-beliefs and importance

The extent of intra-individual differences varied across students, both for students' ratings of their SCA and importance, as well as, although significantly smaller, for teachers' corresponding judgments. Students' gender and prior grades in math and reading contributed to these differences among students. Prior achievements predicted intra-individual differences in teachers' ratings of students' ability, as well as students' ratings of their SCA. This result expands previous evidence on positive effects of students' prior achievement on teachers' perceptions of students' ability (e.g., Jussim et al. 1996) and students' SCA (e.g., Gaspard et al. 2018) within a single domain to differences between domains. Moreover, whereas no significant gender differences emerged in the intra-individual hierarchy of students' ratings of their importance of math and reading, male and female students differed on average in their hierarchy of SCA across domains, with males being more likely to have a higher SCA in math than reading. This finding is partly consistent with prior research showing the overrepresentation of boys in a "math-motivated" profile (e.g., Oppermann et al. 2021). However, our results expand prior evidence by indicating that the extent of these intra-individual differences differs between males and females: among students with higher SCA in math than reading, males compared to females tend to have greater intra-individual differences, whereas among students with higher SCA in reading than math, males tend to have smaller intra-individual differences. This finding supports the previously raised suggestion that not only different cross-domain hierarchies of motivational beliefs for male and female students (e.g., Gaspard et al. 2019), but also different degrees of perceived differences can explain some of the gender disparities in students' career choices in the math vs. verbal domain (cf. Jansen et al. 2021; Wang et al. 2013). Thus, further research that investigates predictive effects of cross-domain, intra-individual differences on students' career aspiration and choices is needed.

Similarly, our results suggest that teachers tend to assign higher abilities and importance in math compared to reading to males and higher abilities and importance in reading compared to math to females. Cross-domain differences in teachers' and students' ratings of ability might reflect underlying gender differences in students' ability across subjects (cf. Wang et al. 2013). However, teachers' perceptions of differences could also be driven by gender stereotypes (Eccles 2009; Jussim et al.

1996) and might influence students' self-perceptions through different instructional behaviors, such as varying levels of support across domains or communicating perceived weaknesses and strengths through domain-specific feedback (cf. Gentrup et al. 2020). Given the importance of intra-individual differences across domains for students' specialization (e.g., career preferences, Gaspard et al. 2019), further studies should examine potential predictive effects of intra-individual differences in teachers' ratings of their students' ability and motivational beliefs across domains on students' motivational beliefs to understand how these gender differences in students' motivational hierarchies emerge.

A large amount of variance in intra-individual differences in ability and importance across domains remained unexplained for both students' and teachers' ratings. Other influencing factors such as parents' values of the different domains might explain differences in the hierarchies of importance across domains among students (Eccles 2009; Eccles and Wigfield 2020). Furthermore, as our analyses focused on the individual student level, no class or teacher characteristics were considered as potential moderators of the cross-domain differences. However, a higher level of teacher expertise may lead to a weaker halo effect (cf. Kozlowski and Kirsch 1987) and could thus intensify intra-individual differences in teachers' perceptions of their students. Moreover, based on social comparison effects, classroom composition factors such as average achievement could influence both teachers' (Dompnier et al. 2006) and students' (Chiu 2012) ratings of students' ability across domains. Thus, future multi-level analyses should examine the potential effects of class or teacher characteristics on intra-individual differences in students' motivational beliefs.

4.3 Limitations

Some limitations of the present study must be considered. First of all, our analyses were restricted to a specific sample of elementary students and teachers in the 1990s in the United States. Although diagnosing individual students' learning needs was already included in the U.S. Standards for Teacher Competence in Educational Assessment of Students back in 1990 (American Federation of Teachers, the National Council on Measurement in Education, and the National Education Association 1990), the focus on individualized instruction (cf. Smale-Jacobse et al. 2019; Tomlinson et al. 2003) and teachers' judgment of individual students has increased in the last decades (cf. Urhahne and Wijnia 2021), and teachers might thus have become more sensitive towards individual student needs. However, studies based on more recent data still provide evidence in support of the halo effect, showing that teachers underestimate the intra-individual variability within their students' motivation and abilities (within a subject; Huber and Seidel 2018; Sanrey et al. 2020). In addition, consistent with our results, recent studies also indicate differing dimensional comparison processes for students' self-ratings and teachers' ratings of their students' abilities (Helm et al. 2018). Moreover, domain-specific gender biases in teachers' judgments of their students' abilities continue to be an issue today. Studies using more recent data have reported that teachers rate boys' compared to girls' ability higher in math (e.g., Heyder et al. 2019) and girls' compared to boys' ability higher in reading (e.g., Muntoni and Retelsdorf 2018), even when girls and boys have sim-

ilar ability levels. However, beliefs about gender norms and equality as well as the related gender gaps in students' math and reading achievement differ across countries (Eriksson et al. 2020). Thus, our findings might not be generalizable across contexts and cultures. Further studies are therefore needed to investigate students' intra-individual cross-domain differences in other school contexts.

Second, we focused on differences in teachers' and students' ratings between two core subjects, namely, math and reading. However, dimensional comparisons depend on the degree of similarity of the subjects being compared (Möller et al. 2020) so that the estimated intra-individual differences are likely to vary across school subjects. Third, we focused on one specific STV facet, namely students' importance of achievement. However, dimensional comparisons seem to vary across different STV (Gaspard et al. 2018). Future studies might therefore investigate intra-individual cross-domain differences in students' other STV perceived by students' and teachers. Fourth, as we used cross-sectional data, it remains unclear how the association between teachers' and students' perceptions of intra-individual differences in students' ability and motivation evolves. Further longitudinal research is necessary to examine how cross-domain differences in teachers' and students' perceptions interact with each other and develop over time.

4.4 Conclusions

This study is the first to systematically analyze intra-individual differences in students' and teachers' perceptions of students' ability and importance across the two domains of math and reading. Overall, the findings suggest that teachers perceive fewer intra-individual differences in students' ability and importance value across domains than do students. Moreover, differences in teachers' ratings were at least partly related to differences in students' motivational beliefs. Thus, further studies on reciprocal influences of intra-individual differences in teachers' and students' perceptions of students' ability and motivation across domains and corresponding effects on teachers' instructional behavior and students' academic development are warranted. Furthermore, the present study demonstrated that latent difference score modeling is a useful method to directly investigate intra-individual hierarchies of students' motivational beliefs, taking into account the perspectives of students and teachers simultaneously. This methodological approach may thus provide further opportunities to examine the implications of dimensional comparisons and intra-individual cross-domain differences for students' academic achievements, choices, and career aspirations.

References

American Federation of Teachers, the National Council on Measurement in Education, and the National Education Association (1990). Standards for teacher competence in educational assessment of students. Washington, DC: The National Council on Measurement in Education. https://buros.org/standards-teacher-competence-educational-assessment-students. Accessed: 16 July 2021.

Asparouhov, T., & Muthén, B. (2010). Plausible values for latent variables using Mplus. Technical report. http://www.statmodel.com/download/Plausible.pdf. Accessed: 10 Mar 2021.

Brandmiller, C., Dumont, H., & Becker, M. (2020). Teacher perceptions of learning motivation and classroom behavior: the role of student characteristics. *Contemporary Educational Psychology*, *63*, 101893. https://doi.org/10.1016/j.cedpsych.2020.101893.

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

Chiu, M.-S. (2012). The internal/external frame of reference model, big-fish-little-pond effect, and combined model for mathematics and science. *Journal of Educational Psychology*, *104*(1), 87–107. https://doi.org/10.1037/a0025734.

Dompnier, B., Pansu, P., & Bressoux, P. (2006). An integrative model of scholastic judgments: Pupils' characteristics, class context, halo effect and internal attributions. *European Journal of Psychology in Education*, *21*, 119–133. https://doi.org/10.1007/BF03173572.

Durik, A.M., Vida, M., & Eccles, J.S. (2006). Task values and ability beliefs as predictors of high school literacy choices: a developmental analysis. *Journal of Educational Psychology*, *98*(2), 382–393. https://doi.org/10.1037/0022-0663.98.2.382.

Eccles, J. (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.

Eccles, J.S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: a developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, *61*, 101859. https://doi.org/10.1016/j.cedpsych.2020.101859.

Eccles, J.S., Adler, T.F., Futterman, R., Goff, S.B., Kaczala, C.M., Meece, J.L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J.T. Spence (Ed.), *Expectancies, values, and academic behaviors* (pp. 75–146). W.H. Freeman.

Eccles, J.S., Wigfield, A., Harold, R.D., & Blumenfeld, P. (1993). Age and gender differences in children's self-and task perceptions during elementary school. *Child Development*, *64*(3), 830–847. https://doi.org/10.2307/1131221.

Eriksson, K., Björnstjerna, M., & Vartanova, I. (2020). The relation between gender egalitarian values and gender differences in academic achievement. *Frontiers in Psychology*, *11*, 1–14. https://doi.org/10.3389/fpsyg.2020.00236.

Gaspard, H., Wigfield, A., Jiang, Y., Nagengast, B., Trautwein, U., & Marsh, H.W. (2018). Dimensional comparisons: How academic track students' achievements are related to their expectancy and value beliefs across multiple domains. *Contemporary Educational Psychology*, *52*, 1–14. https://doi.org/10.1016/j.cedpsych.2017.10.003.

Gaspard, H., Wille, E., Wormington, S.V., & Hulleman, C.S. (2019). How are upper secondary school students' expectancy-value profiles associated with achievement and university STEM major? A cross-domain comparison. *Contemporary Educational Psychology*, *58*, 149–162. https://doi.org/10.1016/j.cedpsych.2019.02.005.

Gaspard, H., Lauermann, F., Rose, N., Wigfield, A., & Eccles, J. (2020). Cross-domain trajectories of students' ability self-concepts and intrinsic values in math and language arts. *Child Development*. https://doi.org/10.1111/cdev.13343.

Geiser, C., Eid, M., Nussbeck, F. W., Courvoisier, D. S., & Cole, D. A. (2010). Analyzing true change in longitudinal multitrait-multimethod studies: application of a multimethod change model to depression and anxiety in children. *Developmental Psychology*, *46*(1), 29–45. https://doi.org/10.1037/a0017888.

Gentrup, S., Lorenz, G., Kristen, C., & Kogan, I. (2020). Self-fulfilling prophecies in the classroom: teacher expectations, teacher feedback and student achievement. *Learning and Instruction*, *66*, 101296. https://doi.org/10.1016/j.learninstruc.2019.101296.

Hardy, I., Decristan, J., & Klieme, E. (2019). Adaptive teaching in research on learning and instruction. *Journal for educational research online*, *11*, 169–191.

Helm, F., Müller-Kalthoff, H., Mukowski, R., & Möller, J. (2018). Teacher judgment accuracy regarding students' self-concepts: affected by social and dimensional comparisons? *Learning and Instruction*, *55*, 1–12. https://doi.org/10.1016/j.learninstruc.2018.02.002.

Heyder, A., Steinmayr, R., & Kessels, U. (2019). Do teachers' beliefs about math aptitude and brilliance explain gender differences in children's math ability self-concept? *Frontiers in Education*, *4*(34), 1–11. https://doi.org/10.3389/feduc.2019.00034.

Huber, S. A., & Seidel, T. (2018). Comparing teacher and student perspectives on the interplay of cognitive and motivational-affective student characteristics. *PLoS One*. https://doi.org/10.1371/journal.pone.0200609.

Jansen, M., Becker, M., & Neumann, M. (2021). Dimensional comparison effects on (gendered) educational choices. *Journal of Educational Psychology*, *113*(2), 330–350. https://doi.org/10.1037/edu0000524.

Jussim, L., Eccles, J., & Madon, S. (1996). Social perception, social stereotypes, and teacher expectations: accuracy and the quest for the powerful self-fulfilling prophecy. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 28, pp. 281–388). Amsterdam: Elsevier. https://doi.org/10.1016/S0065-2601(08)60240-3.

Kozlowski, S. W., & Kirsch, M. P. (1987). The systematic distortion hypothesis, halo, and accuracy: an individual-level analysis. *Journal of Applied Psychology*, *72*(2), 252–261. https://doi.org/10.1037/0021-9010.72.2.252.

Lauermann, F., Chow, A., & Eccles, J. S. (2015). Differential effects of adolescents' expectancy and value beliefs about math and english on math/science-related and human services-related career plans. *International Journal of Gender, Science and Technology*, *7*(2), 205–228.

Lauermann, F., Tsai, Y.-M., & Eccles, J. S. (2017). Math-related career aspirations and choices within Eccles et al.'s expectancy-value theory of achievement-related behaviors. *Developmental Psychology*, *53*(8), 1540–1559. https://doi.org/10.1037/dev0000367.

Marsh, H. W. (1986). Verbal and math self-concepts: an internal/external frame of reference model. *American Educational Research Journal*, *23*(1), 129–149. https://doi.org/10.3102/00028312023001129.

Marsh, H. W., & Hau, K.-T. (1996). Assessing goodness of fit. *The Journal of Experimental Education*, *64*(4), 364–390. https://doi.org/10.1080/00220973.1996.10806604.

Marsh, H. W., Smith, I. D., & Barnes, J. (1984). Multidimensional self-concepts: relationships with inferred self-concepts and academic achievement. *Australian Journal of Psychology*, *36*(3), 367–386. https://doi.org/10.1080/00049538408255318.

Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit in structural equation models. In J. J. McArdle (Ed.), *Contemporary psychometrics: a Festschrift for Roderick P. McDonald* (pp. 275–340). New Jersey: Erlbaum.

Marsh, H. W., Abduljabbar, A. S., Parker, P. D., Morin, A. J. S., Abdelfattah, F., Nagengast, B., Möller, J., & Abu-Hilal, M. M. (2015). The internal/external frame of reference model of self-concept and achievement relations: age-cohort and cross-cultural differences. *American Educational Research Journal*, *52*(1), 168–202. https://doi.org/10.3102/0002831214549453.

Möller, J., & Marsh, H. W. (2013). Dimensional comparison theory. *Psychological Review*, *3*, 544–560. https://doi.org/10.1037/a0032459.

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.

Muntoni, F., & Retelsdorf, J. (2018). Gender-specific teacher expectations in reading—The role of teachers' gender stereotypes. *Contemporary Educational Psychology*, *54*, 212–220. https://doi.org/10.1016/j.cedpsych.2018.06.012.

Nicholson, C.L., & Hibpshman, T.H. (1990). *Administration and scoring manual for the Slosson Intelligence Test-Revised*. East Aurora: Slosson Educational Publications.

Nisbett, R.E., & Wilson, T.D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/0022-3514.35.4.250.

Oppermann, E., Vinni-Laakso, J., Juuti, K., Loukomies, A., & Salmela-Aro, K. (2021). Elementary school students' motivational profiles across Finnish language, mathematics and science: Longitudinal trajectories, gender differences and STEM aspirations. *Contemporary Educational Psychology*, *64*, 101927. https://doi.org/10.1016/j.cedpsych.2020.101927.

Pohlmann, B., Möller, J., & Streblow, L. (2004). Zur Fremdeinschätzung von Schülerselbstkonzepten durch Lehrer und Mitschüler [On students' self-concepts inferred by teachers and classmates]. *Zeitschrift Für Pädagogische Psychologie*, *18*(3/4), 157–169. https://doi.org/10.1024/1010-0652.18.4.157.

Sanrey, C., Bressoux, P., Lima, L., & Pansu, P. (2020). A new method for studying the halo effect in teachers' judgement and its antecedents: Bringing out the role of certainty. *British Journal of Educational Psychology*. https://doi.org/10.1111/bjep.12385.

Scalas, L.F., Marsh, H.W., Morin, A.J.S., & Nagengast, B. (2014). Why is support for Jamesian actual-ideal discrepancy model so elusive? A latent-variable approach. *Personality and Individual Differences*, *69*, 62–68. https://doi.org/10.1016/j.paid.2014.05.010.

Slosson, R.L., Nicholson, C.L., & Hibpshman, T.H. (1991). *Slosson intelligence test-revised*. East Amora: Slosson Educational Publications.

Smale-Jacobse, A.E., Meijer, A., Helms-Lorenz, M., & Maulana, R. (2019). Differentiated instruction in secondary education: a systematic review of research evidence. *Frontiers in Psychology*, *10*, 2366–2366. https://doi.org/10.3389/fpsyg.2019.02366.

Steinmayr, R., Weidinger, A.F., Schwinger, M., & Spinath, B. (2019). The importance of students' motivation for their academic achievement—replicating and extending previous findings. *Frontiers in Psychology*, *10*, 1730–1730. https://doi.org/10.3389/fpsyg.2019.01730.

Tomlinson, C., Brighton, C., Hertberg, H.L., Callahan, C., Moon, T.R., Brimijoin, K., Conover, L.A., & Reynolds, T. (2003). Differentiating instruction in response to student readiness, interest, and learning profile in academically diverse classrooms: a review of literature. *Journal for the Education of the Gifted*, *27*, 119–145. https://doi.org/10.1177/016235320302700203.

Urhahne, D., & Wijnia, L. (2021). A review on the accuracy of teacher judgments. *Educational Research Review*, *32*, 100374. https://doi.org/10.1016/j.edurev.2020.100374.

Van de Schoot, R., Lugtig, P., & Hox, J. (2012). Developmetrics: a checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*, 486–492. https://doi.org/10.1080/17405629.2012.686740.

Wang, M.-T., Eccles, J.S., & Kenny, S. (2013). Not lack of ability but more choice: individual and gender differences in choice of careers in science, technology, engineering, and mathematics. *Psychological Science*, *24*(5), 770–775. https://doi.org/10.1177/0956797612458937.

Wang, S., Rubie-Davies, C.M., & Meissel, K. (2018). A systematic review of the teacher expectation literature over the past 30 years. *Educational Research and Evaluation*, *24*(3–5), 124–179. https://doi.org/10.1080/13803611.2018.1548798.

Wigfield, A., & Eccles, J.S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 68–81. https://doi.org/10.1006/ceps.1999.1015.

Wigfield, A., Eccles, J., Yoon, K.S., Harold, R., Arbreton, A., Freedman-Doan, C., & Blumenfeld, P. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: A 3-year study. *Journal of Educational Psychology*, *89*, 451–469. https://doi.org/10.1037/0022-0663.89.3.451.

Wigfield, A., Eccles, J.S., & Möller, J. (2020). How dimensional comparisons help to understand linkages between expectancies, values, performance, and choice. *Educational Psychology Review*, *32*(3), 657–680. https://doi.org/10.1007/s10648-020-09524-2.

Zhu, M., & Urhahne, D. (2014). Assessing teachers' judgements of students' academic motivation and emotions across two rating methods. *Educational Research and Evaluation*, *20*(5), 411–427. https://doi.org/10.1080/13803611.2014.964261.

Zhu, M., Urhahne, D., & Rubie-Davies, C.M. (2018). The longitudinal effects of teacher judgement and different teacher treatment on students' academic outcomes. *Educational Psychology*, *38*(5), 648–668. https://doi.org/10.1080/01443410.2017.1412399.