Learning and Individual Differences 107 (2023) 102346

Contents lists available at ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

Dynamic relations between motivation and performance across content in a mathematics learning technology

Teomara Rutherford *, Hye Rin Lee, Kerry Duck

School of Education, College of Education and Human Development, University of Delaware, 113 Willard Hall Education Building, 16 West Main Street, Newark, DE 19716, USA

ARTICLE INFO

Keywords:
Mathematics education
Learning technologies
Motivation
Experience sampling
Situated expectancy-value theory

ABSTRACT

Motivation is theorized to be situated and dynamic, changing across contexts and time as students interact with learning materials. To capture these dynamic relations, 9091 third through fifth graders were surveyed multiple times throughout the year as they engaged with mathematics objectives in the learning technology, ST Math. All constructs displayed greater variance in motivation within students than between students among third graders; however, variance within students declined as grade-level increased. Within students, interest and utility positively predicted subsequent performance; effort cost negatively predicted subsequent performance. After partialling out within-student associations, contextual associations of objective-content motivation and performance were smaller and often in the opposite direction; broader measures of mathematics motivation had largely null relationships with performance. Results provide insight into how motivation and performance may relate during decision-making and application of effort in classroom activities, and how these relations may be different than those considering average levels across students.

Educational relevance and implications statement

As elementary students engaged with a year-long mathematics learning technology, when students were more interested in an objective and they noted its content was more useful for their learning, they were more likely to perform well on its post-quiz. The association between usefulness and performance was strongest for fifth graders. Ratings of difficulty for content, which were conceptualized as objective effort cost, had negative associations with performance. By understanding how motivation and performance relate at this level, educators and content developers can make adjustments to their materials and instruction to maximize each. Results suggest that, at this age-level, enhancing student interest in the content and reducing their perceptions of difficulty may both be especially fruitful avenues toward improving performance.

1. Introduction

Motivation has a long-standing relationship with schooling outcomes: students who are confident and who value the activities in school perform better and engage more deeply (Eccles & Wigfield, 2020). Motivation is theorized to operate dynamically in the classroom—students' motivation likely ebbs and flows throughout the school day and across different activities (Kaplan, 2014, 2015; Kaplan & Patrick, 2016). This dynamic nature is difficult to capture—most motivation research relates broader measures of motivation (e.g., self-efficacy for mathematics, Usher & Pajares, 2009) to performance, often focusing on broader measures of performance as well (e.g., end-of-term grades, Lee & Kung, 2018; Weidinger, Steinmayr, & Spinath, 2017; annual standardized tests, Dicke et al., 2018; Garon-Carrier et al., 2015). This work typically characterizes motivation at the student level, looking between students to examine how dispositional, trait-like, motivation relates to performance (e.g., Aunola, Leskinen, & Nurmi, 2006). Research that does examine motivation within students often still examines dispositions, but for different subjects (e.g., motivation for mathematics vs. language, Gaspard et al., 2018). More work is needed that examines how motivation for related tasks fluctuates *within* students and how this fluctuation relates to performance. Researchers focusing on experience sampling methods (ESM) have begun to unpack these dynamic associations (e.g., Beymer, Benden, & Sachisthal, 2022; Dietrich, Schmiedek, & Moeller, 2022; Ketonen, Dietrich, Moeller, Salmela-Aro, & Lonka, 2018). Although ESM work is a growing area of research

* Corresponding author at: University of Delaware, 113 Willard Hall Education Building, 16 West Main Street, Newark, DE 19716, USA.
E-mail address: teomara@udel.edu (T. Rutherford).

https://doi.org/10.1016/j.lindif.2023.102346
Received 22 March 2023; Received in revised form 5 June 2023; Accepted 7 August 2023
Available online 20 August 2023

in dynamic motivation, much of it focuses on high school or university students (e.g., Martin, Mansour, & Malmberg, 2020; Niepel, Marsh, Guo, Pekrun, & Möller, 2022). Exploring the dynamics of motivation in younger children holds value—younger students may experience greater fluctuations in motivation for activities due to less solidified senses of identity (Erikson, 1968) and emerging regulation ability (McClelland, John Geldhof, Cameron, & Wanless, 2015). Understanding dynamic motivation among this age-group can contribute to theories regarding motivation development and positive classroom practice.

In this study, we leverage data collected as students complete content within a mathematics learning technology, Spatial Temporal (ST) Math, to examine topic-specific student reports of motivation (state-based motivation reports) and how these reports relate to performance over the course of a year, both within and between students. This method allows insight into dynamic relations of motivation and performance with an age-group and length of time less often studied in motivation research.

1.1. Theoretical framework

We frame our work within *Situated Expectancy–Value Theory* (SEVT, Eccles & Wigfield, 2020), previously known as *Expectancy–Value Theory*. SEVT posits that student motivation consists of two overarching dimensions: expectancies for success and subjective task values. Expectancies for success are defined in terms of an individual's beliefs about how well they will perform on an upcoming task (Eccles et al., 1983) and are conceptually similar to other self-beliefs, such as self-concept (Marsh, 1990; Marsh & Seaton, 2013), self-efficacy (; Bandura, 1977), and perceived competence (Deci & Ryan, 2012). Subjective task values consist of four components of value: interest (or intrinsic); utility (or usefulness); attainment (or importance); and cost, including the emotional, cognitive, or effort drain involved in an activity (Eccles et al., 1983; Flake, Barron, Hulleman, McCoach, & Welsh, 2015; Perez, Cromley, & Kaplan, 2014).

As conceptualized in Eccles et al. (1983) and further articulated in Eccles and Wigfield (2020), SEVT applies to individuals' engagement with tasks from the broad level of choosing a career to the more close-grained engagement with a particular task (e.g., a specific fractions problem). Individuals may ask themselves a million times a day, "Can I do this [task]?" and "Do I want to do this [task]?" Once an individual has answered these questions, they make a choice of whether and how to engage in a specific task; this engagement has implications for performance and achievement (Eccles & Wigfield, 2020). As Eccles and Wigfield (2020) reiterate, SEVT is entirely situated—in context, in time, in the person and their affiliated groups; Eccles and Wigfield note the promise of methods, such as ESM, for capturing the complexity of individuals' momentary perceptions and choice (see also, Eccles, 2022).

1.2. Mathematics motivation and performance

Prior literature has found statistically significant relations between motivation and performance (e.g., Aunola et al., 2006; Dicke et al., 2018; Marsh et al., 2018). These findings have been supported across different contexts, such as age (Dicke et al., 2018; Pinxten, Marsh, De Fraine, Van Den Noortgate, & Van Damme, 2014), performance indicators (grades versus test scores; Marsh et al., 2016, 2018), country (Lee & Kung, 2018), and gender (Sewasew, Schroeders, Schiefer, Weirich, & Artelt, 2018). These motivation/performance relations are especially strong when considering self-beliefs, such as expectancy or self-efficacy (e.g., Bandura, 1977; Pintrich & De Groot, 1990; Stringer & Heath, 2008). Few studies, however, have examined the relations between motivation and performance within students. As one recent exception, Niepel et al. (2022) examined within-student relations between motivation and *perceived* performance among a sample of German secondary students. They found that mathematics self-concept of ability predicted the subsequent lesson's perceived performance. It is unclear as to whether these results would generalize to more objective measures of performance.

Although there are no studies that have investigated the link between task values and performance using experience sampling, a number of prior studies have examined the relationship between more trait-like mathematics task values and performance (e.g., Hong, Yoo, You, & Wu, 2010; Saw & Chang, 2018; Weidinger, Spinath, & Steinmayr, 2020). For example, Weidinger et al. (2020) found that there was a relationship between mathematics subjective task values (i.e., intrinsic, attainment, and utility value) and grades. They also found that cross-lagged associations between subjective task values and grades did not statistically significantly differ between subjective task value components. Similarly, Saw and Chang (2018) found that mathematics performance statistically significantly predicted subjective task value for the full, Hispanic, White, Black, and Asian sample. But, mathematics subjective task value only statistically significantly predicted performance for the White sample. Although cost is seen as a key subjective task value component, studies in mathematics to date have not typically included cost in models examining the relationship between subjective task value and performance.

Even though many prior studies have reported statistically significant relations between mathematics motivation and performance (e.g., Aunola et al., 2006; Dicke et al., 2018; Marsh et al., 2018), there are also inconsistent results regarding the presence and relative strength of these relationships (e.g., Nuutila, Tuominen, Tapola, Vainikainen, & Niemivirta, 2018). One likely reason for this inconsistency is age. For instance, Weidinger, Steinmayr, and Spinath (2018) found that students' mathematics competence beliefs became more important for predicting performance with age. In particular, at the end of second and beginning of third grade, prior performance was important for later mathematics competence beliefs, whereas prior mathematics competence beliefs did not statistically significantly relate to later performance. Then at the end of fourth grade, there was a statically significant bidirectional relation between competence beliefs and performance. Liu and colleagues (2022) also found that expectancy for mathematics was more predictive of mathematics performance in fifth grade than in third and fourth grades, and that these differences varied across tests of different content, suggesting that variance within students regarding the relationship between motivation and performance may be tied to both age and content.

Additionally, across studies examining the relationship between motivational beliefs and performance, variation in associations may be due to the particular performance metric; scholars have measured performance using both standardized test scores (e.g., Dicke et al., 2018; Garon-Carrier et al., 2015) and grades (e.g., Lee & Kung, 2018; Weidinger et al., 2017). For example, Marsh, Trautwein, Lüdtke, Köller, and Baumert (2005); Marsh et al. (2016, 2018) found that self-concept is more strongly correlated with school grades compared to test scores. However, Preckel et al. (2017) found that the associations between mathematics self-concept and test scores were larger than self-concept and grades. More research is needed to untangle the relationship between motivational beliefs and performance when different outcomes are used across different situations. In particular, with the exception of Niepel et al.'s (2022) study, few studies examine the association between motivation and performance indicators with more formative assessments, such as classroom lessons, activities, or quizzes. Further, studies to date have not systematically compared the association between the different facets of value and various performance indicators in one model.

1.3. The study of dynamic motivation

Given the situated nature of motivation (Eccles & Wigfield, 2020), comparing broad measures of performance and motivation may not capture the dynamic processes at play between motivation and performance across time and contexts. One way researchers have modeled dynamic aspects of motivation is through using data on motivation in the moment, such as through ESM (e.g., Csikszentmihalyi & Larson, 2014; Hektner & Csikszentmihalyi, 1996). Using these methods, researchers collect in-the-moment self-reported data typically leveraging online mobile technology to measure state-like motivational beliefs within real-world settings during or in very close proximity to an activity (Shiffman, Stone, & Hufford, 2008). These multiple observations can then be examined to estimate the stability and variance of constructs across contexts or time through examination of inter- (i.e., between) and intra- (i.e., within) individual variability (Barbot, 2022; Csikszentmihalyi & Larson, 2014; Gabriel et al., 2019).

Although social science research has traditionally focused on inter-individual variability, a growing area of research examines intra-individual variability, as it may contain vital information on parsing out relationships between constructs outside of often unrelated individual differences (Fiske & Rice, 1955). Within motivation research, deeper insights into processes of motivation and engagement can be gathered in the moment than can be gathered with retrospective recall (Dietrich et al., 2022). Prior research using ESM in motivation research has investigated students' self-beliefs (e.g., Nissen & Shemwell, 2016), their achievement goals (e.g., Goetz, Sticca, Pekrun, Murayama, & Elliot, 2016; Lee & Bong, 2022), their engagement (e.g., Martin et al., 2020; Milesi, Perez-Felkner, Brown, & Schneider, 2017; Shernoff et al., 2016; Xie, Heddy, & Greene, 2019), and their values, such as interest (e.g., Beymer, Rosenberg, & Schmidt, 2020; Shumow, Schmidt, & Zaleski, 2013) and utility (e.g., Dietrich, Moeller, Guo, Viljaranta, & Kracke, 2019). Depending on the motivation construct measured, researchers have found differences with respect to whether most of the variance in motivation is between individuals (e.g., Lee & Bong, 2022, achievement goals; Martin et al., 2020, "adaptive motivation" combining self-efficacy, values, and mastery motivation) or within individuals across tasks or time (e.g., Fullagar & Kelloway, 2009, flow; Goetz et al., 2010, academic emotions; Xie et al., 2019, engagement).

The stability of motivation within individuals may also vary depending on age—even outside of experience sampling studies, motivation constructs, such as intrinsic motivation, and related constructs, such as emotional regulation, have demonstrated increasing stability with age (Benson et al., 2019; Gottfried, Fleming, & Gottfried, 2001). However, there have been few developmentally-focused studies of dynamic motivation. Comparing across studies of differing ages may offer some insight, but direct comparisons are challenging, given differences in constructs, measurement intervals, and contexts. Examining junior high students, Martin et al. (2020) found that there was a greater inter- (0.62) rather than intra-individual (0.34) variability in value for mathematics and English across lessons. Among high schoolers, previous studies have found evidence of more within-individual than between-student variation, although not exclusively (cf Becker, Goetz, Morger, & Ranellucci, 2014; Beymer et al., 2020; Goetz et al., 2016; Hausen, Möller, Greiff, & Niepel, 2022; Niepel et al., 2022; Moeller, Brackett, Ivcevic, & White, 2020). For example, Beymer, Rosenberg, Schmidt, and Naftzger (2018) found that there was greater intra- rather than inter-individual variation in engagement, affect, and learning. Similarly, prior studies, such as Goetz et al. (2016) and Moeller et al. (2020) found that there was higher within-person variability in emotions and goals. On the other hand, Hausen et al. (2022) found that only 25 % of their motivation outcome, academic self-concept, varied at the within person-level. Among college students, some studies have found more variation between rather than within individuals (Lee & Bong, 2022), whereas others found more variation within students across situations rather than between individuals (Ketonen et al., 2018, 2019). Studies of students before adolescence are rare.

Student reports of motivation and emotions from ESM can also be examined for their relation between time- and context-varying outcomes. For example, Lee and Bong (2022) reviewed student goal adoptions in the weekend prior to an examination. They found that mastery goals predicted deeper cognitive strategy use between students, but that this association was not seen within students across time. However, within students, ability-approach and avoidance goals predicted strategy use and anxiety. Capelle, Grunschel, Bachmann, Knappe, and Fries (2022) found that as the time of an examination approached, students studied more but also experienced greater salience for the loss of valued alternatives aspect of the cost construct.

In the Eccles and Wigfield (2020) SEVT model, expectancies and values are only shown once, at the far right of the model, as immediate antecedent to achievement-related choices and performance. However, in a model considering a more immediate choice than selecting a college major or similar—for example, deciding how to engage with a specific mathematics learning technology objective—it is likely that more trait-like expectancies and values (e.g., value for mathematics, generally) would also predict engagement and performance through state motivation. There is some evidence to support this assertion. Martin et al. (2020) found that general academic motivation predicted momentary mathematics and English motivation among a sample of junior high students, especially reducing variance in these measures between students. Dietrich et al. (2019) also found relations between start-of-course motivation and in-the-moment profiles of expectancies and values among undergraduates. Generally, higher values at the start of course related to more motivated profiles during the course; however, high utility and attainment at the course start also predicted membership in higher cost profiles. Both studies indicate that more trait-like motivation contributes to state motivation, but the methodological and age differences make it difficult to directly compare the studies or to draw conclusions regarding how a fuller suite of SEVT constructs (expectancy, values, cost) might predict momentary motivation in a younger sample.

The prior research has presented some evidence of the value in state measures of motivation—that these constructs vary across time and context within individuals and that this variation relates to variation in strategy use (Lee & Bong, 2022) and perceived performance (Niepel et al., 2022) speaks to dynamic relations between motivation and outcomes. However, work linking motivational variation with objective measures of time- or context-varying performance is lacking. In particular, even among studies of trait-like motivation and performance, performance on frequent classroom tasks is seldom investigated—except when aggregated as grades, which often include substantial consideration of student behavior (Bowers, 2011). Understanding these links can paint a clearer picture of how motivation and performance may relate in the moment of individual decision making and application of effort in classroom activities. These links may be especially important in learning technologies where students must often engage in independent work, such as when teachers are otherwise engaged as students work in center rotations (see Peddycord-Liu et al., 2019). In addition, the bulk of dynamic motivation research leveraging ESM has been conducted with adolescent and older students. Longitudinal studies of motivation and related concepts suggest greater stability with age (Benson et al., 2019; Gottfried et al., 2001). If this carries through to in-the-moment motivation, younger students are likely to experience greater variation in their motivational states across time or context. This greater variance may reveal different relationships between dynamic motivation and performance than seen with older populations.

1.4. Context and current study

The current study is part of a larger NSF-funded project investigating data from and improvement with the mathematics learning software, Spatial Temporal (ST) Math. ST Math is a year-long web-based supplemental instructional software aligned with grade-level standards. ST Math displays grade-level learning content divided into approximately 30 objectives that cover specific learning goals (e.g., the fourth grade objective of "Fractions—Equivalence and Ordering"). Before and after each objective, students answer objective-content quiz questions; after objectives, students report on their perceptions of the content. Within objectives, students play games that represent mathematics content using a consistent visual representation; games are divided into levels of increasing difficulty; each level contains a number of puzzles where the students solve a visually-represented mathematics problem to help Jiji the penguin leave the screen (e.g., from left to right or from bottom to top). Fig. 1 shows an example ST Math puzzle. ST Math is largely self-paced with students playing either a set curriculum or a curriculum ordered by their teacher. In order to progress through ST Math, students must solve each level in order—they may not move on to the next level until they have completed all puzzles in the current level successfully. Students who fail a level can immediately reattempt the level or play any previously-passed content. Prior studies of ST Math have found that use of ST Math improves student performance on state standardized tests and improves student self-beliefs about mathematics (Rutherford et al., 2014; Rutherford, Liu, Lam, & Schenke, 2019; Wendt, Rice, & Nakamoto, 2019). Within the project, a version of ST Math was implemented that contained enhanced motivational surveys.

We use the term "game" as this is a term used within ST Math itself. ST Math has features of games, such as a set of rules constraining play, interactivity and feedback, and content designed to meet each player where they are by optimizing challenge (Garris, Ahlers, & Driskell, 2002; Gee, 2003). ST Math also as an, albeit surface-level, fantastical storyline, whereby students must help Jiji the penguin overcome a series of obstacles to progress out of each puzzle game screen. Although these elements are those that make ST Math game-like, others may view the platform as an interactive tutorial instead. The focus on our study is the use of ST Math as a context within which to study motivation and therefore we do not elaborate on the specific mechanisms through which ST Math is theorized to support mathematics learning; interested readers may reference Schenke, Rutherford, and Farkas (2014), Graziano, Peterson, and Shaw (1999), Kibrick (2013), and Krumm, Coulson, and Neisler (2022).

A large school district in Florida participating in the NSF project provided data for the current study. Students within the district played ST Math during their normal school day and provided answers about their perceptions of ST Math objectives and their mathematics motivation. These answers, along with data on students' performance within ST Math, were collected through the software, matched with district data, and de-identified before being shared with the researchers. The data contain answers from students about their more dispositional trait-like motivation for mathematics (expectancies and values) and about their state-based perceptions of interest, utility, and effort cost for specific mathematics content within the software (i.e., objectives). With these data, we ask:

1. To what extent does motivation state vary within students across mathematics objectives?

2. (a) To what extent do student reported ratings of objective (state) motivation predict gains from objective pre- to post-quiz within students? For state motivation, we use student ratings of their interest ("fun"), utility (how much they "learned") and cost (how "difficult" it was).

2. (b) To what extent do student reported ratings of objective (state) motivation *and* mathematics (trait) motivation predict gains from objective pre- to post-quiz between students? For trait motivation, we use student ratings of their expectancies, utility, value, and cost of mathematics.

2. (c) Is there evidence that these predictive relationships differ across grade levels?

With these questions, we can understand whether variation in motivation for specific learning content across time predicts performance on that learning content better than individual student averages of motivation for that learning content and better than individual motivation for the broader domain, herein mathematics.

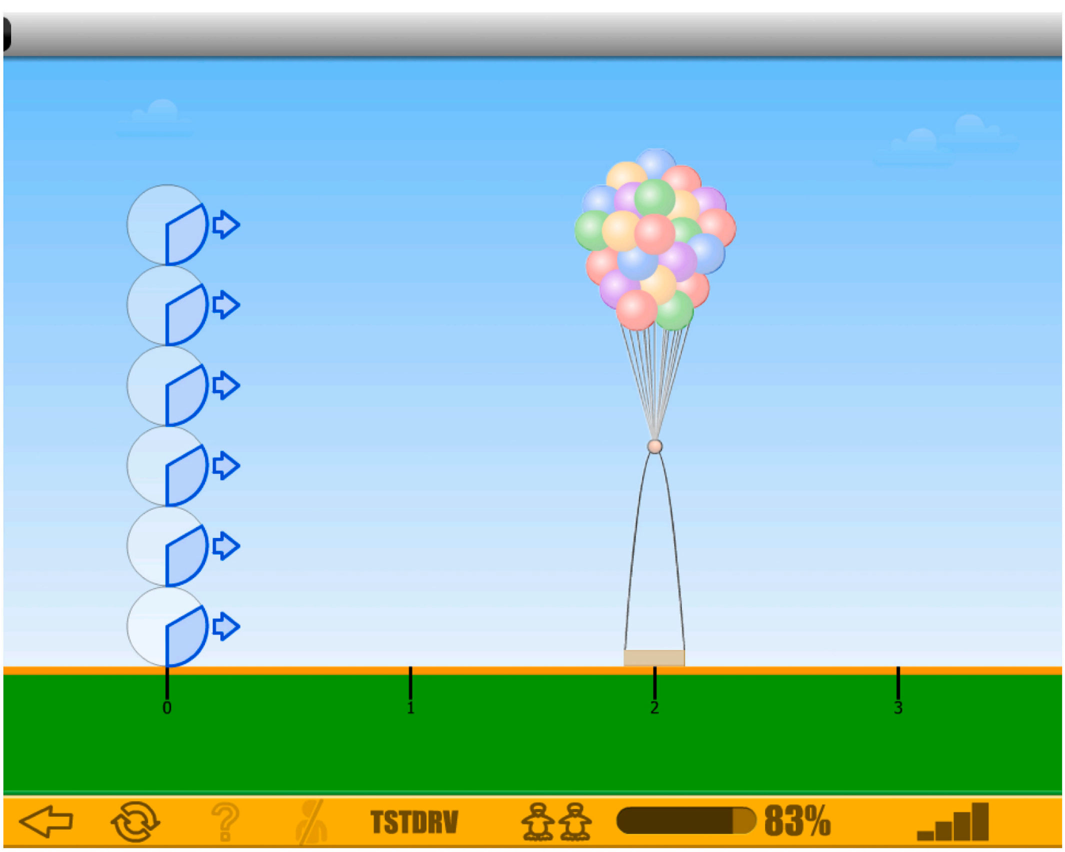

Fig. 1. An example of an ST math puzzle
Note. A puzzle from the objective *Fractions on the Number Line* from the game *Jiji Cycle*. In this puzzle, students must choose the position for the balloon that would match where the circles filled with thirds would unwind if they were spread out. This will allow Jiji the penguin to enter the balloon and exit off the screen.

2. Method

2.1. Participants

Participants were third through fifth grade students (approximately aged eight through 11) within a large Florida school district who played ST Math during the 2017–2018 school year. From district-provided records, 21,884 students were enrolled in the study grades across 100 schools in 2017–2018. Of these, 20,026 were able to be matched with some ST Math data across 96 schools. A number of schools did not fully implement the objective ratings surveys, so only 10,511 students had data on these variables of interest. Finally, 9091 (42 % of the full sample) across 82 schools had the full complement of data, including gameplay data (e.g., objective quiz scores), motivation for mathematics, and motivation ratings of objectives. This final analysis sample was 48 % girls, 4 % Asian, 19 % Black or African American, 19 % Hispanic or Latinx, 53 % White, and 5 % Multi-racial and/or another racial/ethnic category. Nine percent of the students were English Language Learners (ELL), 74 % qualified for federal free or reduced-priced lunch (as a measure of socioeconomic status), and 14 % were classified as having a disability by their school district. Between 30 % and 40 % of students were in each grade, with more in third grade than fourth and fifth. Table 1 compares the district sample to the analysis sample. Although most differences between the two samples were numerically small (within 4 %), the analysis sample was younger, had fewer girls, fewer Asian and White students, more Hispanic/Latinx students, more ELL students, and more students eligible for free/reduced lunch.

2.2. Procedures

All data were passively collected as students engaged with ST Math in their classrooms as part of their normal schoolwork. The structure of the measurement administration is described in Fig. 2. Institutional Review Board approval was secured by the authors before any data analyses were conducted.

2.3. Measures

2.3.1. ST math performance

As students started objectives within ST Math, they took a pre-quiz of five to eight questions covering the content within the objective. After playing the objective, students took a post-quiz with questions using different numbers but mirroring the pre-quiz question-by-question. These two measures were extracted from the log files along with the objective level pass-rate, which was created by dividing the total number of levels passed within the objective by the total number of levels attempted. Students in the analysis sample played an average of 13.37 (SD 8.57) objectives, with a range of one to 35 objectives. Table 2 provides descriptive statistics of these and the other measures below, separately by the analysis sample and the total sample when data are available.

Table 1
Demographics for total vs. analysis sample.

	Full sample	Analysis sample	Difference	
	Percent	Percent	Amount	*p*-value
Grade 3	33.6 %	37.2 %	−3.6 %	<.001
Grade 4	33.2 %	33.6 %	−0.4 %	.311
Grade 5	33.2 %	29.2 %	4.0 %	<.001
Girl	48.9 %	47.9 %	1.0 %	.013
Asian	4.4 %	3.8 %	0.6 %	<.001
Black or African American	19.2 %	19.2 %	0.0 %	.91
Hispanic or Latinx	17.6 %	19.4 %	−1.8 %	<.001
White	53.8 %	52.5 %	1.3 %	.001
Other Race	5.0 %	5.1 %	−0.1 %	.426
English Language Learner	7.7 %	9.0 %	−1.3 %	<.001
Free/Reduced Lunch	70.9 %	73.7 %	−2.8 %	<.001
Reported Disability	13.0 %	13.5 %	−0.5 %	.09
N	21,884	9091		

Note. Chi-squared tests of sample differences compare those included in the analysis sample from those excluded; a more conservative measure than comparing the total full sample to the analysis sample.

2.3.2. Motivation for objective content

As students completed objectives within ST Math, but before they began the objective post-quiz, they were provided with three questions, “How much fun was this objective?”; “How much did you learn?”; and “How difficult was this objective?” Students answered these questions on a zero to five scale representing “Not at all” to “A lot.” Such single-item constructs have been used to represent SEVT and other motivational constructs in prior studies using experience sampling (e.g., Beymer, Robinson, & Schmidt, 2021; Ketonen et al., 2018; Ketonen et al., 2019). Although these surveys were designed by the developers of ST Math without consideration of theory, we align them to the SEVT constructs of interest, utility, and effort cost.

Students in the analysis sample provided these answers on 119,295 objectives total; on average, each student answered questions on 13.12 (SD 18.40) objectives, with a range of one to 34 objectives. These measures are event-contingent (Shiffman et al., 2008), as they are triggered by the particular event (completing the objective). Event-contingent ESM have been used to monitor student enjoyment and other emotions during mathematics tasks (Skwarchuk, 2009), reports of student engagement and self-regulation (Xie et al., 2019), as well as academic self-concept (e.g., Hausen et al., 2022).

2.3.3. Motivation for mathematics

Students in the NSF project districts were given a survey on their motivation for mathematics three times per year as they played ST Math, once within the first weeks of school, immediately after returning from winter break, and within the last two months of school. For this study, students' answers from the first survey at the start of fall 2017 were used. Questions asked students about their expectancy for mathematics (two questions, e.g., “How well do you think you will do on math this year,” alpha .72), their utility and importance value (four questions, e.g., “How useful will math be to you in the future,” alpha .82), and their emotional cost for mathematics (one question with up to three responses). Expectancy and utility/importance value questions were presented using a visual Likert-type scale, wherein students were able to examine changes in faces on a tomato character “tamojis” corresponding to values on the scale (e.g., “Not at all useful”…”Very useful”). The emotional cost question asked students to pick emotionally expressive tamojis in answer to “How does math make you feel?” Tamojis were labeled with and represented seven emotions: bored, challenged, excited, frustrated, happy, hopeful, nervous. A measure of emotional cost for mathematics was created by totaling the number of negative emotions (bored, frustrated, nervous) each student chose, resulting in a variable from zero to three. An expectancy scale was created by mean averaging the two expectancy questions; a positive value scale was created by mean averaging the four utility/importance questions. We use the term “positive value” to reference the combined utility/importance scale; however, we recognize that there may also be times when cost value can serve students positively (see Eccles & Wigfield, 2020). These measures have been previously validated with cognitive interviews (Rutherford, Liu, & Wagemaker, 2021), through relationships with similar constructs (Liu, Rutherford, & Karamarkovich, 2022), and through their inter-relations and relationship with performance (Rutherford, Duck, Rosenberg, & Patt, 2021). For survey images, see Supplemental Materials. See Table 3 for further details on the measures.

2.3.4. Covariates

Student demographics provided by the district were included as covariates in models: grade-level, race, gender, ELL status, free/reduced

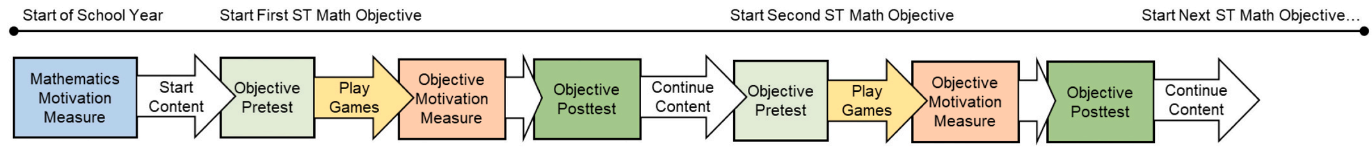

Fig. 2. Illustration of measurement timeline
Note. Measurement instances are shown in colored portions of the figure. Mathematics motivation measurement (blue) occurred once at the start of the academic school year. Before students began each ST Math objective, they took a pretest on the objective content (light green), then played the games where their pass rate per level was recorded (yellow). They then took the objective motivation survey (orange) and the objective posttest (dark green). Only two objectives are shown in the figure, but students in the data for the current study completed up to 34 objectives throughout the year. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Descriptive statistics for total vs. analysis sample.

	All available data			Analysis sample			
Student Level	Mean	SD	Count	Mean	SD	Count	*p* of diff
Mathematics Expectancy	4.185	0.799	19,860	4.188	0.790	9091	.669
Mathematics Value	4.345	0.716	19,860	4.344	0.713	9091	.822
Mathematics Emotional Cost	0.763	0.948	19,860	0.769	0.945	9091	.420
Objective Interest	2.030	1.163	9245	2.025	1.161	9091	<.001
Objective Utility Value	2.176	1.156	9245	2.171	1.154	9091	<.001
Objective Effort Cost	1.485	0.947	9245	1.485	0.943	9091	.045
Objective Pre-Quiz Score	0.624	0.179	19,653	0.637	0.180	9091	<.001
Objective Post-Quiz Score	0.770	0.150	19,646	0.773	0.151	9091	<.001
Objective Level Pass Rate	0.713	0.130	19,654	0.710	0.132	9091	.858
Objective Level	Mean	SD	Count	Mean	SD	Count	*p* of diff
Objective Interest	1.962	1.508	122,184	1.957	1.508	119,295	<.001
Objective Utility Value	2.106	1.486	122,184	2.101	1.486	119,295	<.001
Objective Effort Cost	1.491	1.359	122,184	1.484	1.355	119,295	<.001
Objective Pre-Quiz Score	0.629	0.308	294,676	0.644	0.304	119,295	<.001
Objective Post-Quiz Score	0.789	0.246	291,313	0.792	0.243	119,295	<.001
Objective Level Pass Rate	0.740	0.186	296,375	0.737	0.185	119,295	.068

Note. Not all variables are available for all students in the total sample. Difference reported is from an unpaired *t*-test comparing those included in the analysis sample from those excluded.

Table 3
Measure constructs and example items.

Construct	When measured	Sample item
Objective Interest	After each objective	How much fun was this objective?
Objective Utility		How much did you learn?
Objective Cost		How difficult was this objective?
Mathematics Expectancy	At the beginning of the school year	How good would you be at learning new things in math?
Mathematics Positive Value		How important is math to you now?
Mathematics Emotional Cost		How does math make you feel?

lunch eligibility, and whether the student had a reported disability. In addition, specific objective content and when the student played the content were additional covariates. Play dates were the last day the student played a given objective, which did not perfectly align with objective order (students may go back to replay previously passed content, but our prior work indicates this is rare, Liu, Cody, Barnes, Lynch, & Rutherford, 2017; Zhang & Rutherford, 2022). Student play date was operationalized in models as the number of days since the first entered objective play date in August of 2017. We chose to make this a static date for all students instead of a relative date to capture calendar-specific associations (e.g., holidays).

2.4. Analysis

Little's MCAR test (1988) revealed that the three predictor variables of interest and the post-quiz outcome variable were not missing data completely at random (*ps* < .001). Chi2 tests indicated that teacher predicted missingness for all four variables and that student grade-level and demographics predicted missingness for the three objective motivation outcomes but not objective post-quiz scores. Among other independent variables, although the group failed Little's MCAR test, missing data on these variables were <0.4 % of the sample. Given the large sample size even after removing missing data and given that the largest proportion of students were missing objective content motivation because they did not receive these surveys, we made the decision to run the main analyses with listwise deletion (see Jakobsen, Gluud, Wetterslev, & Winkel, 2017). Because of our use of listwise deletion, our analysis is limited to only those students in the study district who received the objective content surveys and who had other variables in our models.

To examine within student (across objectives) and between student associations, we estimated random-intercepts two-level hierarchical linear models with objective nested within student. To isolate within-student effects and to eliminate bias from unobserved student characteristics (see Allison, 2005; Hamaker & Muthén, 2020), objective-level variables were centered around each student's mean for that variable. For example, for objective pre-quiz, first an average pre-quiz score was calculated for each student. Then, for each objective, this mean was subtracted from the specific objective pre-quiz. Positive values would indicate a pre-quiz higher than the student's typical pre-quiz score; negative values would indicate a pre-quiz lower than typical for that student.

In this way, we could answer questions such as whether the same student made greater pre- to post-quiz gains when they played an objective that they reported was more fun than what they reported as their average level of "fun" for objectives. Models included both the centered variables (to represent within-student associations) and the student mean variables (to represent between-student associations). Within these models, the effect for the individual (Level 2) is the difference between coefficients for the within level and between level—this "compositional" or "contextual" effect is the extent to which the between student effect (Person level) remains once the objective-specific effect within student is controlled (see Allison, 2005; Raudenbush & Bryk, 2002). Allison (2005) claims that the contextual effect coefficient is equal to that for the mean-level variables within a regression where level 1 variables are not group-mean-centered. We confirmed these results by estimating such a model to check whether between-student coefficients were the same as calculated contextual coefficients; in all cases they were. To test for statistical significance of the contextual effect, Wald post-estimation tests were conducted to compare within and between student coefficients for the variables of interest. Differences, if statistically significant, were quantified and expressed as differences in standardized effect sizes. All standardized effect sizes were calculated using the relevant level-specific standard deviation for each variable using the formula: $(B{*}SD_X)/SD_Y$. In text, these are referred to as β.

Within the models at Level 1—the objective level—we included dummy variables for each objective as covariates to represent differences in the specific objective content. The reference objective was the first objective played for each grade level. In third grade, this was "Place Value Concepts," in fourth, this was "Place Value," and in fifth, this was "Whole Numbers." We also included a variable to indicate when each objective was played (number of days since start of play). Unchanging student characteristics (i.e., gender) were entered as covariates at Level 2, along with mathematics motivation and the student objective survey means.

Students in the analysis sample were nested within 839 teachers. Intraclass correlation coefficients (ICCs) for outcome variables across teachers ranged from .04 to .09. In school research, ICCs <.10 are typically considered small (Preacher, Zhang, & Zyphur, 2011). Other studies that have analyzed data structured with students nested within teachers did not include a teacher level with similar ICCs (e.g., Muñoz & Chang, 2007). However, in response to reviewer feedback we have modeled the teacher level in our analyses at Level 3.

Models were built and tested starting with an unconditional model, then moving on to a model with variables of interest and objective controls (content dummies and timing), and finally adding student control variables. At each stage, we determined whether to interpret each model by examining whether changes from model *n*-1 to current model resulted in a better overall fit through statistically significant changes in Deviance statistics and reduction in AIC and BIC.

Because objective content differed across grade-level, we estimated separate models for each grade to allow examination of links between outcomes and specific objective content and to permit associations between variables to vary completely across grade. As noted in our literature review (1.3), we had reason to believe that motivation may operate differently by grade-level across the age range of our sample. We compared coefficients across grade-level models by calculating z scores to represent the difference between each grade pair (e.g., grade 3 vs. grade 5) using the formula [z = (B1-B2)/(sqrt(SE1ˆ2 + SE2ˆ2))]. Z scores greater than the absolute value of 1.96 were considered statistically significant at the $p < .05$ level (see Clogg, Petkova, & Haritou, 1995).

Materials and analysis code for this study are available by emailing the corresponding author. Data can be shared pending district agreement and appropriate IRB.

3. Results

3.1. Zero-order correlations

Table 4 provides correlations between variables at the student level (below diagonal) and among the objective-level variables (above diagonal). Expected correlations emerged between mathematics motivation at the student level, with expectancy and positive value positively correlated and each correlated negatively with emotional cost. Expectancy and positive mathematics value were each also correlated positively with means of student performance variables; cost was correlated negatively with these performance variables. Means of objective motivation were correlated positively with their more similar mathematics counterparts (e.g., objective interest to expectancy and positive value) and negatively correlated with their adverse motivational counterpart (e.g., objective utility to mathematics cost). Interestingly, means of objective content motivation were all positively correlated with each other and negatively correlated with performance variable means, no matter the expected valence (i.e., positive value compared with cost). This same pattern held with the variables at the objective level, except that objective interest was positively correlated with post-quiz score. These correlations present an aggregate of student data. Separate models

Table 4
Correlations among study variables.

		1	2	3	4	5	6	7	8	9
1	Math expectancy	1	–	–	–	–	–	–	–	–
2	Math value	0.546[c]	1	–	–	–	–	–	–	–
3	Math cost	−0.414[c]	−0.390[c]	1	–	–	–	–	–	–
4	Obj Interest	0.150[c]	0.175[c]	−0.149[c]	1	0.508[c]	0.114[c]	−0.019[c]	0.016[c]	−0.062[c]
5	Obj Utility	0.072[c]	0.132[c]	−0.055[c]	0.620[c]	1	0.310[c]	−0.101[c]	−0.064[c]	−0.153[c]
6	Obj Cost	−0.101[c]	−0.0376[c]	0.073[c]	0.188[c]	0.371[c]	1	−0.136[c]	−0.180[c]	−0.186[c]
7	Pre Quiz	0.085[c]	0.074[c]	−0.106[c]	−0.120[c]	−0.197[c]	−0.171[c]	1	0.429[c]	0.337[c]
8	Post Quiz	0.101[c]	0.074[c]	−0.115[c]	−0.125[c]	−0.192[c]	−0.186[c]	0.709[c]	1	0.351[c]
9	Obj Pass Rate	0.108[c]	0.050[c]	−0.129[c]	−0.224[c]	−0.282[c]	−0.162[c]	0.549[c]	0.601[c]	1
10	Girl	−0.059[c]	0.017	0.069[c]	0.030[b]	0.081[c]	0.069[c]	0.086[c]	0.053[c]	−0.018
11	Asian	−0.006	−0.012	−0.022[a]	−0.002	−0.010	0.002	0.050[c]	0.061[c]	0.087[c]
12	Black/African Amer.	0.106[c]	0.094[c]	−0.057[c]	0.166[c]	0.192[c]	0.061[c]	−0.155[c]	−0.185[c]	−0.225[c]
13	Hispanic/Latinx	−0.034[b]	−0.020	−0.022[a]	0.044[c]	0.068[c]	0.024[a]	−0.071[c]	−0.053[c]	−0.062[c]
14	White	−0.055[c]	−0.059[c]	0.073[c]	−0.157[c]	−0.193[c]	−0.056[c]	0.143[c]	0.152[c]	0.178[c]
15	Other Race	0.002	0.012	−0.004	−0.019	−0.019	−0.027[b]	0.037[c]	0.031[b]	0.033[b]
16	ELL	−0.015	−0.028[b]	−0.054[c]	0.121[c]	0.118[c]	0.031[b]	−0.186[c]	−0.170[c]	−0.159[c]
17	Free/Reduced Lunch	0.009	0.032[b]	−0.014	0.167[c]	0.200[c]	0.079[c]	−0.207[c]	−0.236[c]	−0.241[c]
18	Disability	−0.016	−0.057[c]	0.006	0.113[c]	0.102[c]	0.060[c]	−0.226[c]	−0.242[c]	−0.246[c]
19	Grade 3	0.104[c]	0.033[b]	−0.077[c]	0.194[c]	0.101[c]	−0.015	0.130[c]	0.095[c]	−0.070[c]
20	Grade 4	−0.0202	0.013	0.024[a]	−0.032[b]	−0.028[b]	−0.014	−0.078[c]	−0.071[c]	−0.066[c]
21	Grade 5	−0.090[c]	−0.049[c]	0.057[c]	−0.173[c]	−0.078[c]	0.030[b]	−0.058[c]	−0.027[b]	0.143[c]

Note. Values below the diagonal are from student-level (between) correlations of 9091 students. Values above the diagonal are correlations of objective-level (within) measures for 119,295 objectives.
[a] $p < .05$, [b] $p < .01$, [c] $p < .001$.

by grade show similar patterns, except that the positive correlation between objective interest and post-quiz score was only seen for third graders.

Although the majority of the correlations were statistically significant, many correlations across constructs (i.e., from motivation to performance) were small, with many under .1.

3.2. Research question 1

For each of the objective content motivation variables, roughly between 46 % and 67 % of the variance for objective motivation ratings (interest, utility value, effort cost) were within students. These numbers were reduced by grade-level, with older students showing more consistent ratings than younger students. Across grades, student difficulty ratings (effort cost) consistently displayed greater variance within students than did objective interest and utility value. Individual students displayed more variance across objective content in their performance than they did in their motivation, with between 70 % and 81 % of the variance in post-quiz score within students. These performance measures, however, did not show any consistent grade-level patterns. Table 5 displays the variance components for each variable and grade-level, divided by between teacher variance, between student variance, and within student variance.

3.3. Research question 2

For our second research question, we examined post-quiz score for each objective as predicted by both objective motivation and performance, as well as broader mathematics motivation (Table 6 and Fig. 3). Across grades, the addition of the variables of interest explained between 83 % and 92 % of the variance at the between-teacher level, between 74 % and 81 % of the variance at the between-student level, and between 27 % and 36 % of the variance at the within-student level. These models all resulted in statistically significant changes in the deviance statistic from the null model and improved model fit based on AIC and BIC. Adding demographic controls accounted for an additional 13 % to 43 % of the variance at the between-teacher level, 1 % to 2 % of the variance at the between-student level and trace amounts (0.05 % or less) of variance at the within-student level, but still resulted in statistically significant improvements to model fit.

Below, we separately discuss results for questions 2(a), within-student associations, and 2(b), between-student associations. We discuss results related to question 2(c), regarding grade-level differences, within each section.

3.3.1. Within-student associations

Within students, the objective performance variables had positive associations with post-quiz performance. When students performed better on objective content relative to their average performance, they also performed better on the objective post-quiz (betas 0.177 to 0.197, *ps* $< .001$, no statistically significant differences between grades). When students performed better on the objective pre-quiz relative to their average pre-quiz performance, they also performed better on the objective post-quiz (betas 0.151 to 0.175, *ps* $< .001$, third and fourth grade betas statistically significantly different from those in fifth grade, *ps* $< .001$). Objective motivation also had statistically significant associations (all *ps* $< .001$) with objective post-quiz score. Student-reported objective interest and utility were positive associated with post-quiz performance. When students reported greater interest in an objective relative to their average interest, they scored higher on that post-quiz (betas 0.078 to 0.100, no statistically significant differences between grades). Similarly, when they reported an objective was more useful for their learning, they also performed better on its post-quiz (betas 0.051 to 0.071, differences between fifth graders and both third ($p = .003$) and fourth ($p = .015$) graders). Cost had the opposite relation with post-quiz performance within students: when students noted an objective was more difficult for them than typical, they also performed worse on the post-quiz (betas −0.125 to −0.128, differences across grade-levels only statistically significant when comparing third and fifth graders, $p = .028$).

3.3.2. Between-student associations

At the between-student level, the contextual effect (coefficient for student mean minus the coefficient for the centered variables, see Allison, 2005; Hamaker & Muthén, 2020) demonstrated that student-level-average performance metrics were also positive predictors of objective post-quiz performance. Students who performed better on the objective content, on average, scored higher (betas 0.111 to 0.191, statistically significant differences between third graders and fourth and fifth graders, *ps* $< .001$). Students who performed better, on average, on the pre-quizzes also performed better on the post-quizzes (betas 0.159 to 0.214, coefficients statistically significantly different between all grades, *ps* $< .033$). Contextual coefficients for objective motivation showed opposite relations with post-quiz performance than did those for within-student objective motivation and were smaller than their corresponding within-student variable beta. Students who had more positive objective motivation, on average, performed worse. Betas for the contextual association of objective interest and post-quiz performance ranged from −0.068 to −0.074. Those for the contextual association of objective utility and post-quiz performance ranged from −0.023 to 0.035. When students reported that the content was more difficult, on average, they performed better than peers who reported content was less difficult (betas 0.067 to 0.082). No grade-level differences in associations between objective motivation and post-quiz performance were statistically significant.

Mathematics motivation had smaller associations with post-quiz score than did both centered and mean-level objective motivation and performance. Neither mathematics expectancy nor value had statistically significant associations with post-quiz score (*ps* across grade levels $>.311$). Mathematics emotional cost was only statistically significantly associated with post-quiz performance among fifth graders. For these students, a one standard deviation increase in higher reported emotional cost for mathematics was associated with a 0.041 decrease in post-quiz performance ($p < .001$). Finally, the timing of play for each objective mattered little toward predicting objective post-quiz performance. Betas for the number of days since the start of play were close to zero, with all *p*-values above .213. This indicates that students did not get better or worse on ST Math post-quizzes during the academic year when specific content of the material was controlled. Below, we contextualize the results linking motivation and performance, both within and between

Table 5
Intraclass (student-level) correlations for objective performance and motivation.

	3rd			4th			5th		
	bwT	bwSt	wiSt	bwT	bwSt	wiSt	bwT	bwSt	wiSt
Objective Interest	3.62 %	41.63 %	54.75 %	3.02 %	45.95 %	51.03 %	4.89 %	45.42 %	49.69 %
Objective Utility	3.56 %	42.69 %	53.76 %	4.16 %	48.57 %	47.28 %	4.66 %	49.13 %	46.21 %
Objective Cost	1.25 %	32.22 %	66.53 %	1.68 %	37.28 %	61.05 %	0.36 %	40.72 %	58.92 %
Post Quiz Score	4.70 %	14.64 %	80.67 %	5.89 %	23.75 %	70.35 %	6.44 %	19.41 %	74.15 %

Note. bwT is the percent of variance that is between teachers, bwSt is between students, wiSt is within student (across objectives).

Table 6
Objective post-quiz performance predicted by objective motivation.

	Third Grade				Fourth Grade				Fifth Grade			
	B	SE	p-value	Beta	B	SE	p-value	Beta	B	SE	p-value	Beta
Centered Obj Interest	0.015	0.001	<.001	0.100	0.015	0.001	<.001	0.092	0.014	0.001	<.001	0.078
Centered Obj Utility	0.008	0.001	<.001	0.051	0.009	0.001	.000	0.053	0.012	0.001	<.001	0.071
Centered Obj Cost	−0.021	0.001	<.001	−0.128	−0.023	0.001	.000	−0.125	−0.025	0.001	<.001	−0.127
Centered Pass Rate	0.142	0.008	<.001	0.177	0.143	0.008	.000	0.179	0.158	0.009	<.001	0.197
Centered PreQuiz	0.141	0.004	<.001	0.175	0.140	0.004	.000	0.175	0.121	0.004	<.001	0.151
Contextual Obj Interest	−0.014	0.002	<.001	−0.068	−0.016	0.002	<.001	−0.074	−0.017	0.002	<.001	−0.072
Contextual Obj Utility	−0.007	0.002	<.001	−0.033	−0.005	0.002	.027	−0.023	−0.008	0.003	.002	−0.035
Contextual Obj Cost	0.017	0.002	<.001	0.067	0.021	0.002	<.001	0.082	0.020	0.002	<.001	0.074
Contextual Pass Rate	0.150	0.017	<.001	0.111	0.242	0.017	<.001	0.177	0.260	0.021	<.001	0.191
Contextual PreQuiz	0.298	0.011	<.001	0.177	0.332	0.011	<.001	0.214	0.254	0.012	<.001	0.159
Girl	−0.002	0.003	.452	−0.008	0.000	0.003	.880	−0.002	0.005	0.003	.131	0.020
Hispanic/Latinx	0.007	0.004	.109	0.027	0.007	0.004	.098	0.029	−0.001	0.005	.807	−0.004
Asian	0.006	0.007	.382	0.027	0.008	0.007	.228	0.033	0.004	0.008	.670	0.014
Black/African American	0.000	0.004	.961	−0.001	−0.017	0.004	.000	−0.072	−0.020	0.005	<.001	−0.079
Other or Multiple Races	−0.001	0.006	.925	−0.002	−0.008	0.006	.234	−0.032	−0.011	0.007	.145	−0.043
ELL	−0.010	0.005	.059	−0.042	−0.027	0.006	<.001	−0.114	−0.017	0.007	.013	−0.068
Free/Reduced Lunch	−0.013	0.003	<.001	−0.053	−0.013	0.003	<.001	−0.054	−0.011	0.004	.010	−0.043
Reported Disability	−0.021	0.004	<.001	−0.085	−0.008	0.005	.094	−0.031	−0.021	0.006	<.001	−0.085
Math Expectancy	< 0.001	0.002	.957	< 0.001	0.002	0.002	.311	0.007	0.001	0.003	.773	0.003
Math Positive Value	< 0.001	0.002	.989	< 0.001	0.001	0.002	.573	0.004	0.001	0.003	.743	0.003
Math Emotional Cost	0.001	0.002	.519	0.004	0.001	0.002	.438	0.005	−0.010	0.002	<.001	−0.041
Days Since Start	< 0.001	< 0.001	.213	0.007	< 0.001	< 0.001	.213	< 0.001	< 0.001	< 0.001	.932	< 0.001
Constant	0.358	0.016	<.001		0.200	0.016	<.001		0.200	0.020	<.001	
N Teacher	324				273				242			
N Student	3381				3056				2654			
N Objectives	40,471				45,056				33,768			
	Est	SE	95 % Conf.		Est	SE	95 % Conf.		Est	SE	95 % Conf.	
Between Teachers	< 0.001	< 0.001	<0.001	0.001	< 0.001	< 0.001	<0.001	< 0.001	0.001	< 0.001	<0.001	0.001
Between Students	0.002	< 0.001	0.002	0.002	0.003	< 0.001	0.002	0.003	0.003	< 0.001	0.003	0.003
Within Students	0.035	< 0.001	0.034	0.035	0.035	< 0.001	0.034	0.035	0.036	0.000	0.035	0.036

Note. Coefficients for centered variables show within student associations; Coefficients for contextual variables show the between student associations (calculated from coefficients for means of variables minus coefficients for centered variables). Betas calculated using the formula (B*sdx)/sdy using the level-specific (student or means) standard deviations.
Reference group for regressions are boys, who in district records are identified as White, not identified as ELL, not eligible for free/reduced lunch, and without a reported disability

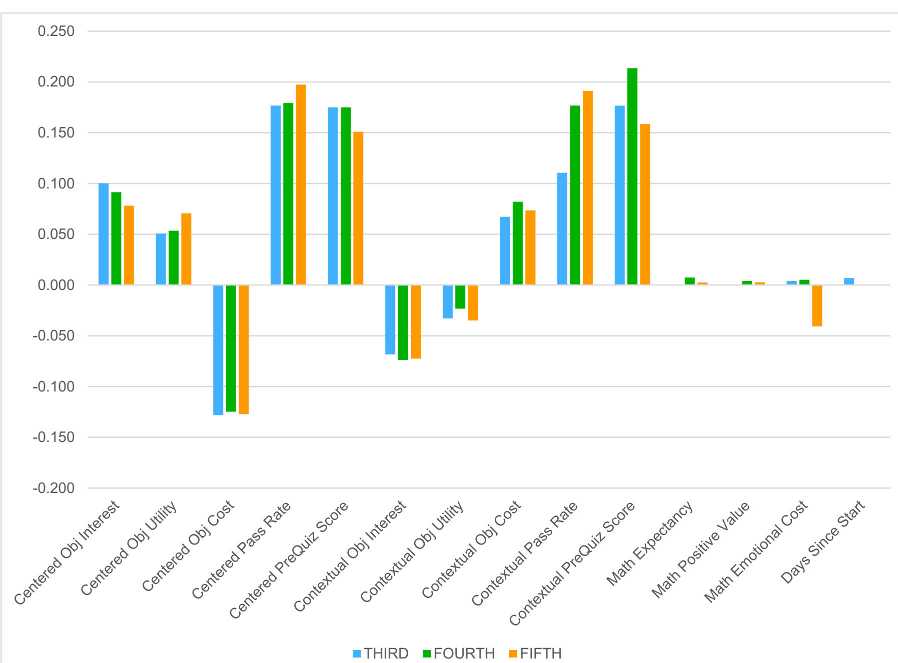

Fig. 3. Association between variables and objective post-quiz score
Note. Y-axis displays standardized betas. Coefficients for centered variables show within student associations; Coefficients for contextual variables show the between student associations (calculated from coefficients for means of variables minus coefficients for centered variables). Betas calculated using the formula (B*sdx)/sdy using the level-specific (student or means) standard deviations.

students, to the broader motivation literature.

4. Discussion

Within this study, we examined how motivation (interest, utility, and cost) varied across different objective content within a mathematics learning software, ST Math, and how state-level motivation related to variations in performance, both within and between students. We found that roughly half of the variance in motivation for mathematics objective content within ST Math was within students, and that this number decreased by grade-level for each of our three measures. The between-student variance was lower than that found with Martin and colleague's (2020) study of values with junior high schoolers. Although our study is not longitudinal across grades, the results are suggestive of the proposition that younger students may experience greater momentary variability in motivation, in line with non-ESM studies of motivation and related constructs (e.g., Benson et al., 2019; Gottfried et al., 2001).

4.1. Association between state motivation and performance

Prior research has consistently demonstrated that motivation for a subject, such as mathematics, predicts performance (Lee & Kung, 2018; Marsh et al., 2018). Within students in our study, both objective interest and utility positively predicted objective math post-quiz, even after controlling for both pre-quiz and objective performance. When students were more interested in objective content and reported they learned more in an objective, they performed better on that objective relative to objectives in which they were less interested and felt they learned less. Beta coefficients for interest were consistently higher than those for utility, and in the case of third graders, were nearly two times as high. Much of the prior research examining the link between task values and performance has used combined measures of value (e.g., Jiang, Rosenzweig, & Gaspard, 2018; Liem, Lau, & Nie, 2008), not allowing for comparisons between value components. Other research has shown both stronger links between utility and performance (e.g., Guo, Marsh, Parker, Morin, & Yeung, 2015) and stronger links between interest and performance (e.g., Chung & Kim, 2022). Peixoto et al. (2022) is the most closely-related study to our own in that they investigated mathematics motivation and performance in a similar age group; they also found stronger associations between interest and performance than between utility and performance. As for mechanisms driving this comparative strength—given that our measures of motivation and performance were tightly coupled in time and content, the link between interest and short-term engagement (see Renninger & Bachrach) may have been determinative.

Within students, perceived effort cost was negatively associated with performance, which may indicate that students accurately perceived more difficult objectives as requiring more effort. However the direction of this association was different at the between-students level. Those students who, on average, perceived ST Math as more difficult were more likely to perform well, on average. These same reversals in coefficients were also seen for interest and utility, which had negative associations with performance at the between-student level, indicating that students who found the games more interesting or useful, on average, had average performance worse than those students who found the games less interesting or useful. Although this reversal of coefficients is possible in these types of models (e.g., Heatly, Bachman, & Votruba-Drzal, 2015), we nevertheless ran robustness checks examining purely within-student fixed effects models (see McNeish & Kelley, 2019); within-student coefficients for our variables of interest were within 0.01 of our current models, with most differences smaller. We hesitate to interpret the contextual objective motivation coefficients beyond noting that they are consistently smaller than their within-student counterparts. Allison (2017) notes that these coefficients are less meaningful in longitudinal within-between models than in one with other nesting structures. Although our main focus is within-student relationships, we chose to present the hybrid within/between models because of their ability to also provide estimates for purely between-student variables (e.g., mathematics expectancy).

4.2. Association between mathematics motivation and performance

Examining these between-student motivation variables, neither mathematics expectancy nor positive value for mathematics (utility, attainment) were statistically significant predictors of post-quiz score. Both variables were positively correlated with objective motivation variables and have conceptual overlap with these variables as well—value for mathematics includes utility questions, and expectancy is likely most related to perceptions of difficulty (Watt, 2004). The literature is limited with respect to studies that include broader measures of motivation together with ESM motivation measures (cf Martin et al., 2020), and none include both sets of variables in a model predicting performance. Our results suggest that the more immediately relevant links between same-time motivation and performance subsume broader measures of same or similar-construct motivation. Emotional cost for mathematics presented a more complicated picture, as it was a statistically significant negative predictor of performance, but only for fifth graders. There are few studies examining the association of emotional cost and performance among elementary-aged students; however, studies of anxiety (one aspect of our emotional cost measure) may be instructive. Prior studies have found links between mathematics anxiety and performance in even younger children (e.g., Gunderson, Park, Maloney, Beilock, & Levine, 2018; Wu, Barth, Amin, Malcarne, & Menon, 2012), but studies including other motivation variables in models have found that these links are indirect (e.g., Meece, Wigfield, & Eccles, 1990). Prior developmental work has found no grade-level differences in these associations (e.g., Sorvo et al., 2019). Without further research, we cannot know if our fifth grade result is an outlier.

4.3. Developmental changes associating objective motivation and performance

Examining grade-level changes in predicting post-quiz performance from objective content motivation, within students across grade levels, coefficients for objective content interest declined with older grades, but these differences were not statistically significant. Coefficients for utility, in contrast, increased with older grades, but differences were only statistically significant in comparing fifth graders to the other two grades. Early adolescence, such as during fifth grade, is when students begin to form more stable pictures of their identity (see Meeus, Van de Schoot, Keijsers, Schwartz, & Branje, 2010)—these may influence the relation between utility value and performance. Early conceptualizations of SEVT noted that utility and attainment may become especially important as students enter middle school (Wigfield, 1994). Turning to effort cost, although the associations between earlier performance and effort cost were stronger in older grades, there was no clear grade-related pattern between objective effort cost and post-quiz performance.

4.4. Limitations and future directions

Within this study, we leveraged secondary data from within a learning technology, ST Math, to investigate how motivation dynamically relates to performance across mathematics objectives within a year-long curriculum. Although this approach presented many opportunities to study a large and diverse population of elementary students—a population not typically represented in ESM research—relying on secondary data also presented limitations. We were unable to specify and develop our measures in ways more typical of traditional researcher-generated studies, and this may present some construct validity issues. We have presented arguments and prior literature justification for why student reports of "fun," "learning," and "difficulty" align with task interest, utility, and effort cost; however, had we been developing these instruments anew, we likely would have made different choices in their construction. The same can be said for our broader measures of mathematics motivation; however, the construct validity of these measures has been established through a cognitive interview process (Rutherford, Liu, & Wagemaker, 2021). All in all, we assert that the balance between specificity and access falls on the positive side of *pragmatic measurement* (see Kosovich, Hulleman, & Barron, 2017). Construct validity issues are important in linking results to theory, and although our work can inform theory, it might especially have value in demonstrating that variance in reports of meaningful student perceptions are related to variance in performance over and above immediately-prior performance and broader measures of student perception.

Educational games have previously been found to be motivating for students (e.g., Fadda, Pellegrini, Vivanet, & Zandonella Callegher, 2022), as have learning technologies more generally (e.g., Higgins, Huscroft-D'Angelo, & Crawford, 2019). ST Math itself has been found to improve student self-beliefs about mathematics (Rutherford et al., 2019). Therefore, the results found within this study may be limited to or especially relevant for learning technology or other motivating learning environments.

The study of dynamic motivation using ESM is quickly expanding; however, it has yet to reach a saturation point across subjects, grade-levels, and conceptualizations of both motivation and performance. Our research expands this work to provide insights into how motivation relates to performance, both between and within elementary students in a mathematics technology context. Diversity in samples, outcomes, and construct conceptualization add strength to theorized links between constructs (Shadish, Cook, & Campbell, 2002); future work bracketing our operationalizations can lead to more firmly grounded conclusions and recommendations.

5. Conclusion

Within this study we investigated how motivation (task interest, utility, and effort cost) for specific mathematics objective content varied across content and grade level and how that objective content motivation predicted objective post-quiz performance after engaging with the content. We found that when students were more interested in an objective and found it more useful for their learning, they were more likely to perform well at post-quiz. The latter association strengthened with age. Ratings of difficulty for content, which we conceptualized as objective effort cost, had negative associations with performance. Each measure of objective motivation was correlated with more trait-level mathematics motivation. With one exception, this trait-level motivation did not predict performance when considered in a model with state-level motivation.

Results contribute to expanding understanding of the dynamic relationship between motivation and performance, especially at an age-level not typically studied with momentary measures. Our use of classroom technology measures of performance can provide more information about how motivation can inform student actions and performance within day-to-day learning activities. By understanding how motivation and performance relate at this level, educators and content developers can make adjustments to their materials and instruction to maximize each. Our results suggest that enhancing student interest in the content and reducing their perceptions of difficulty may both be especially fruitful avenues toward improving performance.

Credit author statement

Teomara Rutherford: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Resources, Data Curation, Writing (Original), Writing (Review & Editing), Visualization, Supervision, Project Administration, Funding Acquisition.

Hye Rin Lee: Methodology, Writing (Original).

Kerry Duck: Methodology, Software, Data Curation, Writing (Original).

Declaration of competing interest

The authors have no conflicts of interest to declare.

Acknowledgements

This work was supported by National Science Foundation [grant numbers 1544273 and 1845584]. The opinions expressed are those of the authors and do not represent views of the National Science Foundation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.lindif.2023.102346.

References

Allison, P. D. (2005). *Fixed effects regression methods for longitudinal data using SAS*. Cary, NC: SAS Institute, Inc.
Aunola, K., Leskinen, E., & Nurmi, J. E. (2006). Developmental dynamics between mathematical performance, task motivation, and teachers' goals during the transition to primary school. *The British Journal of Educational Psychology, 76*(1), 21–40. https://doi.org/10.1348/000709905X51608
Allison, P. D. (2017, July 26). Using "between-within" models to estimate contextual effects. *Statistical Horizons.* https://statisticalhorizons.com/between-within-contextual-effects/.
Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191
Barbot, B. (2022). Intra-individual variability in creativity: Nature, measurement, and prospects. *European Psychologist, 27*(3), 154–164. https://doi.org/10.1027/1016-9040/a000470
Becker, E. S., Goetz, T., Morger, V., & Ranellucci, J. (2014). The importance of teachers' emotions and instructional behavior for their students' emotions–An experience sampling analysis. *Teaching and Teacher Education, 43*, 15–26.
Benson, L., English, T., Conroy, D. E., Pincus, A. L., Gerstorf, D., & Ram, N. (2019). Age differences in emotion regulation strategy use, variability, and flexibility: An experience sampling approach. *Developmental Psychology, 55*(9), 1951.
Beymer, P. N., Benden, D. K., & Sachisthal, M. S. (2022). Exploring the dynamics of situated expectancy-value theory: A panel network analysis. *Learning and Individual Differences, 100*, Article 102233.
Beymer, P. N., Robinson, K. A., & Schmidt, J. A. (2021). Classroom activities as predictors of control, value, and state emotions in science. *The Journal of Educational Research, 114*(6), 550–561.
Beymer, P. N., Rosenberg, J. M., & Schmidt, J. A. (2020). Does choice matter or is it all about interest? An investigation using an experience sampling approach in high school science classrooms. *Learning and Individual Differences, 78*, Article 101812.
Beymer, P. N., Rosenberg, J. M., Schmidt, J. A., & Naftzger, N. J. (2018). Examining relationships among choice, affect, and engagement in summer STEM programs. *Journal of Youth and Adolescence, 47*, 1178–1191.
Bowers, A. J. (2011). What's in a grade? The multidimensional nature of what teacher-assigned grades assess in high school. *Educational Research and Evaluation, 17*(3), 141–159.
Capelle, J. D., Grunschel, C., Bachmann, O., Knappe, M., & Fries, S. (2022). Multiple action options in the context of time: When exams approach, students study more and experience fewer motivational conflicts. *Motivation and Emotion, 46*(1), 16–37.
Chung, Y., & Kim, S. (2022). Expectancies, task values, and perceived costs: Reciprocal effects on math-related career intention and achievement among middle school girls. *Learning and Individual Differences, 96*, Article 102145.
Clogg, C. C., Petkova, E., & Haritou, A. (1995). Statistical methods for comparing regression coefficients between models. *American Journal of Sociology, 100*(5), 1261–1293.
Csikszentmihalyi, M., & Larson, R. (2014). Validity and reliability of the experience-sampling method. In M. Csikszentmihalyi (Ed.), *Flow and the foundations of positive psychology* (pp. 35–54). Dordrecht: Springer.
Deci, E. L., & Ryan, R. M. (2012). *Self-determination theory*.
Dicke, T., Marsh, H. W., Parker, P. D., Pekrun, R., Guo, J., & Televantou, I. (2018). Effects of school-average achievement on individual self-concept and achievement: Unmasking phantom effects masquerading as true compositional effects. *Journal of Educational Psychology, 110*(8), 1112–1126. https://doi.org/10.1037/edu0000259
Dietrich, J., Moeller, J., Guo, J., Viljaranta, J., & Kracke, B. (2019). In-the-moment profiles of expectancies, task values, and costs. *Frontiers in Psychology, 10*, 1662.
Dietrich, J., Schmiedek, F., & Moeller, J. (2022). Academic motivation and emotions are experienced in learning situations, so let's study them. Introduction to the special issue. *Learning and Instruction, 101623*.
Eccles, J. S. (2022). Commentary on within-person designs and motivational science. *Learning and Instruction, 81*, Article 101662.

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.), *Achievement and achievement motives: Psychological and sociological approaches* (pp. 75–146). San Francisco, CA: W. H. Freeman.

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*, Article 101859.

Erikson, E. H. (1968). *Identity: Youth and crisis*. Norton.

Fadda, D., Pellegrini, M., Vivanet, G., & Zandonella Callegher, C. (2022). Effects of digital games on student motivation in mathematics: A meta-analysis in K-12. *Journal of Computer Assisted Learning, 38*(1), 304–325.

Fiske, D. W., & Rice, L. (1955). Intra-individual response variability. *Psychological Bulletin, 52*(3), 217–250. https://doi.org/10.1037/h0045276

Flake, J. K., Barron, K. E., Hulleman, C., McCoach, B. D., & Welsh, M. E. (2015). Measuring cost: The forgotten component of expectancy-value theory. *Contemporary Educational Psychology, 41*, 232–244.

Fullagar, C. J., & Kelloway, E. K. (2009). Flow at work: An experience sampling approach. *Journal of Occupational and Organizational Psychology, 82*(3), 595–615.

Gabriel, A. S., Podsakoff, N. P., Beal, D. J., Scott, B. A., Sonnentag, S., Trougakos, J. P., & Butts, M. M. (2019). Experience sampling methods: A discussion of critical trends and considerations for scholarly advancement. *Organizational Research Methods, 22*(4), 969–1006.

Garon-Carrier, G., Boivin, M., Guay, F., Kovas, Y., Dionne, G., Lemelin, J. P., … Tremblay, R. E. (2015). Intrinsic motivation and achievement in mathematics in elementary school: A longitudinal investigation of their association. *Child Development, 87*(1), 165–175. https://doi.org/10.1111/cdev.12458

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming, 33*(4), 441–467.

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.

Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. New York: Palgrave/Macmillan.

Goetz, T., Frenzel, A. C., Stoeger, H., & Hall, N. C. (2010). Antecedents of everyday positive emotions: An experience sampling analysis. *Motivation and Emotion, 34*(1), 49–62.

Goetz, T., Sticca, F., Pekrun, R., Murayama, K., & Elliot, A. J. (2016). Intraindividual relations between achievement goals and discrete achievement emotions: An experience sampling approach. *Learning and Instruction, 41*, 115–125.

Gottfried, A. E., Fleming, J. S., & Gottfried, A. W. (2001). Continuity of academic intrinsic motivation from childhood through late adolescence: A longitudinal study. *Journal of Educational Psychology, 93*(1), 3.

Graziano, A. B., Peterson, M., & Shaw, G. L. (1999). Enhanced learning of proportional math through music training and spatial-temporal training. *Neurological Research, 21*(2), 139–152.

Gunderson, E. A., Park, D., Maloney, E. A., Beilock, S. L., & Levine, S. C. (2018). Reciprocal relations among motivational frameworks, math anxiety, and math achievement in early elementary school. *Journal of Cognition and Development, 19*(1), 21–46.

Guo, J., Marsh, H. W., Parker, P. D., Morin, A. J., & Yeung, A. S. (2015). Expectancy-value in mathematics, gender and socioeconomic background as predictors of achievement and aspirations: A multi-cohort study. *Learning and Individual Differences, 37*, 161–168.

Hamaker, E. L., & Muthén, B. (2020). The fixed versus random effects debate and how it relates to centering in multilevel modeling. *Psychological Methods, 25*(3), 365. https://doi.org/10.1037/met0000239

Hausen, J. E., Möller, J., Greiff, S., & Niepel, C. (2022). Students' personality and state academic self-concept: Predicting differences in mean level and within-person variability in everyday school life. *Journal of Educational Psychology, 114*(6), 1394–1411.

Heatly, M. C., Bachman, H. J., & Votruba-Drzal, E. (2015). Developmental patterns in the associations between instructional practices and children's math trajectories in elementary school. *Journal of Applied Developmental Psychology, 41*, 46–59.

Hektner, J. M., & Csikszentmihalyi, M. (1996, April). *A longitudinal exploration of flow and intrinsic motivation in adolescents*. Paper presented at the annual meeting of the. New York City: American Educational Research Association.

Higgins, K., Huscroft-D'Angelo, J., & Crawford, L. (2019). Effects of technology in mathematics on achievement, motivation, and attitude: A meta-analysis. *Journal of Educational Computing Research, 57*(2), 283–319.

Hong, S., Yoo, S. K., You, S., & Wu, C. C. (2010). The reciprocal relationship between parental involvement and mathematics achievement: Autoregressive cross-lagged modeling. *The Journal of Experimental Education, 78*(4), 419–439. https://doi.org/10.1080/00220970903292926

Jakobsen, J. C., Gluud, C., Wetterslev, J., & Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials—A practical guide with flowcharts. *BMC Medical Research Methodology, 17*(1), 1–10.

Jiang, Y., Rosenzweig, E. Q., & Gaspard, H. (2018). An expectancy-value-cost approach in predicting adolescent students' academic motivation and achievement. *Contemporary Educational Psychology, 54*, 139–152.

Kaplan, A. (2014). Theory and research on teachers' motivation: Mapping an emerging conceptual terrain. In P. W. Richardson, S. Karabenick, & H. M. G. Watt (Eds.), *Teacher motivation: Theory and practice* (pp. 52–66). Routledge.

Kaplan, A. (2015, April). Motivation and complexity: The merits of a dynamic systems approach to theory and research on motivation in education. In *Symposium at the annual convention of the American Educational Research Association, Chicago, IL*.

Kaplan, A., & Patrick, H. (2016). Learning environments and motivation. In *Handbook of motivation at school* (pp. 251–274). Routledge.

Ketonen, E. E., Dietrich, J., Moeller, J., Salmela-Aro, K., & Lonka, K. (2018). The role of daily autonomous and controlled educational goals in students' academic emotion states: An experience sampling method approach. *Learning and Instruction, 53*, 10–20.

Ketonen, E. E., Malmberg, L. E., Salmela-Aro, K., Muukkonen, H., Tuominen, H., & Lonka, K. (2019). The role of study engagement in university students' daily experiences: A multilevel test of moderation. *Learning and Individual Differences, 69*, 196–205.

Kibrick, M. C. (2013). *Cognitive hierarchies of mathematics learning: An analysis of the ST math curriculum*. Irvine: University of California.

Kosovich, J. J., Hulleman, C. S., & Barron, K. E. (2017). Measuring motivation in educational settings: A case for pragmatic measurement. *The Cambridge handbook on motivation and learning*, 39–60.

Krumm, A., Coulson, A., & Neisler, J. (2022, March). Defining productive struggle in ST math: Implications for developing indicators of learning behaviors and strategies in digital learning environments. *In LAK22: 12th international learning analytics and knowledge conference* (pp. 514-520).

Lee, C. Y., & Kung, H. Y. (2018). Math self-concept and mathematics achievement: Examining gender variation and reciprocal relations among junior high school students in Taiwan. *EURASIA journal of mathematics, science and technology. Education, 14*(4). https://doi.org/10.29333/ejmste/82535

Lee, M., & Bong, M. (2022). Patterns of motivation and learning during test-preparation of Korean adolescents: An experience sampling approach. *Educational Psychology, 42*(2), 200–221.

Liem, A. D., Lau, S., & Nie, Y. (2008). The role of self-efficacy, task value, and achievement goals in predicting learning strategies, task disengagement, peer relationship, and achievement outcome. *Contemporary Educational Psychology, 33*(4), 486–512.

Liu, A. S., Rutherford, T., & Karamarkovich, S. M. (2022). Numeracy, cognitive, and motivational predictors of one-year growth in elementary mathematics achievement. *Journal of Educational Psychology, 114*(7), 1589–1607. https://doi.org/10.1037/edu0000772

Liu, Z., Cody, C., Barnes, T., Lynch, C., & Rutherford, T. (2017). The antecedents of and associations with elective replay in an educational game: Is replay worth it? In X. Hu, T. Barnes, A. Hershkovitz, & L. Paquette (Eds.), *Proceedings of the 10th International Conference on Educational Data Mining* (pp. 40–47). Wuhan, China.

Marsh, H. W. (1990). The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology, 82*(4), 623–636.

Marsh, H. W., Pekrun, R., Lichtenfeld, S., Guo, J., Arens, A. K., & Murayama, K. (2016). Breaking the double-edged sword of effort/trying hard: Developmental equilibrium and longitudinal relations among effort, achievement, and academic self-concept. *Developmental Psychology, 52*(8), 1273–1290. https://doi.org/10.1037/dev0000146

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

Marsh, H. W., & Seaton, M. (2013). Academic self-concept. In *International guide to student achievement* (pp. 62–63). Routledge.

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development, 76*(2), 397–416. https://doi.org/10.1111/j.1467-8624.2005.00853.x

Martin, A. J., Mansour, M., & Malmberg, L. E. (2020). What factors influence students' real-time motivation and engagement? An experience sampling study of high school students using mobile technology. *Educational Psychology, 40*(9), 1113–1135. https://doi.org/10.1080/01443410.2018.1545997

McClelland, M. M., John Geldhof, G., Cameron, C. E., & Wanless, S. B. (2015). Development and self-regulation. *Handbook of Child Psychology and Developmental Science, 1-43*. https://doi.org/10.1002/9781118963418.childpsy114

McNeish, D., & Kelley, K. (2019). Fixed effects models versus mixed effects models for clustered data: Reviewing the approaches, disentangling the differences, and making recommendations. *Psychological Methods, 24*(1), 20.

Meece, J. L., Wigfield, A., & Eccles, J. S. (1990). Predictors of math anxiety and its influence on young adolescents' course enrollment intentions and performance in mathematics. *Journal of Educational Psychology, 82*(1), 60.

Meeus, W., Van de Schoot, R., Keijsers, L., Schwartz, S. J., & Branje, S. (2010). On the progression and stability of adolescent identity formation: A five-wave longitudinal study in early-to-middle and middle-to-late adolescence. *Child Development, 81*(5), 1565–1581.

Milesi, C., Perez-Felkner, L., Brown, K., & Schneider, B. (2017). Engagement, persistence, and gender in computer science: Results of a smartphone ESM study. *Frontiers in Psychology, 8*, 602. https://doi.org/10.3389/fpsyg.2017.00602

Moeller, J., Brackett, M. A., Ivcevic, Z., & White, A. E. (2020). High school students' feelings: Discoveries from a large national survey and an experience sampling study. *Learning and Instruction, 66*, Article 101301.

Muñoz, M. A., & Chang, F. C. (2007). The elusive relationship between teacher characteristics and student academic growth: A longitudinal multilevel model for change. *Journal of Personnel Evaluation in Education, 20*(3), 147–164. https://doi.org/10.1007/s11092-008-9054-y

Niepel, C., Marsh, H. W., Guo, J., Pekrun, R., & Möller, J. (2022). Revealing dynamic relations between mathematics self-concept and perceived achievement from lesson to lesson: An experience-sampling study. *Journal of Educational Psychology, 114*(6), 1380–1393. https://doi.org/10.1037/edu0000716

Nissen, J. M., & Shemwell, J. T. (2016). Gender, experience, and self-efficacy in introductory physics. *Physical Review Physics Education Research, 12*(2), Article 020105. https://doi.org/10.1103/PhysRevPhysEducRes.12.020105

Nuutila, K., Tuominen, H., Tapola, A., Vainikainen, M. P., & Niemivirta, M. (2018). Consistency, longitudinal stability, and predictions of elementary school students' task interest, success expectancy, and performance in mathematics. *Learning and Instruction, 56*, 73–83. https://doi.org/10.1016/j.learninstruc.2018.04.003

Peddycord-Liu, Z., Catete, V., Vandenberg, J., Barnes, T., Lynch, C., & Rutherford, T. (2019). A field study of teachers using a curriculum-integrated digital game. In S. Brewster, G. Fitzpatrick, A. Cox, & V. Kostakos (Eds.), *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (Paper 428). Glasgow, UK*.

Peixoto, F., Radišić, J., Krstić, K., Hansen, K. Y., Laine, A., Baucal, A., & Mata, L. (2022). Contribution to the validation of the expectancy-value scale for primary school students. *Journal of Psychoeducational Assessment, 07342829221144868*.

Perez, T., Cromley, J. G., & Kaplan, A. (2014). The role of identity development, values, and costs in college STEM retention. *Journal of Educational Psychology, 106*(1), 315.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*(1), 33–40. https://doi.org/10.1037/0022-0663.82.1.33

Pinxten, M., Marsh, H. W., De Fraine, B., Van Den Noortgate, W., & Van Damme, J. (2014). Enjoying mathematics or feeling competent in mathematics? Reciprocal effects on mathematics achievement and perceived math effort expenditure. *The British Journal of Educational Psychology, 84*(1), 152–174. https://doi.org/10.1111/bjep.12028

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevel SEM. *Structural Equation Modeling, 18*(2), 161–182. https://doi.org/10.1080/10705511.2011.557329

Preckel, F., Schmidt, I., Stumpf, E., Motschenbacher, M., Vogl, K., & Schneider, W. (2017). A test of the reciprocal-effects model of academic achievement and academic self-concept in regular classes and special classes for the gifted. *Gifted Child Quarterly, 61*(2), 103–116. https://doi.org/10.1177/001698621668782

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Sage.

Rutherford, T., Duck, K., Rosenberg, J., & Patt, R. (2021). Leveraging mathematics software data to understand student learning and motivation during the COVID-19 pandemic. Within a special issue of *Journal of Research on Technology in Education, 1-38*. https://doi.org/10.1080/15391523.2021.1920520.

Rutherford, T., Farkas, G., Duncan, G., Burchinal, M., Graham, J., Kibrick, M., … Martinez, M. E. (2014). A randomized trial of an elementary school mathematics software intervention: Spatial-Temporal (ST) math. *Journal of Research on Educational Effectiveness, 7*(4), 358–383. https://doi.org/10.1080/19345747.2013.856978

Rutherford, T., Liu, A. S., Lam, A. C., & Schenke, K. (2019). Impact on mathematics self-beliefs from a mastery-based mathematics software. *Journal of Research on Technology in Education, 52*(1), 79–94. https://doi.org/10.1080/15391523.2019.1689210

Rutherford, T., Liu, A. S., & Wagemaker, M. (2021). "I chose math because…": Cognitive interviews of a motivation measure. *Contemporary Educational Psychology, 66*, 1–24. https://doi.org/10.1016/j.cedpsych.2021.101992

Saw, G., & Chang, C. N. (2018). Cross-lagged models of mathematics achievement and motivational factors among hispanic and non-Hispanic high school students. *Hispanic Journal of Behavioral Sciences, 40*(2), 240–256. https://doi.org/10.1177/0739986318766511

Schenke, K., Rutherford, T., & Farkas, G. (2014). Alignment of game design features and state mathematics standards: Do results reflect intentions? *Computers & Education, 76*, 215–224. https://doi.org/10.1016/j.compedu.2014.03.019

Sewasew, D., Schroeders, U., Schiefer, I., Weirich, S., & Artelt, C. (2018). Development of sex differences in math achievement, self-concept, and interest from grade 5 to 7. *Contemporary Educational Psychology., 54*, 55–65. https://doi.org/10.1016/j.cedpsych.2018.05.003

Shadish, W., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference* (pp. 103–134). Boston, MA: Houghton Mifflin.

Shernoff, D. J., Kelly, S., Tonks, S. M., Anderson, B., Cavanagh, R. F., Sinha, S., & Abdi, B. (2016). Student engagement as a function of environmental complexity in high school classrooms. *Learning and Instruction, 43*, 52–60. https://doi.org/10.1016/j.learninstruc.2015.12.003

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4*, 1–32.

Shumow, L., Schmidt, J. A., & Zaleski, D. J. (2013). Multiple perspectives on student learning, engagement, and motivation in high school biology labs. *The High School Journal*, 232–252.

Skwarchuk, S. L. (2009). How do parents support preschoolers' numeracy learning experiences at home? *Early Childhood Education Journal, 37*(3), 189–197.

Sorvo, R., Koponen, T., Viholainen, H., Aro, T., Räikkönen, E., Peura, P., & Aro, M. (2019). Development of math anxiety and its longitudinal relationships with arithmetic achievement among primary school children. *Learning and Individual Differences, 69*, 173–181.

Stringer, R. W., & Heath, N. (2008). Academic self-perception and its relationship to academic performance. *Canadian Journal of Education, 31*(2), 327–345.

Usher, E. L., & Pajares, F. (2009). Sources of self-efficacy in mathematics: A validation study. *Contemporary Educational Psychology, 34*(1), 89–101.

Watt, H. M. (2004). Development of adolescents' self-perceptions, values, and task perceptions according to gender and domain in 7th-through 11th-grade Australian students. *Child Development, 75*(5), 1556–1574. https://doi.org/10.1111/j.1467-8624.2004.00757.x

Weidinger, A. F., Spinath, B., & Steinmayr, R. (2020). The value of valuing math: Longitudinal links between students' intrinsic, attainment, and utility values and grades in math. *Motivation Science, 6*(4), 413–422. https://doi.org/10.1037/mot0000179

Weidinger, A. F., Steinmayr, R., & Spinath, B. (2017). Math grades and intrinsic motivation in elementary school: A longitudinal investigation of their association. *The British Journal of Educational Psychology, 87*(2), 187–204. https://doi.org/10.1111/bjep.12143

Weidinger, A. F., Steinmayr, R., & Spinath, B. (2018). Changes in the relation between competence beliefs and achievement in math across elementary school years. *Child Development, 89*. https://doi.org/10.1111/cdev.12806

Wendt, S., Rice, J., & Nakamoto, J. (2019). *A cross-state evaluation of MIND research Institute's ST math program and math performance*. WestEd.

Wigfield, A. (1994). Expectancy-value theory of achievement motivation: A developmental perspective. *Educational Psychology Review, 6*(1), 49–78.

Wu, S. S., Barth, M., Amin, H., Malcarne, V., & Menon, V. (2012). Math anxiety in second and third graders and its relation to mathematics achievement. *Frontiers in Psychology, 3*, 162.

Xie, K., Heddy, B. C., & Greene, B. A. (2019). Affordances of using mobile technology to support experience-sampling method in examining college students' engagement. *Computers & Education, 128*, 183–198.

Zhang, Q., & Rutherford, T. (2022). Grade 5 students' elective replay after experiencing failures in learning fractions in an educational game: When does replay after failures benefit learning?. In *Proceedings of the 12th International Conference on Learning Analytics and Knowledge* (pp. 98–106). The Association for Computing Machinery. Held virtually.