Journal of Research on Technology in Education

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ujrt20

Leveraging mathematics software data to understand student learning and motivation during the COVID-19 pandemic

Teomara Rutherford, Kerry Duck, Joshua M. Rosenberg & Raymond Patt

To cite this article: Teomara Rutherford, Kerry Duck, Joshua M. Rosenberg & Raymond Patt (2021): Leveraging mathematics software data to understand student learning and motivation during the COVID-19 pandemic, Journal of Research on Technology in Education, DOI: 10.1080/15391523.2021.1920520

To link to this article: https://doi.org/10.1080/15391523.2021.1920520

Published online: 16 Jun 2021.

JOURNAL OF RESEARCH ON TECHNOLOGY IN EDUCATION
https://doi.org/10.1080/15391523.2021.1920520

Leveraging mathematics software data to understand student learning and motivation during the COVID-19 pandemic

Teomara Rutherford[a], Kerry Duck, Joshua M. Rosenberg[b] and Raymond Patt

[a]School of Education, University of Delaware, Newark, Delaware, USA; [b]STEM Education/Science, College of Education, Health, and Human Sciences, University of Tennessee, Knoxville, USA

ABSTRACT
School closures during the COVID-19 pandemic presented a threat to student learning and motivation. Suspension of achievement testing created a barrier to understanding the extent of this threat. Leveraging data from a mathematics learning software as a substitute assessment, we found that students had lower engagement with the software during the pandemic, but students who did engage had increased performance. Students also experienced changes in motivation: lowered mathematics expectancy, but also lower emotional cost for mathematics. Results illustrate the potential and pitfalls of using educational technology data in lieu of traditional assessments and draw attention to access and motivation during at-home schooling.

KEYWORDS
Emergency distance learning; mathematics education; motivation; learning technologies

ARTICLE HISTORY
Received 18 September 2020
Revised 27 March 2021
Accepted 19 April 2021

As schools across the U.S. closed in response to the COVID-19 pandemic, continuity of learning was at the forefront of discussions surrounding education (e.g., Andrew et al., 2020; Kaffenberger, 2020; Kuhfeld & Tarasawa, 2020). Online education provided a ready substitute, but an imperfect one, as teachers had to hastily craft plans for emergency remote instruction (see Greenhow et al., 2020; Trust & Whalen, 2020). From the start, there were indicators many students would have difficulty accessing online curricula, due to lack of devices, Internet, and adult support (Bacher-Hicks et al., 2020; Dynarski, 2020; Lai & Widmar, 2020). Further, evidence from studies of online education suggests that virtual instruction may especially harm students from already marginalized communities (e.g., Agostinelli et al., 2020; Parolin & Lee, 2021; Xu & Jaggars, 2014). As schools relied more on parental support of learning, parents balanced work and schooling responsibilities; those from families with fewer resources or who spoke languages other than English faced particular challenges (Bacher-Hicks et al., 2020; Garbe et al., 2020; Sugarman & Lazarín, 2020). These concerns are especially pressing given the eventuality of more online instruction, either in response to the current pandemic or as educators deal with the next disruption or crisis. It is critical that researchers understand both the educational impact of the pandemic and how educators can support student online learning in the future.

There are challenges in meeting both of these goals. The first is that along with school closures, states across the U.S. suspended end-of-year academic assessments (Gewertz, 2020), removing a metric by which researchers could compare post- and pre-pandemic academic achievement. The second is that school closures also meant the end of many school-based research projects, leaving the specifics of much pandemic education unknowable. To surmount these challenges, we use data collected from a mathematics educational technology, Spatial Temporal (ST) Math, to provide insight into the engagement, performance, and motivation of a sample of fourth-grade students as they experienced pandemic-induced online schooling. Additionally, we link ST Math data with curricular information provided by Central[1] school

CONTACT Teomara Rutherford teomara@udel.edu School of Education, University of Delaware, 113 Willard Hall Education Building, 16 West Main Street, Newark, DE 19716, USA

district to understand how student use of and performance within ST Math during the pandemic aligned with teacher-led instruction. Evidence of alignment may indicate the extent to which our results on ST Math engagement and achievement can serve as stand-ins for broader instructional access and learning during the pandemic and can contribute further insights into the black box of schooling during emergency remote instruction.

Spatial temporal (ST) math as context

ST Math, created by MIND Research Institute (MIND), is a year-long standards-aligned supplemental mathematics tutorial program currently used by over 1.2 million students in 48 states across the United States. Within ST Math, the curriculum is divided into math content objectives, each bookended by a pre- and post-test. Each objective contains games that present math content using a similar game mechanic. Within each game, students complete between one and eight levels that each contain a number of puzzles representing specific math problems. ST Math focuses on student mastery of content by providing highly-scaffolded tutorials and practice opportunities, allowing multiple attempts with the provision of detailed visual feedback (Rutherford et al., 2014; 2019; Wendt et al., 2019). As ST Math gathers digital sources of data—what we refer to as log data—while students interact with the platform, it provides insight into student engagement and performance on the provided content.

Insights into achievement from software data

We draw on two sources of evidence to establish that data from educational technology can provide value toward understanding student learning experiences: (1) evidence that educational technology interventions produce gains on standardized assessments (Cheung & Slavin, 2013) and (2) evidence that performance within educational technology is itself related to other classroom or standardized measures of performance (e.g., Liu et al., 2017). We focus on K-12 mathematics as the area most relevant to our study.

Regarding the first stream of evidence, meta-analyses of the impacts of mathematics educational technology reveal small positive effects from such programs—effects range depending on the nature of the program, assessment, study design, and age of students (Cheung & Slavin, 2013; Clark et al., 2016; Higgins et al., 2019; Pellegrini et al., 2020). In particular, use of tutorial and practice programs, such as DreamBox, ASSISTments, and Cognitive Tutor, have been linked with improved mathematics test scores outside of the platform (Pane et al., 2014; Roschelle et al., 2016; Wang & Woodworth, 2011). Although these studies provide evidence that educational technology can improve student performance, relative effects depend on how students interact with the technology (Bullock et al., 2015; Kosko, 2017; Moyer-Packenham et al., 2016; Yuan et al., 2010). These interactions may be driven by features outside the technology (e.g., making connections between technology and mathematics content), but are also related to features within the technology, such as accuracy feedback and providing multiple attempts (Filsecker & Hickey, 2014; Hwang & Lai, 2017; Moyer-Packenham et al., 2019; Syal & Nietfeld, 2020). In considering ST Math specifically, a number of studies have demonstrated that use of ST Math is associated with positive gains on standards-aligned assessments across 16 states, and gains may be especially pronounced where alignment in strong between content covered within ST Math and that assessed (e.g., Schenke et al., 2014; Wendt et al., 2019).

Regarding the second stream of evidence, studies utilizing data mining have revealed links between performance within educational technology and measures of performance obtained from beyond the technology itself (e.g., Ritter et al., 2013; San Pedro et al., 2015; Umer et al., 2019). For example, San Pedro and colleagues (2015) and Pardos and colleagues (2014) have found that performance within the ASSISTments tutorial platform positively correlated with state exam scores. In online courses, Agudo-Peregrina et al. (2014) have found that log data regarding course interactions positively related to student course performance in online courses. Within ST Math, Liu and colleagues (2017) found that student performance on game levels was predictive of year-over-year gains in state standardized achievement tests. Mirroring the variance in impacts from technology interventions, the predictive power on in-platform assessments depends on alignment between platform content, test items, and classroom instruction (Schenke et al., 2014; DiCerbo et al., 2017). In addition, when students engage more meaningfully with the platform, metrics of this engagement are more likely to be related to measures of performance outside of the platform (Brezovszky et al., 2019; Filsecker & Hickey, 2014; Syal & Nietfeld, 2020).

Insights into student motivation and engagement from software data

Student motivation is of practical significance to educators and administrators—students are more likely to fully engage with content and to achieve more when they are motivated (Schunk et al., 2012; Simpkins et al., 2006). Also, when motivational measures are included as predictors of achievement, they add explanatory power over and above measures of prior achievement (e.g., Anderman et al., 2010; Hough et al., 2017).

Within one prominent theory of motivation, Expectancy-Value Theory (EVT; Eccles et al., 1983; Wigfield & Eccles, 2002), students are more likely to engage with and succeed in activities if they expect to achieve positive outcomes and if they value the activities (Bong et al., 2012; Denissen et al., 2007; Marshall & Brown, 2004; Simpkins et al., 2006; Trautwein et al., 2012). Within EVT (Wigfield & Eccles, 2002), a student is said to value an activity if they find it interesting or enjoyable (intrinsic value), useful (utility value), and important to them (attainment value), and if they do not perceive it as requiring large costs (e.g., effort, time, frustration). We frame our current study with EVT, but recognize the relevance of research situated in other motivational theories (e.g., Achievement Goals; Ames & Archer, 1988) given the overlapping nature of many motivational constructs (Linnenbrink-Garcia & Patall, 2016).

Motivation is itself an important outcome malleable to teaching and schooling effects (e.g., Ruzek et al., 2015; Schiefele & Schaffner, 2015; Urdan & Schoenfelder, 2006). There is evidence that student motivation may have been influenced by the change in environment from the shift to emergency remote instruction; for example, motivation may have declined due to not being in class or from having to adapt to remote learning (e.g., Garbe et al., 2020; Smith et al., 2021). Having fewer social interactions, as well as personal interactions with the teacher may also influence intrinsic value and utility value for a given task (Agostinelli et al., 2020; Bergin, 1999; Lazarides et al., 2019). Lastly, reduced interactive instruction, including teacher feedback and class activities, may have resulted in declines in student self-efficacy (Bandura, 1997; Hattie & Timperley, 2007).

Educational technology can provide information related to students' motivation via digital data sources logged within the software. Data about choice behaviors may be inferred (e.g., time on task, returning to a task for remediation, help seeking, attentional focus) or from measures situated within the technology (e.g., embedded surveys). Within-platform measures of motivation are often used together with measures of platform performance to explain more variance in external assessments—both the San Pedro et al. (2015) and Pardos et al. (2014) studies noted above included such measures developed from log data within software; these measures explained unique variance in state test scores. Educational technology can also provide a platform for examining student motivation at scale—Ostrow and Heffernan (2018) leveraged ASSISTments to reach a broad sample of students in their validation study of a measure of intrinsic value. This ability to give voice to students is even more critical during remote instruction.

As both complementary to and distinct from motivation, engagement is itself predictive of learning and achievement (Baroody et al., 2016; Fredricks et al., 2004; Putwain et al., 2019). Fredricks et al. (2004) describe a three-dimensional conceptualization of engagement comprising emotional, cognitive, and behavioral engagement. Expectancies for success as well as attainment value, as in EVT, have been linked to behavioral engagement, and in turn linked to achievement (Fan, 2011; Putwain et al., 2019). Behavioral engagement is likely the aspect most amenable to measurement with data from educational technology—data generated as students interact with technology can be analogous to the classroom observations more traditionally used to measure behavioral engagement (e.g., Dotterer & Lowe, 2011). Relevant data from educational technology include factors such as time on task, number of sessions played, and number of attempts, which have been used in prior research as measures of behavioral engagement (e.g., Ghergulescu & Muntean, 2016; Syal & Nietfeld, 2020).

Technology integration and temporal alignment

In considering schooling during the pandemic, many districts relied on a combination of educational technology programs, enhanced instructional resources (i.e., e-textbooks with linked resources), and teacher-led instruction (Amador et al., 2020; Peterson et al., 2020). The synergy between these instructional formats can be thought of as their alignment. Much of the prior research on the topic of alignment has considered how resources, including educational technology, are aligned with the content of the curriculum (e.g., Xie et al., 2018) or curricular standards (e.g., Schenke et al., 2014; Resnick et al., 2004). This is a critical step in determining the value of technology resources, but alignment extends beyond these aspects—it also applies to time as represented by alignment between technology and teacher instruction *at that moment*. For the purposes of our study, we refer to this contemporaneous matching of technology coverage and teacher instruction as *temporal alignment*. Temporal alignment between instruction and technology use is an important facet to be explored, and may help with reducing barriers to using games and other technologies within the classroom (Baek, 2008; De Grove et al., 2012; Watson & Yang, 2016). For example, teachers can use the data from technology to inform their practice, make recommendations for the students, and provide students with additional practice (Admiraal et al., 2020). Prior research shows that teachers find integration through aligning technology and classroom instruction as challenging but desirable (e.g., Baek, 2008; McCulloch et al., 2018; Watson & Yang, 2016).

ST Math and similar tutorials (e.g., DreamBox, Imagine Math) can provide content to drive classroom instruction—for example, supporting student's engagement with the software and their mathematical learning outside of the technology through whole-class and small group discussions around mathematical concepts (Anderson-Pence et al., 2020). In order to capitalize on this synergy, it is necessary that the game content covers the same or similar topics at the same time as the content is covered in the classroom. Such temporal alignment may be especially important in mathematics, where the developmental progression of skills requires a specific ordering of instruction (Clements & Sarama, 2004). As self-paced supplemental technology programs for mathematics proliferate, temporal alignment becomes increasingly challenging. Prior research on ST Math has focused on the practice of reordering objectives to align in-game instruction with classroom curriculum (e.g., Callaghan et al., 2017; Peddycord-Liu et al., 2017; Rutherford et al., 2020). Callaghan and colleagues (2017) found that this practice was both rare and associated with better student achievement outcomes. Peddycord-Liu and colleagues (2019) interviewed and observed a small sample of teachers using ST Math in their classrooms and reported that strong ST Math teachers reordered objectives to match their classroom instructional sequence, but also intervened to move lagging students forward or slow down students who were too far ahead of the classroom curriculum. If teachers are engaging in these practices within Central, it may indicate that ST Math is a more integrated element of instruction; such temporal alignment may also strengthen student performance within the platform, as students will have received contemporaneous instruction on relevant skills from multiple sources. Evidence that ST Math is not temporally aligned with other simultaneous instruction could weaken our conclusions that ST Math performance is representative of pandemic learning; however, if both ST Math and other instruction are aligned to the content in state standards, ST Math may still be representative of performance vis-a-vis these standards.

Current study

We leverage data from within ST Math to gain insight into fourth grade students' learning experiences during spring 2020 of the COVID-19 pandemic and to understand how the pandemic may have impacted student performance, engagement, and motivation. Because ST Math was used both before and during the pandemic, it provides a ready comparison group assessed with the same within-platform metrics.

In evaluating student performance, engagement, and motivation during the pandemic, we acknowledge that threats to internal validity (Shadish et al., 2002) will make it difficult to disentangle associations with the pandemic from those of cohort or grade. We use a difference in difference approach (Wing et al., 2018) to examine how the change in these metrics from third to fourth grade differ between students who experience online instruction during the pandemic in the spring of their fourth grade year from the prior cohort, who completed fourth grade in 2019, before the pandemic. We specifically ask, *Question 1: In what ways did fourth grade student engagement, performance, and motivation within ST Math change from before to after the pandemic-induced move to remote instruction, and how do those changes compare to a similar period of time in a previous cohort that did not experience a pandemic?* We then focus this question on two demographic groups who may be hit particularly hard by the pandemic interruption to schooling (Agostinelli et al., 2020; Parolin & Lee, 2021), to ask, *Does this change vary depending on student free/reduced lunch eligibility (as a measure of socioeconomic status) and English Language Learner status?*

We then attempt to understand how students' experiences with ST Math during the pandemic may have fit within their broader instructional experiences. As students played ST Math through the pandemic, they may have used their time on the platform to play levels related to what they were learning from remote instruction as a means to reinforce learning, which would represent a tighter integration between the educational technology and other learning. To explore this, we link ST Math data with curricular information provided by Central to determine the extent to which students' ST Math play aligned with outside-of-ST Math instruction. We ask, *Question 2: (a) Is student engagement with ST Math temporally aligned with district instruction during the pandemic? (b) Is alignment associated with student performance and motivation?*

Method

Study context

Data for this study were gathered from the mathematics learning platform, ST Math, as part of a larger NSF-funded multi-year project investigating ST Math within a number of school districts. The ST Math curriculum was consistent year-over-year during our study period, which enables us to compare students who engaged with ST Math in the spring of 2020 to students of the same grade who engaged with ST Math in the spring of 2019.

We focus on Central School District, a district located in the South-Central United States with established ST Math use. Central closed their campuses in response to the Coronavirus pandemic in mid-March with remote instruction beginning approximately one week later. For remote instruction, the district provided a weekly instructional sheet with recommended videos and websites. In these weekly instructional sheets, there was an insert that recommended student use of ST Math for 60-90 minutes each per week and an additional 60-90 minutes per week of a different mathematics program, Imagine Math. Televised lessons were approximately 30 minutes each, twice per week.

Procedure

Log data from within ST Math were collected passively as students interacted with the software, except for the motivation surveys, which were embedded in the platform and presented to students three times a year: at the beginning of the school year, at the first log-in January after winter break, and a month before the end of the school year. These data were combined with district-collected data on student demographics by MIND, who de-identified the data before sharing with our research team. The surveys of interest in the current study are those administered one month before the end of the school year. For all students, this came after the move to emergency online instruction due to the pandemic.

Data on district instruction were collected weekly via the district's public website. The weekly instructional sheets were saved as pdfs, and the instructional videos were screen captured using an audio/visual recording software (Camtasia).

Participants

Participants were students within Central district who were in fourth grade in either the 2019-2020 school year (pandemic cohort) or the 2018-2019 school year (prior cohort). Central provided demographics for all students within the district for the 2017-2018 and 2018-2019 school years. We derived our sample from the 2018-2019 demographics, with the pandemic cohort in third grade that year and the prior cohort in fourth. Table 1 provides demographics for both the complete district-provided sample and the matched sample, across both year cohorts. 3,363 students were provided for the pandemic cohort and 3,302 for the prior cohort. Of these, 2,924 (87%) and 2,715 (82%) respectively were able to be matched with ST Math data. An additional 186 students were excluded because they repeated a grade or were working on an off-grade ST Math curriculum, leaving a sample of 2,800 for the pandemic cohort and 2,653 for the prior cohort, 83% and 80% of the starting sample. There were some small differences in demographics between the total district-provided sample and the analysis sample that varied depending on cohort. There were no statistically significant demographic differences between the pandemic and prior cohorts (*ps* > .05).

Measures

Each of the in-game measures described below were calculated for each year of ST Math relevant to the study: 4th grade in 2019-2020 and 3rd grade in 2018-2019 for the pandemic cohort and 4th grade in 2018-2019 and 3rd grade in 2017-2018 for the prior cohort.

Engagement

We measured student behavioral engagement with the software in two ways. First, we used annual total minutes logged within ST Math. Unfortunately, minutes were provided as an annual

Table 1. Demographics of Pandemic and Prior Cohorts, for Total District Sample and Sample Matched with ST Math Data for Analysis

	Total District Sample		Sample for Analysis		*p*-value from tests of differences		
Demographic Category	Pand.	Prior	Pand.	Prior	Pand.: Tot vs. Sample	Prior: Tot vs. Sample	Pand. vs. Prior
Hispanic	30%	31%	30%	32%	0.831	0.085	0.059
Black/African Amer.	35%	34%	35%	33%	0.054	0.128	0.073
White	23%	22%	22%	21%	0.007	0.133	0.588
Two or More Races	8%	8%	8%	9%	0.59	0.154	0.719
Other Race	5%	5%	5%	5%	0.453	0.5	0.6
Boy	51%	52%	51%	51%	0.969	0.356	0.965
Eng. Lang. Learner	12%	13%	12%	13%	0.312	0.129	0.248
Free/Reduced Lunch	65%	65%	65%	65%	0.143	0.647	0.548
Special Education	14%	14%	13%	12%	<.001	<.001	0.576
Talented/Gifted	5%	5%	5%	6%	0.96	<.001	0.284
N	3,302	3,363	2,800	2,653			

Note. *p*-values from Chi2 tests of statistical significance. Total vs. Sample represent tests comparing those excluded from those retained. Pand. is the pandemic cohort who were in fourth grade during the 2019-2020 school year.

total, not divided into before/after mid-March, so this measure was less sensitive to timing of the pandemic, as it included minutes prior to the March switch to online instruction. As a more temporally-sensitive measure, we used the number of levels attempted after the March switch to online instruction. This measure provided a different picture of use/engagement and may be less precise in that each level varied in difficulty and specific content, and therefore varied in time required to complete. By using measures that vary in their level of sensitivity and sample included, we are able to present a more complete picture of students' experiences—the yearlong minutes measure allows us to include all fourth grade students, even those who stopped playing after school closures, whereas the level attempts measure focuses on only those students who played at least once after the closure.

Performance

Similar to the engagement measures, our performance measures varied in precision and specific focus. To match our year-long minutes measure, we examined end-of-year content progress. This can be seen as a combined measure of behavioral engagement and performance—progress depends heavily on how often students log on to ST Math, but also on their successful completion of levels to move through the curriculum. We used two other measures of performance less dependent on student logins: average level attempt score for levels not previously passed and average objective posttest score. Average level attempt score for levels not previously passed is similar to the measure used in Liu et al. (2017). To create this measure, we limited the attempts to only those before a student had passed the particular level and we then averaged the score for each of these attempts. Both level attempt score and posttest score are presented as percentage correct and only include data from after the mid-March online transition.

Motivation for mathematics

The surveys embedded within ST Math were designed around EVT (Eccles et al., 1983; Wigfield & Eccles, 2002). In the ST Math survey, expectancy, utility, and importance for mathematics were each measured with two questions answered with a five-point Likert-type scale illustrated with animated tomato characters ("tamojis") that changed expression depending on the Likert scale value. For each of these constructs, we took the average of the two questions (alphas expectancy: .73, utility: .71, importance: .73).

Interest was measured with a single question focusing on enjoyment, asking students to rank order nine subjects from most favorite (1) to least favorite (9). We reverse scored the rank-order of mathematics so that a higher number represents greater enjoyment.

Cost was measured with students' reported academic emotions. Within the ST Math survey, students were asked to choose up to three emotions for each of three subjects: their most favorite (from the enjoyment question), their least favorite, and mathematics. Student emotions were coded for the presence of the negative emotions of frustrated, nervous, and bored for the mathematics sub-question. In this way, students could have values between zero (no negative emotions) to three (all negative emotions). See Figure 1 for question format examples.

Temporal alignment

Temporal alignment between curriculum encountered in ST Math and instruction during spring 2020 was based on grade-level instructional sheets and videos. Each week's lessons were coded with the relevant state mathematics content standards (e.g., Determine the measure of an unknown angle formed by two non-overlapping adjacent angles given one or both angle measures). An alignment guide provided by MIND Research Institute (2015) allowed the matching of content standards with objectives within the ST Math fourth grade curriculum. Some content standards were linked to more than one ST Math objective; when this was the case, all relevant ST Math objectives were listed with the associated standard. The list of ST Math Objective topics and relevant state mathematics standards are presented in Table 2.

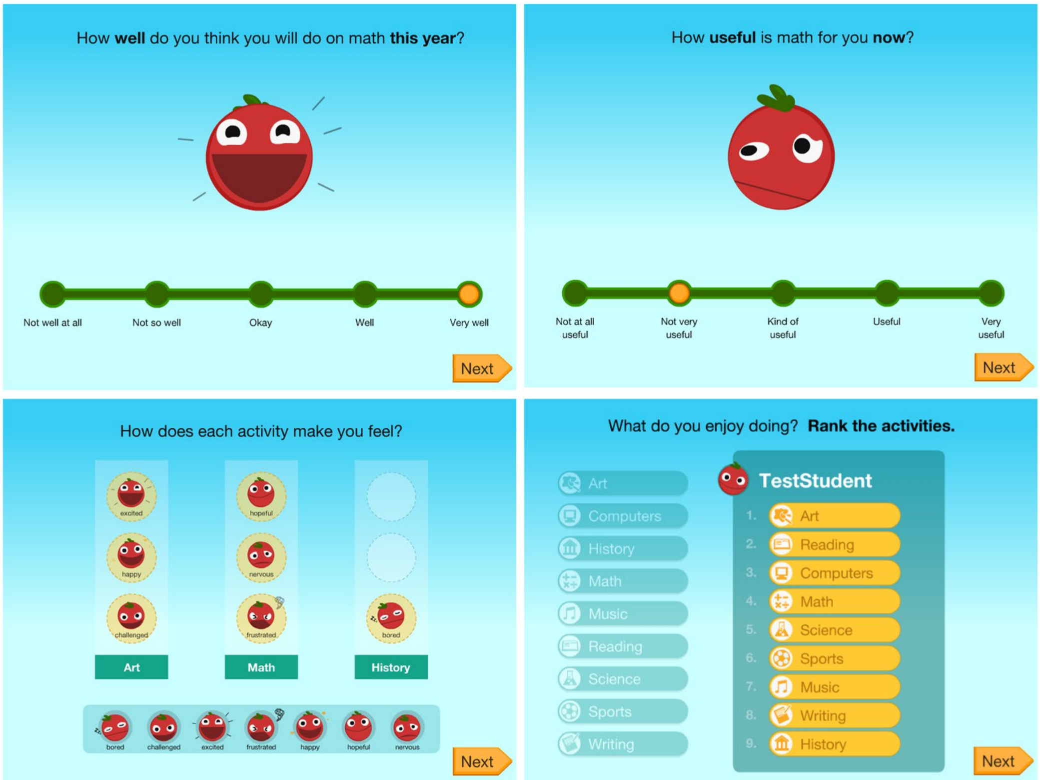

Figure 1. Examples of question presentation for surveys within ST Math.

Note. Top left shows an example of an expectancy question where student's answer would be coded as a "5" on the scale. Top right shows a utility question with an answer that would be coded as "2." Bottom left shows the emotional cost question—the focus subject is math, student would be given a 2 out of 3 for negative emotions. The bottom right panel shows the enjoyment question; math would be given a ranking of 4, which would be reverse-scored for analysis to 6.

Student ST Math level-plays after the end-of-March release of Central online materials (pandemic instruction week 2) were coded as covering content that matched that week's topic or covering content that did not match that week's topic.

Covariates

District-provided demographic information was used to create student-level covariates within the analyses. The district provided information on student ethnicity, gender, English Language Learner (ELL) status, disability status, whether the student was identified as talented and/or gifted, and whether they were eligible for the free/reduced lunch program. The ethnicity variable was combined by the district from two questions asked of parents: (1) Is the student Hispanic/Latino? And (2) What race is the student?. The latter had the options of American Indian/Alaskan Native, Asian, Black/African American, Native Hawaiian/Pacific Islander, or White. Individuals were allowed to specify more than one race. Individuals who answered yes to the first question were classified as Hispanic regardless of the answer to the second question. Individuals who answered no to the first question were classified by their race as specified in the second question (if only one) or as "Two or More Races" if more than one was selected. Gender was coded as male or female, which state data codebooks identified as "the gender of the person." We recoded gender to represent whether the participant was a boy or a girl.

Table 2. ST Math Objective Topics and Relevant Keywords

Week	Weekly Instructional Sheet Description	State Standards (Pulled from Mind, 2015)	ST Math Objective(s)
1	No Instructional Content		
2	Students are reviewing their understanding of points, lines, and angles in order to identify and classify two dimensional figures.	6D: Classify two-dimensional figures based on the presence or absence of parallel or perpendicular lines or the presence or absence of angles of a specified size. 6A: Identify points, lines, line segments, rays, angles, and perpendicular and parallel lines.	8, 14, 15, 16
3	Students are learning about the measurement of angles. This includes determining the approximate measures of angles and measuring angles as part of a circle.	7A: Illustrate the measure of an angle as the part of a circle whose center is at the vertex of the angle that is 'cut out' by the rays of the angle. Angle measures are limited to whole numbers. 7C: Determine the approximate measures of angles in degrees to the nearest whole number using a protractor.	8
4	Students will continue to learn about the measurement of angles. Students are learning how to determine the measure of an unknown angle formed by two adjacent angles when given one or both angle measures and how to draw an angle with a given measure.	7E: Determine the measure of an unknown angle formed by two non-overlapping adjacent angles given one or both angle measures.	8
5	Students will be working on making connections between various math concepts that they have learned in previous units. This week, students will review their understanding of adding/subtracting fractions with equal denominators. Students will also review input/output tables.	3: The student applies mathematical process standards to represent and generate fractions to solve problems. 3E: Represent and solve addition and subtraction of fractions with equal denominators using objects and pictorial models that build to the number line and properties of operations.	10,11
6	Students will be working on making connections between various math concepts that they have learned in previous units. This week, students will review solving problems with elapsed time and the measurement of length. Students will also review representing data in various ways.	8C: Solve problems that deal with measurements of length, intervals of time, liquid volumes, mass, and money using addition, subtraction, multiplication, or division as appropriate.	31
7	Students will be reviewing essential understandings of fractions. This week, students will focus on comparing two fractions with different numerators and denominators.	3D: compare two fractions with different numerators and different denominators and represent the comparison using the symbols for greater than, less than, and equal.	7
8	Students will be reviewing essential understandings of fractions. This week, students will be learning to relate decimals to fractions that name tenths and hundredths.	3G: Represent fractions and decimals to the tenths or hundredths as distances from zero on a number line.	18, 19
9	Students will be reviewing essential understandings of all operations, to include problem solving. This week, the focus will be on addition and subtraction of decimals, as well as multiplication and division of whole numbers.	4: The student applies mathematical process standards to develop and use strategies and methods for whole number computations and decimal sums and differences in order to solve problems with efficiency and accuracy. 4H: Solve with fluency one- and two-step problems involving multiplication and division, including interpreting remainders.	20,21,22
10	Students will be reviewing their understanding of problem-solving using money. Students will also be learning about personal financial literacy, including how to calculate profit and how to distinguish between fixed and variable expenses.	2E: Represent decimals, including tenths and hundredths, using concrete and visual models and money.	18,31

Analysis

For *Question 1*, each student contributed two years of data—their third grade year and their fourth grade year. We regressed each outcome on cohort, grade-level, student demographics, and an interaction term for grade-by-cohort. The statistical significance of the interaction term within each regression indicates how the change in outcome from third to fourth grade is different between the pandemic and prior cohorts, in short, the *association* of the pandemic and our outcome variables: Positive values of the coefficient indicate that the outcome's value was greater during the pandemic, whereas negative values indicate that the outcome's value was lesser. Hereafter we refer to this as the grade X pandemic interaction. The data did not allow for a traditional multilevel model (Raudenbush & Bryk, 2002; West et al., 2014) wherein each timepoint is nested within the student and each student is nested within the teacher, because, due to the multi-year nature of the data, students and teachers are cross-classified across years; students are not uniquely nested within a single teacher. To account for this cross-classification, we used a cross-classified multilevel model (Raudenbush & Bryk, 2002; West et al., 2014), with observations per student nested within students and students cross-classified with teachers. In more detail, using the notation outlined by Gelman and Hill (2006), our model was specified as follows:

$$\text{outcome}_i \sim N(\alpha_{j[i],k[i]} + \beta_1(\text{gr4}) + \beta_2(\text{pandemic}) + \beta_3(\text{pandemic X gr4}), \sigma^2)$$

$$\alpha_j \sim N(r_0^{\alpha} + r_{1-10}^{\alpha}(\text{covariates}), \sigma_{\alpha_j}^2), \text{for student j=1,\dots,J}$$

$$\alpha_k \sim N(\mu_{\alpha_k}, \sigma_{\alpha k}^2), \text{for teacher k=1,\dots,k}$$

Where represents the model intercept, which randomly varied between both students and teachers. The first line of the equation represents within-student variation, which was assumed generated by a normal distribution with a mean structure according to when the student was in grade four, when the pandemic occurred, and their interaction. The second line of the equation represents between-student variation, which was similarly assumed generated from a normal distribution and included additional student-level covariates, including student demographic variables. These covariates account for between-student (rather than within-student) variation. Finally, the intercept is also specified as randomly varying between teachers, and this variation was assumed generated by a normal distribution.

To estimate this model, we used the *lme4* package (Bates et al., 2015) using the R (R Core Team, 2020) statistical software.[2] For each analysis, we present standardized coefficients using the formula $B^*(SD_x)/SD_y$, where $SD_x = 1$ for dummy variables, using the standard deviations from the fourth grade non-pandemic cohort. This places our coefficients in the units of each dependent variable's control group standard deviation.

For each grade X pandemic interaction term that met a $p < .05$ statistical significance threshold, we also calculated the percent bias to invalidate the inference (PBI) values using the robustness analysis technique developed by Frank et al. (2013) via the konfound (Rosenberg et al., 2020) R package, which calculates how robust an inference is to alternative explanations and sources of bias, including sample- and measurement-related bias and omitted, confounding variables. One approach to this is to quantify how much of the effect would need to be due to bias for the effect to be invalidated. The output of Frank et al.'s (2013) technique is a percentage with the range of 0-100%; values approaching 100% would indicate that nearly all of the effect would need to be due to bias for an inference to not be made about the effect; values approaching 0% indicate that nearly none of the effect would need to be due to bias for an inference to be invalidated. Thus, higher PBI values indicate that an effect may be more robust to sources of bias, although these values should be interpreted in the context of each study.

After determining the association between the pandemic and each of our outcomes, we then extended our analyses in Question 1 to examine demographic moderators of the association between the pandemic and outcomes. Namely, we asked whether the coefficient for the grade-by-cohort variable differed depending on free/reduced lunch status or ELL status by adding interaction terms between each moderator variable and cohort, grade, and grade-by-cohort, the latter creating a three-way interaction.

To address *Question 2*, we used data only from Spring 2020. First, we limited objectives to those that were taught at any point during spring after the switch to online instruction. We visually examined the alignment between these objectives and teacher-led instruction by graphing the percent of play on each objective by week and how this percentage matched with other instruction (greater percentage match would indicate greater temporal alignment). We then used the percentage match to predict engagement and achievement in regression analyses. These analyses are exploratory; our specific analytic methods emerged after examination of the data, therefore they are more fully described within the results section.

Results

Descriptive statistics by cohort

With the March 2020 switch to online instruction, nearly 30% of students (783) ceased logging on to ST Math. These students were more likely to be Black, boys, eligible for free/reduced lunch, and labeled as having a disability—differences presented in this section are from Chi2 tests of differences between samples. Prior springs also saw a reduction in the scale of ST Math use, but a much smaller one—in the spring of 2019, only 2% of students stopped logging in mid-March; in the spring of 2018, only 1% stopped logging on. There were also demographic differences in these earlier years between who continued playing ST Math after March and who did not. Figure 2 displays a comparison of the percentage of students who last logged into ST Math on a given week, broken down by the three years of data we examined. Table 3 displays demographics of those who continued playing and notes statistically significant differences between these students and those who stopped. Across years, students labeled as having a disability were more likely to stop playing ST Math before the end of the year. Table 4 displays our variables of interest by cohort and grade. Data available for each outcome differs depending on the last day of each student's log on. For example, students who did not log on after March school closures do not have spring motivation measures.

Before turning to our main questions, we examined correlations between the variables, expecting that each of our outcome variables would be more correlated with other outcomes within its category. Table 5 displays these correlations. We do see that our motivation variables were most strongly correlated with each other and two of our engagement variables and two of our performance variables were most strongly correlated. Progress presented an interesting case,

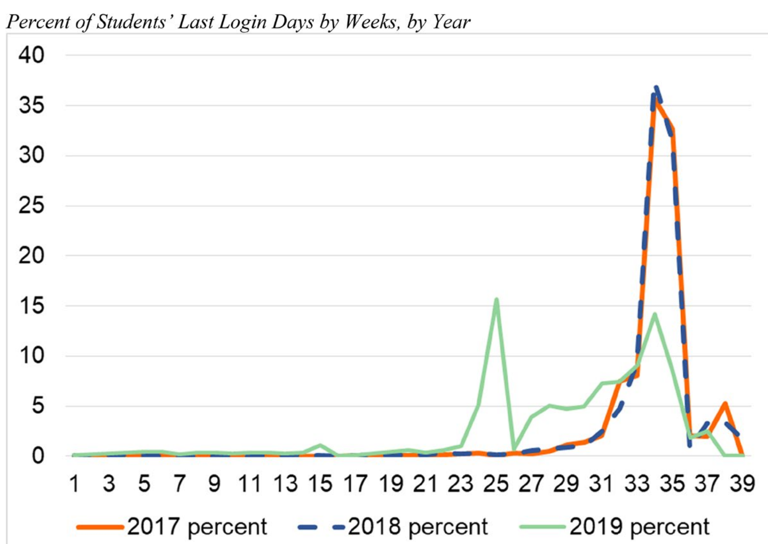

Figure 2. Percent of students' last login days by weeks, by year.

Note. 2017 includes students in third grade from the prior cohort, 2018 includes third graders (target/pandemic cohort) and fourth graders (prior cohort), 2019 includes fourth graders from the target/pandemic cohort. School closures occurred in week 25.

Table 3. Demographics of Pandemic and Prior Cohorts, by Grade and by Whether Student Logged on After Mid-March Each Year

	Fourth Grade				Third Grade			
	Pandemic Cohort		Prior Cohort		Pandemic Cohort		Prior Cohort	
Demographic Category	Sample	vs. Tot	Sample	vs. Tot	Sample	vs. Tot	Sample	vs. Tot
Hispanic	30%	0.416	32%	0.212	30%	0.461	32%	0.329
Black/African Amer.	***34%***	0.039	33%	0.092	35%	0.865	33%	0.133
White	22%	0.237	21%	0.632	22%	0.121	***21%***	0.03
Two or More Races	8%	0.660	***8%***	0.023	8%	0.171	8%	0.211
Other Race	5%	0.866	***5%***	0.015	5%	0.980	5%	0.743
Boy	***48%***	<.001	***51%***	0.023	51%	0.057	51%	0.144
Eng. Lang. Learner	12%	0.764	13%	0.163	12%	0.160	13%	0.421
Free/Reduced Lunch	***62%***	<.001	***65%***	<.001	65%	0.554	65%	0.119
Special Education	***12%***	0.017	12%	0.378	***12%***	<.001	***12%***	0.022
Talented/Gifted	6%	0.071	***5%***	<.001	***5%***	<.001	***6%***	<.001
N	2,017	2,800	2,585	2,653	2,759	2,800	2,627	2,653

Note. The pandemic cohort are those who were in fourth grade during the 2019-2020 school year and third grade during the 2018-2019 school year. Each "vs. Tot" column presents the *p*-value from Chi2 tests between the sample shown and the complete analysis sample for that grade/cohort, including those who stopped logging in after mid-March. Values bolded indicate the sample of students that logged on after mid-march each year had more students who represent the relevant category; bolded and italics indicate the sample had fewer students who represent the relevant category.

Table 4. Descriptive Statistics for Engagement, Performance, and Motivation Variables, by Cohort and Grade.

		Pandemic Cohort					Prior Cohort				
	GR	Mean	SD	Min	Max	Count	Mean	SD	Min	Max	Count
Content Progress Total	3	0.75	0.30	0.00	1.00	2,800.00	0.77	0.26	0.02	1.00	2,653.00
	4	0.63	0.32	0.00	1.00	2,800.00	0.75	0.28	0.00	1.00	2,653.00
Minutes Total	3	2848.19	973.00	0.00	6,060.00	2,800.00	3023.39	859.08	75.00	6,671.00	2,653.00
	4	2062.19	910.34	0.00	6,698.00	2,800.00	2678.94	868.01	0.00	6,789.00	2,653.00
PostTest Ave (Spring)	3	0.75	0.17	0.00	1.00	2,589.00	0.75	0.16	0.00	1.00	2,499.00
	4	0.77	0.19	0.00	1.00	1,555.00	0.74	0.18	0.00	1.00	2,397.00
Levels Attempted (Spring)	3	188.43	124.45	1.00	999.00	2,745.00	230.75	125.84	1.00	840.00	2,606.00
	4	98.65	124.12	1.00	1,480.00	1,929.00	158.27	116.56	1.00	769.00	2,576.00
Ave Level Score (Spring)	3	0.70	0.15	0.00	1.00	2,693.00	0.71	0.14	0.00	1.00	2,569.00
	4	0.74	0.20	0.00	1.00	1,878.00	0.72	0.15	0.00	1.00	2,455.00
Expectancy	3	4.18	0.92	1.00	5.00	2,487.00	4.19	0.92	1.00	5.00	2,427.00
	4	3.92	0.92	1.00	5.00	1,104.00	4.08	0.95	1.00	5.00	2,368.00
Utility	3	5.49	2.63	1.00	9.00	2,487.00	5.37	2.60	1.00	9.00	2,427.00
	4	5.26	2.61	1.00	9.00	1,104.00	5.17	2.59	1.00	9.00	2,368.00
Importance	3	4.35	0.92	1.00	5.00	2,487.00	4.37	0.88	1.00	5.00	2,427.00
	4	4.30	0.87	1.00	5.00	1,104.00	4.34	0.90	1.00	5.00	2,368.00
Enjoyment	3	4.38	0.89	1.00	5.00	2,487.00	4.40	0.87	1.00	5.00	2,427.00
	4	4.30	0.89	1.00	5.00	1,104.00	4.33	0.92	1.00	5.00	2,368.00
Cost	3	1.01	1.06	0.00	3.00	2,487.00	1.01	1.05	0.00	3.00	2,427.00
	4	1.04	0.91	0.00	3.00	1,104.00	1.16	1.08	0.00	3.00	2368.00
N						2,800					2,653

Table 5. Correlations between Outcome Variables.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) Minutes Total	1								
(2) Levels Attempted	0.45***	1							
(3) Content Progress	0.69***	0.31***	1						
(4) Ave Level Score	−0.02*	−0.04***	0.28***	1					
(5) Post Test Ave	−0.01	−0.11***	0.25***	0.43***	1				
(6) Expectancy	0.11***	0.04***	0.23***	0.10***	0.11***	1			
(7) Enjoyment	0.04***	0.01	0.23***	0.11***	0.11***	0.38***	1		
(8) Utility	0.07***	0.02	0.14***	0.06***	0.10***	0.55***	0.27***	1	
(9) Importance	0.07***	0.03**	0.11***	0.02*	0.06***	0.51***	0.29***	0.70***	1
(10) Cost	−0.04***	−0.01	−0.21***	−0.10***	−0.11***	−0.46***	−0.60***	−0.34***	−0.36***

Note. Correlations are calculated from comparison within grade and cohort. Engagement variables are color-coded blue (variables 1-2), performance are yellow (variables 3-5), and motivation are green (variables 6-10).
*$p < .05$; **$p < .01$; ***$p < .001$.

being more correlated with engagement measures than with performance measures, although progress is moderately correlated with each. Our sense that progress may be more reflective of engagement than the other performance measures is borne out in these data.

In addition, for each outcome, we examined the intraclass correlations for teacher and student. *ICC*s ranged from .014 to .460 for teachers and .086 to .498 for students. Generally, student *ICC*s were higher for the motivation measures based upon self-report data, whereas the teacher *ICC*s were higher for the achievement and engagement measures using log data. Full ICC results are presented in Table 6.

Question 1: Differences in changes from third to fourth grade engagement, performance, and motivation between pandemic and prior cohort

Marginal (for only the fixed effects) and conditional (for both the fixed and random effects) R^2 values for mixed models were calculated based on Nakagawa et al.'s (2017) via the performance R package (Lüdecke et al., 2020). Marginal R^2 values ranged from .006-.120; conditional R^2 values ranged from .314-.584. Results are provided in Table 7.

Table 6. Intraclass Correlations for Outcome Variables

Outcome Variable	Student	Teacher
Minutes Total	0.12	0.46
Levels Attempted	0.09	0.30
Content Progress	0.42	0.24
Ave Level Score	0.23	0.10
Post Test Ave	0.29	0.12
Expectancy	0.35	0.04
Enjoyment	0.50	0.03
Utility	0.30	0.01
Importance	0.33	0.02
Cost	0.39	0.03

Note. Intraclass correlations calculated from null cross-classified multilevel models (Raudenbush & Bryk, 2002; West et al., 2014), estimated separately for each outcome, with observations per student nested within students and students cross-classified with teachers models. Models estimated using the *lme4* package (Bates et al., 2015) using the R (R Core Team, 2020) statistical software.

Table 7. Regression Results

A. Engagement Outcomes

	(1)				(2)			
	Minutes Played				Level Attempts			
Predictors	B	SE	*p*	Beta	B	SE	*p*	Beta
Grade 4	**−319.05**	46.63	<0.001	**−0.37**	**−72.31**	5.63	<0.001	**−0.62**
Pandemic Cohort	**−161.85**	23.91	<0.001	**−0.19**	**−37.47**	3.59	<0.001	**−0.32**
Gr.4 BY Pandemic	**−396.46**	32.69	<0.001	**−0.46**	**−12.78**	5.30	0.016	**−0.11**
Hispanic	−32.18	23.06	0.163	−0.04	4.30	3.58	0.230	**0.04**
Black/African Amer.	31.09	22.16	0.161	0.04	9.74	3.43	0.005	**0.08**
Two or More Races	−23.34	31.37	0.457	−0.03	−4.71	4.86	0.332	−0.04
Other Race	13.46	38.46	0.726	0.02	−8.3	5.97	0.165	**−0.07**
Boy	**−118.54**	15.34	<0.001	**−0.14**	−3.45	2.38	0.148	**−0.03**
Eng. Lang. Learner	−27.11	30.12	0.368	−0.03	**13.10**	4.54	0.004	0.11
Free/Reduced Lunch	**−42.69**	17.41	0.014	**−0.05**	3.48	2.68	0.194	**0.03**
Special Education	**−185.87**	24.86	<0.001	**−0.21**	**16.34**	3.83	<0.001	**0.14**
Talented/Gifted	**−137.72**	35.77	<0.001	**−0.16**	**−40.88**	5.66	<0.001	**−0.35**
Constant	**3061.28**	40.55	<0.001	**−0.37**	**225.17**	5.06	<0.001	**−0.62**
Random Effects								
σ2	416,323.12				10,845.04			
τ00Student	97,862.25				1,333.09			
Teacher	366,466.75				3,157.26			
NsTeacher	642				636			
Student	5,451				5,420			
Obs.	10,902				9,854			
Marg R2/Cond R2	0.120 / 0.584				0.118 / 0.376			
PBI: Gr.4XPandemic	83.83				18.7			

B. Performance Results

	(3)				(4)				(5)			
	Content Progress				Level Score Ave				Posttest Ave			
Predictors	B	SE	*p*	Beta	B	SE	*p*	Beta	B	SE	*p*	Beta
Grade 4	**−0.03**	0.01	0.013	**−0.11**	**0.01**	0.01	0.008	**0.07**	<0.01	0.01	0.523	< 0.01
Pandemic Cohort	**−0.02**	0.01	0.003	**−0.07**	−0.01	<0.01	0.097	−0.07	<0.01	<0.01	0.359	< 0.01
Gr.4 BY Pandemic	**−0.08**	0.01	<0.001	**−0.29**	**0.02**	0.01	0.002	**0.13**	**0.02**	0.01	0.004	**0.11**
Hispanic	**−0.03**	0.01	0.001	−0.11	**−0.01**	<0.01	0.017	**−0.07**	−0.01	0.01	0.073	−0.06
Black/African Amer.	**−0.04**	0.01	<0.001	−0.14	**−0.03**	<0.01	<0.001	**−0.20**	**−0.03**	0.01	<0.001	**−0.17**
Two or More Races	−0.02	0.01	0.147	−0.07	0.01	0.01	0.428	0.07	<0.01	0.01	0.909	0.00
Other Race	**0.03**	0.02	0.036	0.11	−0.01	0.01	0.214	−0.07	0.01	0.01	0.335	0.06
Boy	**0.06**	0.01	<0.001	0.21	0.01	<0.01	0.056	0.07	**−0.01**	<0.01	<0.001	**−0.06**
Eng. Lang. Learner	−0.01	0.01	0.641	**−0.04**	<0.01	0.01	0.949	0.00	**−0.02**	0.01	0.001	**−0.11**
Free/Reduced Lunch	**−0.03**	0.01	<0.001	−0.11	**−0.01**	<0.01	0.011	**−0.07**	**−0.02**	<0.01	<0.001	**−0.11**
Special Education	**−0.15**	0.01	<0.001	**−0.54**	**−0.08**	0.01	<0.001	**−0.53**	**−0.10**	0.01	<0.001	**−0.56**
Talented/Gifted	**0.13**	0.01	<0.001	**0.46**	**0.08**	0.01	<0.001	**0.53**	**0.12**	0.01	<0.001	**0.67**
Constant	**0.77**	0.01	<0.001	**−0.11**	**0.73**	0.01	<0.001	**0.07**	**0.78**	0.01	<0.001	**< 0.01**
Random Effects												
σ2	0.03				0.02				0.02			
τ00Student	0.03				<0.01				0.01			
Teacher	0.02				<0.01				0.00			
NsTeacher	642				636				628			
Student	5,451				5,368				5,301			
Obs.	10,902				9,593				9,038			
Marg R2/Cond R2	0.089 / 0.669				0.069 / 0.326				0.083 / 0.405			
PBI: Gr.4XPandemic	75.49				32.46				32.36			

C. Motivation Outcomes, Expectancy, Enjoyment

	(6)				(7)			
	Expectancy				Enjoyment			
Predictors	B	SE	*p*	Beta	B	SE	*p*	Beta
Grade 4	**−0.12**	0.03	<0.001	**−0.13**	**−0.24**	0.07	0.001	**−0.26**
Pandemic Cohort	−0.01	0.03	0.696	−0.01	0.11	0.08	0.154	0.12
Gr.4 BY Pandemic	**−0.15**	0.04	<0.001	**−0.16**	−0.04	0.10	0.697	−0.04
Hispanic	−0.06	0.03	0.074	−0.06	0.08	0.10	0.434	0.09
Black/African Amer.	0.01	0.03	0.772	0.01	0.12	0.09	0.191	0.13
Two or More Races	0.01	0.05	0.898	0.01	−0.11	0.13	0.420	−0.12
Other Race	<0.01	0.06	0.960	< 0.01	0.11	0.16	0.494	0.12
Boy	**0.05**	0.02	0.016	**0.05**	**0.64**	0.07	<0.001	**0.70**
Eng. Lang. Learner	0.03	0.04	0.510	0.03	**0.23**	0.11	0.039	**0.25**
Free/Reduced Lunch	<0.01	0.03	0.924	< 0.01	0.14	0.07	0.051	0.15
Special Education	−0.06	0.03	0.106	−0.06	**0.20**	0.10	0.045	**0.22**
Talented/Gifted	**0.30**	0.05	<0.001	**0.32**	**0.57**	0.15	<0.001	**0.62**
Constant	**4.16**	0.04	<0.001	**−0.13**	**4.81**	0.10	<0.001	**−0.26**
Random Effects								
σ2	0.53				3.21			
τ00Student	0.30				3.27			
Teacher	0.03				0.17			
NsTeacher	602				602			
Student	5,095				5,095			
Obs.	8,385				8,385			
Marg R2/Cond R2	0.017 / 0.395				0.024 / 0.529			
PBI: Gr.4XPandemic	47.72				*ns*			

Starting with engagement, there were both cohort and grade differences not attributed to the pandemic. Comparisons are discussed in terms of the reference group, third graders from the cohort that was prior to the pandemic. Fourth graders and those in the pandemic cohort played fewer minutes during the year ($\beta = -0.37$, $p < .001$; $\beta = -0.19$, $p < .001$) and made fewer level attempts in the spring ($\beta = -0.62$, $p < .001$; $\beta = -0.32$, $p < .001$). Our coefficient of interest to

D. Motivation Outcomes, Utility, Importance, Emotional Cost

	(8)				(9)				(10)			
	Utility				Importance				Cost			
Predictors	B	SE	*p*	Beta	B	SE	*p*	Beta	B	SE	*p*	Beta
Grade 4	−0.03	0.02	0.190	−0.01	**−0.06**	0.02	0.008	**−0.07**	**0.16**	0.03	<0.001	**0.15**
Pandemic Cohort	−0.02	0.03	0.376	−0.01	−0.02	0.03	0.42	−0.02	<0.01	0.03	0.982	0.00
Gr.4 BY Pandemic	−0.02	0.04	0.603	−0.01	−0.01	0.04	0.767	−0.01	**−0.09**	0.04	0.024	**−0.08**
Hispanic	−0.01	0.03	0.771	< 0.01	0.01	0.03	0.873	0.01	−0.01	0.04	0.873	−0.01
Black/African Amer.	0.04	0.03	0.142	0.02	0.06	0.03	0.046	0.07	<0.01	0.04	0.915	< 0.01
Two or More Races	0.06	0.04	0.154	0.02	0.04	0.04	0.331	0.04	0.01	0.05	0.892	0.01
Other Race	**0.14**	0.05	0.011	**0.05**	**0.11**	0.05	0.035	**0.12**	−0.07	0.06	0.273	−0.06
Boy	0.01	0.02	0.691	< 0.01	−0.04	0.02	0.096	−0.04	**−0.19**	0.03	<0.001	**−0.18**
Eng. Lang. Learner	0.04	0.04	0.253	0.02	0.03	0.04	0.441	0.03	**−0.11**	0.04	0.010	**−0.10**
Free/Reduced Lunch	<0.01	0.02	0.961	0.00	0.03	0.02	0.151	0.03	−0.03	0.03	0.253	−0.03
Special Education	**−0.21**	0.03	<0.001	**−0.08**	**−0.12**	0.03	<0.001	**−0.13**	**−0.14**	0.04	<0.001	**−0.13**
Talented/Gifted	0.02	0.05	0.751	0.01	−0.03	0.05	0.565	−0.03	**−0.33**	0.06	<0.001	**−0.31**
Constant	**4.36**	0.03	<0.001	**−0.01**	**4.37**	0.03	<0.001	**−0.07**	**1.18**	0.04	<0.001	**0.15**
Random Effects												
σ2	0.55				0.52				0.64			
τ00Student	0.24				0.26				0.42			
Teacher	0.01				0.01				0.02			
NsTeacher	602				602				602			
Student	5,095				5,095				5,095			
Obs.	8,385				8,385				8,385			
Marg R2/Cond R2	0.009 / 0.314				0.006 / 0.347				0.022 / 0.421			
PBI: Gr.4XPandemic	*ns*				*ns*				14.49			

Note. Unstandardized coefficients represented as "B." To place coefficients in the units of each dependent variable's standard deviation, standardized coefficients (Betas) were calculated using the formula B*(SDx)/SDy (SDx = 1 for dummy variables), using the standard deviations from the fourth grade non-pandemic cohort. Bolded Bs and Betas are statistically significant at the $p < .05$ threshold. Reference group is students who are White, girls, not eligible for free/reduced lunch, and not designated as English Language Learners, special education, or talented/gifted. Ns for each model vary depending on the number of students who completed each outcome. All students had minutes and progress, level attempts includes all attempts, even those that were replays of previously-passed levels, level score only includes plays of levels not previously passed. Each student contributes one or two years of data. Random effects for student and teacher are estimated. Marginal (for only the fixed effects) and conditional (for both the fixed and random effects) R^2 values for mixed models are calculated based on Nakagawa et al. (2017) via the performance R package (Lüdecke, 2020). Percent bias to invalidate the inference (PBI) values are calculated using the robustness analysis technique developed by Frank et al. (2013) via the [name removed for peer review] (Authors 2020) R package.

the research question was the grade X pandemic interaction; this was statistically significant. Fourth graders in the pandemic cohort played fewer minutes that year (interaction $\beta = -0.46$, $p < .001$) and made fewer level attempts (interaction $\beta = -0.11$, $p = .016$). These differences were over and above grade-level or cohort differences. Figure 3 displays tiled interaction graphs for all of the outcomes that were statistically significantly predicted by the grade X pandemic interaction. In the graph for total minutes played, the solid line shows the trend from third to fourth grade of the cohort that completed fourth grade in 2018-2019. There is a dip in minutes played across grades. The dashed line displays the trend from third to fourth for the pandemic cohort that completed fourth grade in 2019-2020. Even in third grade, they played fewer minutes than the prior cohort; however, their fourth grade decline is steeper. This difference in rate of decline is represented by the statistically significant interaction term (Table 8).

Our three performance measures tell different stories. Annual content progress appears to follow a trend similar to that shown in the engagement measures. Fourth graders make less progress ($\beta = -0.11$, $p = .013$), as do those in the pandemic cohort ($\beta = -0.07$, $p = .003$). During fourth grade, the pandemic cohort makes the least progress ($\beta = -0.29$, $p < .001$). The two purer performance outcomes show the opposite trend. Those who played fourth grade ST Math during the pandemic had higher average level scores (interaction terms $\beta = 0.13$, $p = .002$) and posttest scores (interaction term: $\beta = 0.11$, $p = .004$) than would be expected from the prior cohort's third to fourth grade trend. There were fewer grade or cohort differences for spring

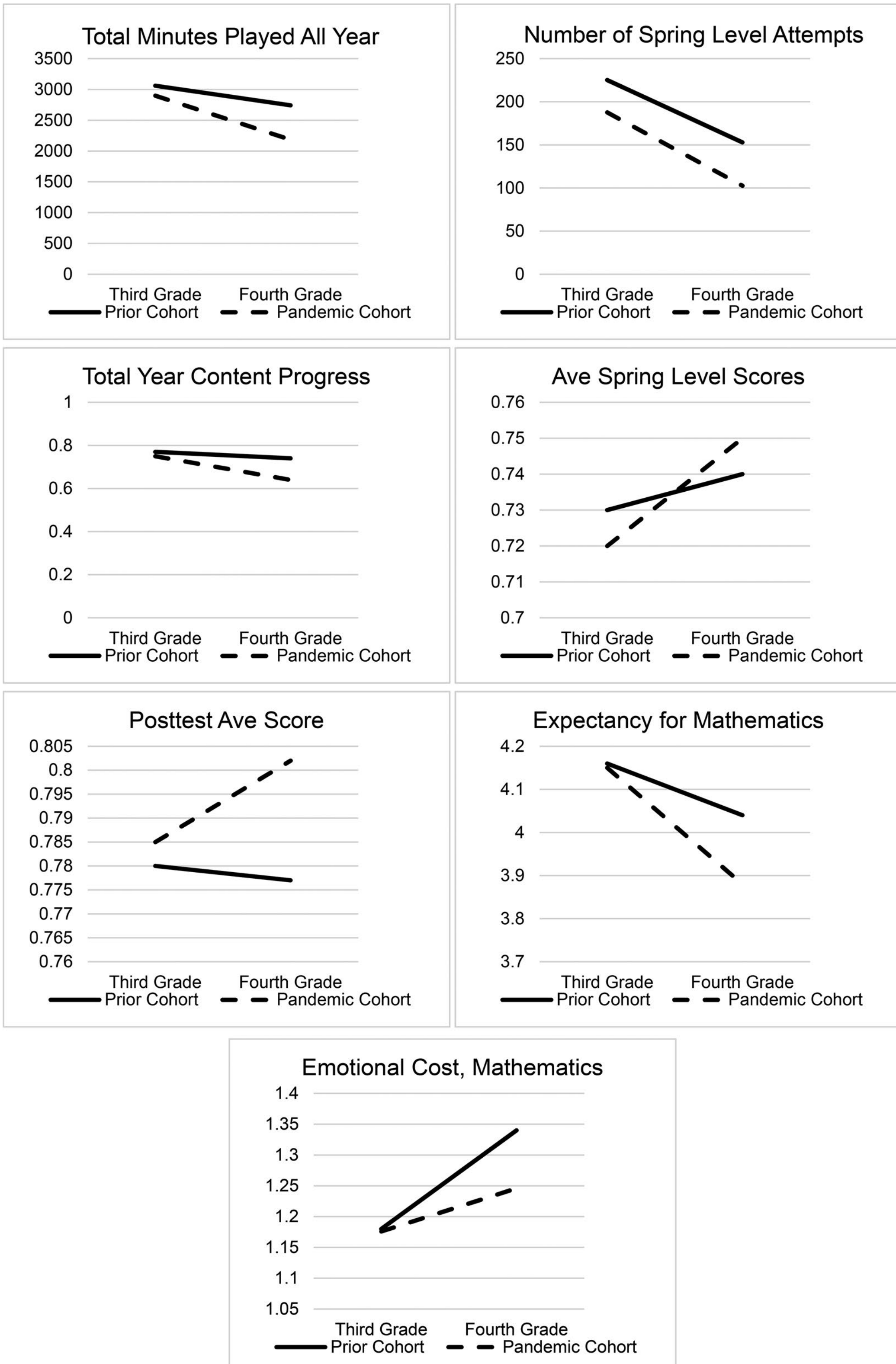

Figure 3. Regression-adjusted results by grade and cohort.

Note. Figures drawn from regression-adjusted values based on unstandardized coefficients (see Table 7). Only statistically significant interactions (at the p < .01 level*)* are shown.

performance—the only statistically significant difference outside of the interaction term was that fourth graders had higher spring level score averages than third graders ($\beta = 0.07$, $p = .008$); this may reflect content differences.

Finally, looking at our motivation variables, only expectancy for mathematics and cost of mathematics displayed statistically significant associations with the pandemic. Expectancy, along with enjoyment and importance, was lower for fourth graders than third graders ($\beta = -0.13$, $p < .001$). This dip in expectancy was even lower for students in fourth grade during the pandemic (interaction $\beta = -0.16$, $p < .001$). Interestingly, cost also showed negative associations with the pandemic—but due to the nature of the variable, this has different implications. There was greater reported emotional cost for mathematics when comparing fourth to third grade ($\beta = 0.15$, $p < .001$); however, during the pandemic, fourth graders reported less emotional cost (interaction $\beta = -0.08$, $p = .024$).

We have presented associations between our outcomes and the pandemic that met a statistical significance threshold of $p < .05$. We note that given our multiple comparisons, a Bonferroni-adjusted[3] p-value would be $p < .005$. This adjustment would exclude our findings regarding level attempts ($p = .016$) and cost value ($p = 0.024$); these results should be interpreted with caution. We present PBI as an explication of these concerns: 8.7% of the effect for level attempts and 14.5% of the effect for cost value would need to be due to bias for these effects to be invalidated; these are relatively small (Frank et al., 2013). In contrast, the other statistically significant effects had PBI values at or above the median from Frank et al. (2013), with the PBI for minutes played and content progress over 75%. For minutes played, invalidation would require that nearly 84% of the effect was due to bias.

Moderators of the cohort differences in third to fourth grade change

For each of our outcomes, we investigated two potential moderators: ELL status and eligibility for free/reduced price lunch. Only engagement metrics and content progress were statistically significantly moderated by either ELL status and free/reduced price lunch eligibility. The negative association between the pandemic and both minutes played and content progress was greater for those students eligible for free/reduced priced lunch. Figure 4 displays tiled graphs of the interactions. Here, as in the main effects models, content progress displays patterns more similar to engagement measures than achievement measures. For ELL students, the pandemic was only differentially associated with level attempts. There were no statistically significant ELL interactions for grade or cohort, but ELL fourth graders during the pandemic experienced a steeper decline in level attempts from third to fourth grade than did their prior cohort peers. A table of these results is in Appendix A.

Question 2: Temporal alignment between instruction and ST math

Description of temporal alignment

Each week's fourth grade Central district content aligned with between one and four ST Math objectives; 13 objectives total were aligned to at least one week of Central instruction. Figure 5 displays the percent of play time dedicated to each of the 13 objectives compared to all other objectives played. On average, the 13 objectives represented around 50% of the total play during each week. Within a week, the objective that aligned with that week's curriculum is represented with a dark border. We examined this another way by investigating what percent of play in a given week was on the week's curriculum-aligned objectives (Figure 6). The distribution of Central's curriculum over a number of objectives per week made it difficult to understand how closely ST Math play was aligned with contemporaneous district instruction. We chose two objectives to examine in more detail, Objective 8 (Angles and Triangles) and Objective 31 (Measurement and Conversions). We chose these two objectives because they appeared more than once each in the Central curriculum over the study period, and because they were at

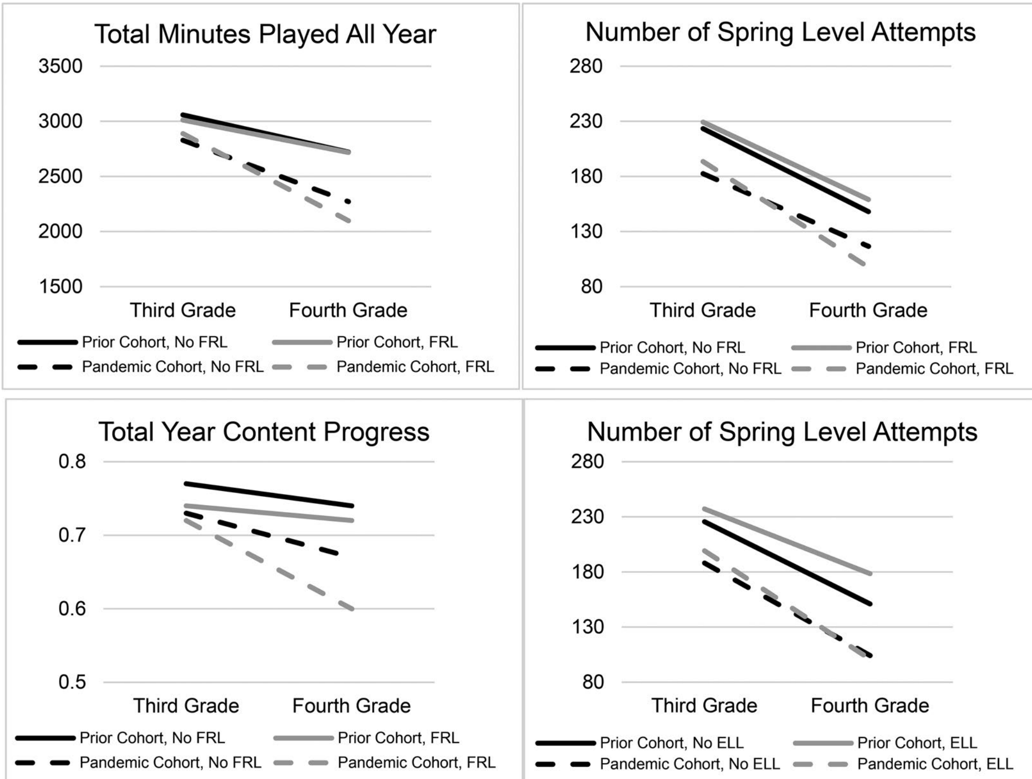

Figure 4. Regression-adjusted results interacted with free/reduced lunch or ELL Status.

Note. Figures drawn from regression-adjusted values based on unstandardized coefficients. Only statistically significant interactions (at the $p < .05$ level) are shown.

different places within the ST Math curriculum—if students were following the prescribed order,

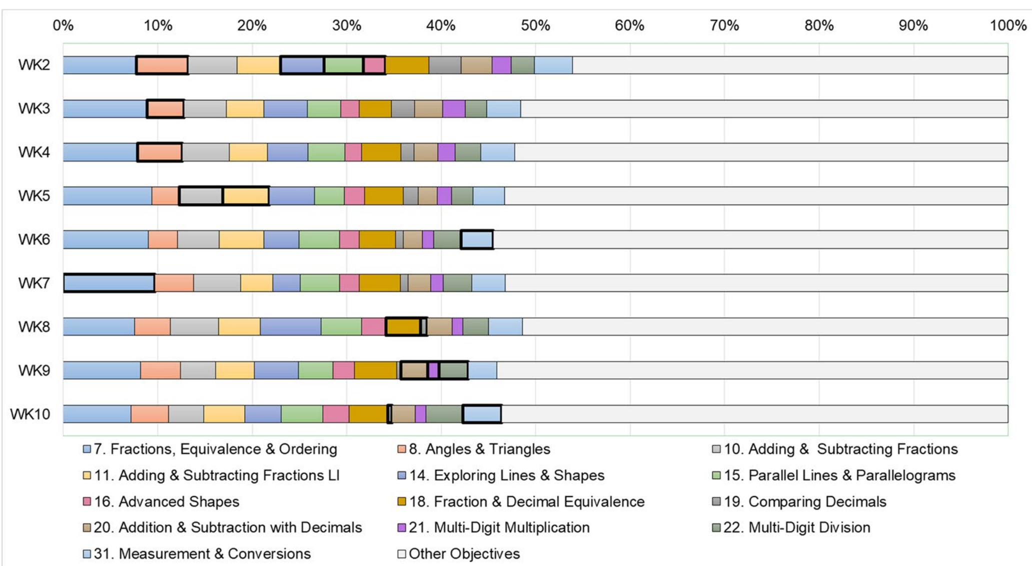

Figure 5. Alignment between ST math objective and other curriculum during pandemic at-home instruction.

Note. Only objectives that were taught at some point during the spring of 2020 are included. Objectives in dark outlines are those that cover content taught in other instruction during that same week.

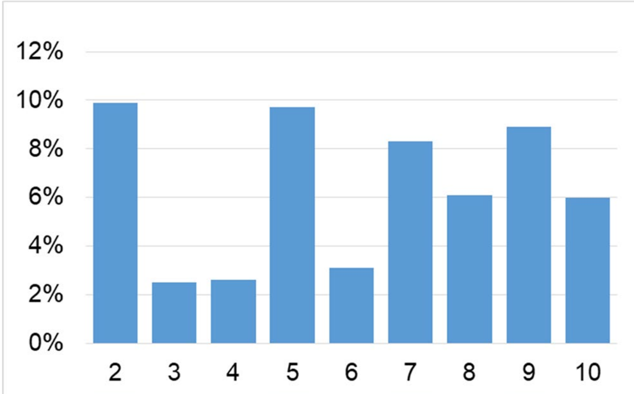

Figure 6. Percent of plays aligned to central curriculum by week.

they would be unlikely to play these objectives closely together. For these analyses, we structured the data by week, calculating a percentage of weekly plays attributed to Objective 8 and 31. Separately by objective, we regressed this percentage on week (an ordinal variable to represent the curriculum order) and a dummy variable indicating whether the content aligned with that objective was taught in that week. If the alignment dummy variable was a statistically significant predictor of percentage of play, it would provide evidence that the content of teacher-led instruction and ST Math play were temporally aligned. We included demographic covariates and a variable indicating the total number of levels played by the student that week. Standard errors were clustered on student to account for each student's contribution of multiple weeks. Temporal alignment was not predictive of percentage of play for either objective. We further examined percent play as a function of week as a series of dummy variables; statistically significant associations would indicate that—on average—students were playing certain objectives more in different weeks. Week was not a statistically significant predictor of how much students played either objective. The table of results is in Appendix C.

Predicting achievement from temporal alignment

Even if students were not more likely to play objectives that were aligned to Central's lessons, they may perform better when playing objectives that were matched to their outside-of-ST Math instruction. To answer this question, we examined student level scores on levels they had not previously passed and compared plays within students to examine whether students performed better on levels when those levels were matched with instruction as compared to their performance of levels that were not matched with instruction (Appendix D). Because objective difficulty likely differed due to the content of the objective, we included a series of dummy variables to represent objective. Match was not predictive of level score (p = .971).

Discussion

Pandemic associations with engagement, performance, and motivation

The COVID-19 pandemic and the resulting school closures during the spring of 2020 presented a threat to student learning and motivation (Domina et al., 2021; Ewing & Cooper, 2021; Smith et al., 2021; Steinmayr et al., 2021). For example, Catalano et al. (2021) found that approximately 30 percent of students of interviewed teachers were not engaged consistently with online coursework. The decision to suspend state-wide achievement testing and most school-based research created a substantial barrier to understanding the extent of this threat (Bailey et al., 2021). Given the online nature of pandemic schooling, many districts continued to use educational technology programs, including ST Math. We used data from ST Math to serve as a substitute assessment for student performance and to provide insight into student engagement and motivation while students completed school activities at home.

Our difference-in-difference models presented estimates that accounted for individual and grade-level factors. Our analyses showed declines in engagement for fourth graders during the pandemic over and above what was expected based on the prior cohort's grade-level decline. This decline was especially pronounced for students who were eligible for free/reduced lunch or who were designated as ELLs. The former were more likely to have stopped playing ST Math when schools closed—given that our total minutes played analysis included all students, regardless of whether they logged in after mid-March, much of the association between pandemic and minutes may be due to this drop-out. Those who were ELLs also showed a steeper decline for the number of level attempts. The level attempts measure was more sensitive to spring effects than was total annual minutes, but these analyses were limited only to those students who had attempted at least one level after the closure. Both measures are likely representative of student ability to access the software. During the pandemic, students may have faced difficulties in accessing online materials due to numerous reasons, including internet speeds, technology access, and competition (i.e., sharing resources with siblings); researchers have identified that these challenges are associated with socioeconomic status (e.g., Bacher-Hicks et al., 2020; Domina et al., 2021). Students designated as ELLs also had more pronounced negative associations between the pandemic and number of level attempts. ELLs may face technology access issues over and above those faced by students eligible for free/reduced lunch—merely having access to technology and Internet may not be enough—ELLs may be more greatly impacted from a divide in use of technology (see Warschauer et al., 2004); there is some evidence that teachers faced challenges during the pandemic supporting their ELL students in working with educational technology (Catalano et al., 2021; Hebert et al., 2020; Walters, 2020).

Performance measures painted a different picture. Those students who did play at least some ST Math during the pandemic spring obtained level and objective posttest scores greater than expected based on trends from third to fourth grade of the prior cohort. This may in-part be a selection effect: those who were retained in the sample of spring 2020 players may have been qualitatively different than those who contributed only to the 2018 and 2019 estimates. There were demographic differences between those who played during the spring and those who didn't, but these differences were small (e.g., 65% eligible for free/reduced lunch for the full sample and 62% for the spring sample) and demographic differences were also seen in other years/cohorts between the full-year sample and the spring sample. Although we included student demographic covariates in our models, the presence of these observed variable differences may indicate unobserved selection effects are driving results. Analyses showing relations between these pandemic players and increased future ST Math or external mathematics assessments may lend credence to these results. Within our own data, future analyses using propensity score or other matching techniques could further limit bias due to selection. If these positive performance associations with pandemic play are a true effect, it may be that flexibility or other aspects of the home environment (e.g., help from parents) improved performance within ST Math. More qualitative work or more detailed accounting on time spent on ST Math could illuminate this result. For example, a more temporally-detailed logging system would permit comparison of time on each puzzle between at-home pandemic play and prior at-school play.

Content progress did not follow the pattern of the other performance measures; fourth graders in the pandemic cohort made less progress than they did in third grade and made less progress than the prior cohort fourth graders. We had conceptualized progress as a variable that represented a combined engagement/performance measure, but had categorized it as performance due its reliance on students passing levels to make progress. In our correlation results, progress was most highly correlated with minutes played, but this may have been due to the fact that both were year-long measures and not limited only to spring data. Progress' correlation with the other engagement measure (level attempts) was similar in magnitude to its correlation with performance measures. In ST Math, content progress is an easily-obtained high-level metric of students' interaction with the platform; this is likely true of other technology-based tutorials. The characterization of progress as an engagement rather than achievement metric may have implications for future research considering such a measure.

As students completed school work from home they were away from the motivating influence of friends and special classroom activities and may not have received as much individualized feedback from teachers; this may have impacted their expectancies and values for mathematics. Both cohorts experienced the expected grade-over-grade decline in mathematics self-beliefs (see Fredricks & Eccles, 2002; Gottfried et al., 2007), but the pandemic cohort experienced a sharper decline. The sources contributing to student self-beliefs (Bandura, 1977, 1997) may have been impacted during the pandemic. For example, without teachers offering verbal persuasion, student self-beliefs may have suffered—parents may not have been able to serve this role while juggling their own work and childcare. The impact of being removed from peers may have hurt self-beliefs in some students as they were unable to witness peer success and gain confidence through vicarious experience; however, this is likely dependent on student relative positioning—social comparisons can also serve to reduce self-beliefs (Müller-Kalthoff et al., 2017). Dimensional comparisons (e.g., Marsh et al., 2015) may have also functioned differently during at-home schooling; some school subjects (e.g., reading) may have been easier to complete without in-class teacher scaffolding, leading to reduced mathematics self-beliefs from between-subject comparative effects. Finally, it is unlikely that the decline in mathematics expectancy was due to mastery experiences or physiological reactions—our performance results indicated that students who played in the spring (and therefore took the survey) experienced more success within ST Math than in prior years, and our emotional cost results indicated that negative mood was not increased. Prior research has shown that use of ST Math improves student mathematics self-beliefs (Rutherford et al., 2020); our results may underestimate the association between the pandemic and reduced expectancy if students over-weighted ST Math in their consideration of math motivation or experienced ST Math-specific amelioration of self-efficacy decline.

Cost was the only value measure that demonstrated statistically significant associations with the pandemic. Emotional cost increased from third to fourth grade in both the pandemic cohort and the prior cohort. This increase was attenuated for the pandemic cohort, however, and the third to fourth-grade increase in cost for this cohort was only a little more than half that of the prior cohort. Prior research has shown that value for mathematics declines across grades in elementary school (e.g., Jacobs et al., 2002), but these declines may not be uniform with respect to aspects of value; for example, utility/importance and interest display different patterns (Wigfield & Eccles, 1994). There is scant research on developmental trends for cost. One exception is Gaspard et al. (2017) who found that cost for mathematics, including emotional cost, increased over the high school years. Outside of an EVT framework, prior research in achievement emotions has found that negative emotions increase with age generally (e.g., Vierhaus et al., 2016) and in mathematics specifically (e.g., Ahmed et al., 2013). Our results are consistent with the expected grade-over-grade increase, that this increase was less during the pandemic indicates students perceived a difference in the school environment. Considering antecedents of academic emotions, Pekrun (2017) specifies that perceptions of control (e.g., expectancies) and value influence emotions. It is unlikely that increased control in the form of positive self-beliefs reduced negative emotions given the negative association between the pandemic and mathematics expectancies. None of the other facets of mathematics value, which all had non-statistically significant associations, seem likely explanatory candidates either. Other aspects of control (e.g., choice of when and how much to work) may be implicated in the switch to online instruction.

Temporal alignment of ST math play with Teacher-Led instruction

We set out to paint a more complete picture of student experiences during spring 2020 with our investigation into alignment between the pandemic curriculum at Central and student engagement with ST Math. The lack of alignment between the teacher-led curriculum and the objectives students played left many unanswered questions regarding the role of educational technology programs like ST Math and how representative of the schooling experience engagement and performance in such programs might be. Based on prior research on ST Math, it is not unusual for curriculum and platform temporal alignment to be low (Peddycord-Liu et al., 2019). Additionally, the self-paced nature of programs like ST Math mean that content distance between students is likely to grow as the year continues. We might have seen greater alignment if at-home instruction had occurred in the fall before student progress had been allowed to spread. Prior research has also shown that teacher management of student pacing through ST Math is variable (Callaghan, 2017; Peddycord-Liu et al., 2019); our results present only averages; it could be that for some classes, alignment was greater. Although our results that alignment was low call into question the value of data from ST Math for providing a window into other concurrent instruction, if ST Math and typical mathematics instruction are both aligned to the same content as assessments, this result does not necessarily invalidate our use of ST Math data as a stand-in for external assessments.

Despite the supposition that temporal alignment between educational technology and classroom curriculum is beneficial for student learning (McCulloch et al., 2018), we did not find evidence that students performed better during weeks where they played objectives that matched teacher-led instruction. These analyses were limited to just the nine weeks of at-home instruction during the pandemic and to those students who logged on after school went online—the limited time-frame and sample may have influenced our results. Experienced teachers of ST Math note the desire for temporal alignment (Peddycord-Liu et al., 2019)—the desired result from alignment may be improvement in ST Math and other similar educational technology platforms, or it may be improved performance on classroom assessments. We were not able to assess the latter.

Limitations

The data gathered from educational technology like ST Math helps to address the data desert created by the suspension of testing during the pandemic. However, these data are not without limitations. One important limitation was revealed in analyzing temporal alignment between ST Math and teacher-led instruction: namely, educational technology may be a poor representation of the totality of students' learning, during a pandemic or at other times. Nevertheless, the relation between educational technology data and more traditional measures of achievement give credence to the idea that it holds at least some value.

Another caveat arises: as students interact with educational technology, their performance partially relies on their skill or comfort with the mechanics of the gaming environment itself (Maertens et al., 2015). This is true of ST Math—prior research has shown that difficulty of ST Math games appears to be a feature of both math content and mechanics (Akintunde et al., 2020). In this way, educational technology data are representative of both less than and more than traditional learning metrics.

Our conclusions are also limited by the changes in our samples across measures. As students dropped out of the learning stream, they also dropped out of some of our outcomes, especially objective posttests and motivation measures. We chose to allow the sample to vary to paint a picture of the experiences of each group—those whose access was limited and those who were able to engage at home. Caution should be used in generalizing results from the more limited groups to the larger group of students.

Further, some cautions should be exercised due to the nature of our analyses. If we consider a conservative statistical significance threshold, our results regarding level attempts and cost value should be subject to particular scrutiny; future replications will be needed to determine how robust these inferences are to potential sources of bias (and type 1 error) with other students and at other points in time. In addition, the marginal R^2 values (for the variance explained by the fixed effects) for many of our models was on the small side, although we think R^2 value were still notable given the absence of many covariates or a prior measure for the outcome; these may be important to interrogate in future work. It may be more important to situate our effect sizes, some of which can be benchmarked against prior research. For example, the achievement effect sizes can be benchmarked against studies using achievement measures. According to Kraft (2020), achievement effect sizes between 0.05 and 0.20 can be considered "medium." Our level score effect sizes of 0.13 and posttest score effect size of 0.11 fall squarely in this range. Similarly, the expectancy effect size of −0.16 can be benchmarked against studies involving the oft-cited gender difference in self-beliefs; which was .18 from a meta-analysis of self-efficacy (Huang, 2013). We are less able to benchmark our engagement measures or our cost measure, as the context for comparison is muddier.

Implications for practice

As the COVID-19 pandemic continues and schools extend at-home instruction into the 2020-2021 school year, communities are grappling with the best way to serve students. Our work demonstrates that there are disparities in engagement with at-home instruction, and these disparities are correlated with marginalized status, such as eligibility for free/reduced lunch. Among students who were able to access ST Math, performance did not suffer; this demonstrates that the use of digital tools during online instruction has potential. To alleviate potential learning loss during this pandemic or during times of emergency schooling in the future, it will be key to ensure that all students have the access necessary to engage fully.

With statewide standardized tests and other forms of school-based assessments suspended, schools may wish to leverage data from educational technology to monitor student performance and engagement. We set out to determine whether data from educational technology, such as ST Math, could serve as a proxy. Our results are equivocal. Although the data from ST Math provided insights into students' engagement, achievement, and motivation, these insights were largely limited to those students who meaningfully engaged with the platform after the switch to online schooling. Because marginalized status was associated with drop-out, reliance on such data may leave out students for which schools may most want information. Nevertheless, our work presents a model for how to operationalize and analyze platform data.

Lastly, although our alignment measure did not predict performance, schools may wish to consider how educational technology fits within the complete package of instruction provided to students. Greater temporal alignment may make both use of and data from educational technology programs more meaningful to students and teachers.

Implications for theory and research

Our work contributes to the growing research area regarding how log data from contexts, including educational technology, can be used to make inferences about student learning, engagement, and performance (Papamitsiou & Economides, 2014; Teasley, 2019). Within this study, our operationalizations of engagement and performance were interrelated in expected ways and had meaningful associations with grade-level and with the shift to emergency online instruction during the pandemic. As researchers seek ways to utilize data from student interactions with programs such as ST Math, these operationalizations may prove beneficial. More specifically, our research demonstrates the potential and limitations of using log data to estimate effects from interruptions to traditional classroom instruction.

Our work also contributes to understanding of EVT as situated within online schooling during the pandemic. Eccles and Wigfield (2020) have recently renamed EVT to situated expectancy-value theory (SEVT) in order to stress the situative nature of motivation. Within SEVT, the people and experiences with which individuals interact influence their expectancies and values. At-home schooling during the pandemic presents a unique situation within which to study expectancies and values. Our work links the pandemic with a decline in expectancies despite an increase in performance, coupled with a reduction in emotional cost in the form of negative emotions. As discussed above, specific features of the at-home schooling environment may contribute to these results. Future work can further explicate the processes by which motivation is impacted.

Conclusion

As schools across the U.S. closed in response to the COVID-19 pandemic, many students moved online for emergency at-home instruction. Analogous situations from prior research presented a dire picture of how students, especially those from marginalized communities, might be impacted (Dynarski, 2020; Kuhfeld & Tarasawa, 2020). With the suspension of state-wide achievement tests and other school-based assessments, assessing these impacts became a challenge. Leveraging data collected from within an educational technology platform, ST Math, offered a partial solution. We were able to estimate associations between the pandemic and measures of engagement, performance, and motivation for some students; however, our conclusions were limited by substantial (30%) drop-out from any platform use after the shift to online schooling. We found that even students who did continue using the platform had lower engagement with ST Math during the pandemic, but among those students, performance within the software increased. Changes in student motivation were also associated with the pandemic: students had lower mathematics expectancy, but also lower emotional cost for mathematics. To broaden our understanding of student experiences, we collected public lesson plans from district websites to examine temporal alignment between these plans and student engagement with ST Math. We found little alignment and no performance benefit from alignment, indicating that student experiences with ST Math during the pandemic may stand apart from other pandemic instruction. Our results illustrate the potential and pitfalls of using educational technology data in lieu of more traditional assessments and draw attention to issues of access and motivation during at-home schooling.

Notes on contributors

Teomara (Teya) Rutherford is a faculty member in the Learning Sciences program within the University of Delaware School of Education. She holds a PhD in Learning, Cognition, and Development from the University of California, Irvine and a Juris Doctor from Boston University School of Law. Her research focuses on learning and motivation in digital environments. Twitter: @DrTeyaR

Kerry Duck is a postdoctoral researcher at the University of Delaware. He completed his Ph.D. in Educational Psychology at the University of Northern Colorado. His research interests include student and teacher motivation within STEM disciplines. Additionally, he has methodological interests within the use of Ecological Momentary Assessments/Interventions to facilitate student motivation and learning. ORCID id:https://orcid.org/0000-0002-3602-5205 Twitter: @DrKerryDuck Linkedin: www.linkedin.com/in/KerryDuckPhD

Joshua M. Rosenberg (PhD, Michigan State University) is an assistant professor of STEM education and faculty fellow at the Center for Enhancing Education in Mathematics and Sciences at the University of Tennessee, Knoxville. His research focuses on how learners think of and with data, particularly in science education settings. Twitter: @JRosenberg6432

Ray Patt is a doctoral student at the University of Delaware. He received his B.S. in Neuroscience and Linguistics from The Ohio State University. He is broadly interested in how children learn in digital environments, and specifically interested in the role language plays in learning.Twitter: @RayPatt

Notes

1. pseudonym
2. For the average level outcome, the estimation did not converge using the default optimizer. We chose to use the optimx package optimizer (Nash, 2014) for this outcome, which led to the estimation converging (and yielded identical effects as those from when we used the default optimizer).
3. We calculated the Boneferroni-adjusted p-value by dividing the conventional alpha value of .05 by the number of outcomes (10) to arrive at .005.

Disclosure statement

No potential conflcit of interest was reported by the authors.

Funding

This work was supported by the National Science Foundation under Grant XX and Grant XX [grant details redacted for review].

ORCID

Kerry Duck http://orcid.org/0000-0002-3602-5205

Data availability statement

Those interested in the data associated with this research should contact [author redacted] at [contact redacted].

References

Admiraal, W., Vermeulen, J., & Bulterman-Bos, J. (2020). Teaching with learning analytics: how to connect computer-based assessment data with classroom instruction?*Technology, Pedagogy and Education*, 1–15.

Agostinelli, F., Doepke, M., Sorrenti, G., Zilibotti, F. (2020). *When the Great Equalizer Shuts Down: Schools, Peers, and Parents in Pandemic Time* (NBER Working Paper No. 28264). National Bureau of Economic Research. https://www.nber.org/papers/w28264

Agudo-Peregrina, Á. F., Iglesias-Pradas, S., Conde-González, M. Á., & Hernández-García, Á. (2014). Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning. *Computers in Human Behavior*, *31*, 542–550. https://doi.org/10.1016/j.chb.2013.05.031

Ahmed, W., Van der Werf, G., Kuyper, H., & Minnaert, A. (2013). Emotions, self-regulated learning, and achievement in mathematics: A growth curve analysis. *Journal of Educational Psychology*, *105*(1), 150–161. https://doi.org/10.1037/a0030160

Amador, J. M., Wallin, A., Keehr, J., & Chilton, C. (2020). Supporting children's mathematical understanding through a hyperlinked book of mathematics games. In R. E. Ferding, E. Baumgartner, R. Hartshorne, R. Kaplan-Rakowski, & C. Mouza (Eds.), *Teaching, technology, and teacher education during the COVID-19 pandemic: Stories from the field* (pp. 655–660) AACE-Association for the Advancement of Computing in Education.

Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology*, *80*(3), 260–267. https://doi.org/10.1037/0022-0663.80.3.260

Anderman, E. M., Anderman, L. H., Yough, M. S., & Gimbert, B. G. (2010). Value-added models of assessment: Implications for motivation and accountability. *Educational Psychologist*, *45*(2), 123–137. https://doi.org/10.1080/00461521003703045

Anderson-Pence, K. L., Tygret, J. A., & Crocker, L. (2020). Leveraging tutorial instructional software to enhance classroom mathematical discussions: An exploratory mixed-methods study. *Educational Research: Theory and Practice*, *31*(3), 74–93.

Andrew, A., Cattan, S., Costa Dias, M., Farquharson, C., Kraftman, L., Krutikova, S., Phimister, A., & Sevilla, A. (2020). Inequalities in children's experiences of home learning during the COVID-19 lockdown in England. *Fiscal Studies*, *41*(3), 653–683. https://doi.org/10.1111/1475-5890.12240

Bacher-Hicks, A., Goodman, J., Mulhern, C. (2020). Inequality in household adaptation to schooling shocks: Covid-induced online learning engagement in real time *(No. w27555)*. National Bureau of Economic Research. https://www.nber.org/papers/w27555

Baek, Y. K. (2008). What hinders teachers in using computer and video games in the classroom? Exploring factors inhibiting the uptake of computer and video games. *Cyberpsychology & Behavior: The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society*, *11*(6), 665–671. https://doi.org/10.1089/cpb.2008.0127

Bailey, D. H., Duncan, G. J., Murnane, R. J., & Yeung, N. A. (2021). *Achievement Gaps in the Wake of COVID-19 (Ed WorkingPaper No. 21-346)*. Annenberg Institute at Brown University. https://doi.org/10.26300/y7pp-sa02

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

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Freeman.

Baroody, A. E., Rimm-Kaufman, S. E., Larsen, R. A., & Curby, T. W. (2016). A multi-method approach for describing the contributions of student engagement on fifth grade students' social competence and achievement in mathematics. *Learning and Individual Differences*, *48*, 54–60. https://doi.org/10.1016/j.lindif.2016.02.012

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bergin, D. A. (1999). Influences on classroom interest. *Educational Psychologist*, *34*(2), 87–98. https://doi.org/10.1207/s15326985ep3402_2

Bong, M., Cho, C., Ahn, H. S., & Kim, H. J. (2012). Comparison of self-beliefs for predicting student motivation and achievement. *The Journal of Educational Research*, *105*(5), 336–352. https://doi.org/10.1080/00220671.2011.627401

Brezovszky, B., McMullen, J., Veermans, K., Hannula-Sormunen, M. M., Rodríguez-Aflecht, G., Pongsakdi, N., Laakkonen, E., & Lehtinen, E. (2019). Effects of a mathematics game-based learning environment on primary school students' adaptive number knowledge. *Computers & Education*, *128*, 63–74. https://doi.org/10.1016/j.compedu.2018.09.011

Bullock, E. P., Moyer-Packenham, P., Shumway, J. F., MacDonald, B., Watts, C. (2015). Effective teaching with technology: Managing affordances in iPad apps to promote young children's mathematics learning. In *Society for Information Technology & Teacher Education International Conference* (pp. 2648–2655). Association for the Advancement of Computing in Education (AACE).

Callaghan, M., Long, J. J., vanEs, E. A., Reich, S., & Rutherford, T. (2017). How teachers integrate a math computer game: Professional development use, teaching practices, and effects on student achievement. *Journal of Computer Assisted Learning*. https://doi.org/10.1111/jcal.12209

Catalano, A. J., Torff, B., & Anderson, K. S. (2021). Transitioning to online learning during the COVID-19 pandemic: differences in access and participation among students in disadvantaged school districts. *The International Journal of Information and Learning Technology*, *38*(2), 258–270. https://doi.org/10.1108/IJILT-06-2020-0111

Cheung, A. C., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review*, *9*, 88–113. https://doi.org/10.1016/j.edurev.2013.01.001

Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of Educational Research*, *86*(1), 79–122. https://doi.org/10.3102/0034654315582065

Clements, D. H., & Sarama, J. (2004). Learning trajectories in mathematics education. *Mathematical Thinking and Learning*, *6*(2), 81–89. https://doi.org/10.1207/s15327833mtl0602_1

De Grove, F., Bourgonjon, J., & Van Looy, J. (2012). Digital games in the classroom? A contextual approach to teachers' adoption intention of digital games in formal education. *Computers in Human Behavior*, *28*(6), 2023–2033. https://doi.org/10.1016/j.chb.2012.05.021

Denissen, J. J. A., Zarrett, N. R., & Eccles, J. S. (2007). I like to do it, I'm able, and I know I am: longitudinal couplings between domain-specific achievement, self-concept, and interest. *Child Development*, *78*(2), 430–447. https://doi.org/10.1111/j.1467-8624.2007.01007.x

DiCerbo, K. E., Xu, Y., Levy, R., Lai, E., & Holland, L. (2017). Modeling student cognition in digital and nondigital assessment environments. *Educational Assessment*, *22*(4), 275–297. https://doi.org/10.1080/10627197.2017.1382343

Domina, T., Renzulli, L., Murray, B., Garza, A. N., & Perez, L. (2021). Remote or removed: Predicting successful engagement with online learning during COVID-19. *Socius: Sociological Research for a Dynamic World*, *7*, 237802312098820–237802312098815. https://doi.org/10.1177/2378023120988200

Dotterer, A. M., & Lowe, K. (2011). Classroom context, school engagement, and academic achievement in early adolescence. *Journal of Youth and Adolescence*, *40*(12), 1649–1660. https://doi.org/10.1007/s10964-011-9647-5

Dynarski, S. (2020). The school year really ended in March. New York Times. https://www.nytimes.com/2020/05/07/business/school-education-online-money.html

Eccles, J., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J., & Midgley, C. (1983). Expectancies, values and academic behaviors. In J. T. Spence (Ed.), *Achievement and Achievement Motives*. 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*, 101859. https://doi.org/10.1016/j.cedpsych.2020.101859

Ewing, L. A., & Cooper, H. B. (2021). Technology-enabled remote learning during COVID-19: perspectives of Australian teachers, students and parents. *Technology, Pedagogy and Education*, 1–17.

Fan, W. (2011). Social influences, school motivation and gender differences: an application of the expectancy-value theory. *Educational Psychology*, *31*(2), 157–175. https://doi.org/10.1080/01443410.2010.536525

Filsecker, M., & Hickey, D. T. (2014). A multilevel analysis of the effects of external rewards on elementary students' motivation, engagement and learning in an educational game. *Computers & Education*, *75*, 136–148. https://doi.org/10.1016/j.compedu.2014.02.008

Frank, K. A., Maroulis, S. J., Duong, M. Q., & Kelcey, B. M. (2013). What would it take to change an inference? Using Rubin's causal model to interpret the robustness of causal inferences. *Educational Evaluation and Policy Analysis*, *35*(4), 437–460. https://doi.org/10.3102/0162373713493129

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59–109. https://doi.org/10.3102/00346543074001059
Fredricks, J. A., & Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: Growth trajectories in two male sex-typed domains. *Developmental Psychology*, *38*(4), 519–533. https://doi.org/10.1037/0012-1649.38.4.519
Garbe, A., Ogurlu, U., Logan, N., & Cook, P. (2020). Parents' experiences with remote education during COVID-19 school closures. *American Journal of Qualitative Research*, *4*(3), 45–65.
Gaspard, H., Häfner, I., Parrisius, C., Trautwein, U., & Nagengast, B. (2017). Assessing task values in five subjects during secondary school: Measurement structure and mean level differences across grade level, gender, and academic subject. *Contemporary Educational Psychology*, *48*, 67–84. https://doi.org/10.1016/j.cedpsych.2016.09.003
Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.
Gewertz, C. (2020). It's official: All states have been excused from statewide testing this year. Education Week. https://blogs.edweek.org/edweek/campaign-k-12/2020/04/coronvirus-no-state-testing-essa.html
Ghergulescu, I., & Muntean, C. H. (2016). ToTCompute: A novel EEG-based TimeOnTask threshold computation mechanism for engagement modelling and monitoring. *International Journal of Artificial Intelligence in Education*, *26*(3), 821–854. https://doi.org/10.1007/s40593-016-0111-2
Gottfried, A. E., Marcoulides, G. A., Gottfried, A. W., Oliver, P. H., & Guerin, D. W. (2007). Multivariate latent change modeling of developmental decline in academic intrinsic math motivation and achievement: Childhood through adolescence. *International Journal of Behavioral Development*, *31*(4), 317–327. https://doi.org/10.1177/0165025407077752
Greenhow, C., Lewin, C., & Staudt Willet, K. B. (2020). The educational response to Covid-19 across two countries: a critical examination of initial digital pedagogy adoption. *Technology, Pedagogy and Education*, 1–19.
Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. [Database] https://doi.org/10.3102/003465430298487
Hebert, M., Goodrich, J. M., Namkung, J. M. (2020). *Characterizing Remote Instruction Provided by Elementary School Teachers during School Closures due to COVID-19*. https://edarxiv.org/vsx4q/download
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. https://doi.org/10.1177/0735633117748416
Hough, H., Kalogrides, D., & Loeb, S. (2017). Using surveys of students' social-emotional learning and school climate for accountability and continuous improvement. *Policy Analysis for California Education, PACE*. https://edpolicyinca.org/publications/using-surveys-students-social-emotional-skills-and-school-climate-accountability-and
Huang, C. (2013). Gender differences in academic self-efficacy: A meta-analysis. *European Journal of Psychology of Education*, *28*(1), 1–35. https://doi.org/10.1007/s10212-011-0097-y
Hwang, G.-J., & Lai, C.-L. (2017). Facilitating and bridging out-of-class and in-class learning: An interactive E-bookbased flipped learning approach for Math courses. *Educational Technology & Society*, *20*(1), 184–197.
Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development*, *73*(2), 509–527. https://doi.org/10.1111/1467-8624.00421
Kaffenberger, M. (2020). Modelling the long-run learning impact of the Covid-19 learning shock: Actions to (more than) mitigate loss. *International Journal of Educational Development*, 81, 102326.
Kosko, K. W. (2017). Effects of student-reported gameplay strategy related to growth in multiplicative reasoning. *Electronic Journal of Mathematics & Technology*, *11*(3), 184–193.
Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798
Kuhfeld, M., Tarasawa, B. (2020). The COVID-19 slide: What summer learning loss can tell us about the potential impact of school closures on student academic achievement. Northwest Evaluation Association. https://www.nwea.org/content/uploads/2020/05/Collaborative-Brief_Covid19-Slide-APR20.pdf
Lai, J., & Widmar, N. O. (2020). Revisiting the Digital Divide in the COVID-19 Era. *Applied Economic Perspectives and Policy*, 43(1), 458–464.
Lazarides, R., Gaspard, H., & Dicke, A. L. (2019). Dynamics of classroom motivation: Teacher enthusiasm and the development of math interest and teacher support. *Learning and Instruction*, *60*, 126–137. https://doi.org/10.1016/j.learninstruc.2018.01.012
Linnenbrink-Garcia, L., & Patall, E. (2016). Motivation. In L. Corno & E. M. Anderman (Eds.), *Handbook of educational psychology* (3rd ed., pp. 91–103). Routledge.
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?. *International Educational Data Mining Society*.
Lüdecke, D., Makowski, D., Waggoner, P., & Patil, I. (2020). Assessment of regression models performance. *CRAN. R Package Version*, *0.4*, *8*. https://cran.r-project.org/web/packages/performance/index.html
Maertens, M., Vandewaetere, M., Cornillie, F., & Desmet, P. (2015). Performance in educational math games: Is it a question of math knowledge? In *Describing and studying domain-specific serious games* (pp. 117–131). Springer.

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

Marshall, M. A., & Brown, J. D. (2004). Expectations and realizations: The role of expectancies in achievement settings. *Motivation and Emotion*, *28*(4), 347–361. https://doi.org/10.1007/s11031-004-2388-y

McCulloch, A. W., Hollebrands, K., Lee, H., Harrison, T., & Mutlu, A. (2018). Factors that influence secondary mathematics teachers' integration of technology in mathematics lessons. *Computers & Education*, *123*, 26–40. https://doi.org/10.1016/j.compedu.2018.04.008

MIND Research Institute. (2015). *ST Math scope and sequence with standards alignment for fourth grade Texas 2015-2016 edition*.

Moyer-Packenham, P. S., Bullock, E. K., Shumway, J. F., Tucker, S. I., Watts, C. M., Westenskow, A., Anderson-Pence, K. L., Maahs-Fladung, C., Boyer-Thurgood, J., Gulkilik, H., & Jordan, K. (2016). The role of affordances in children's learning performance and efficiency when using virtual manipulative mathematics touch-screen apps. *Mathematics Education Research Journal*, *28*(1), 79–105. https://doi.org/10.1007/s13394-015-0161-z

Moyer-Packenham, P. S., Lommatsch, C. W., Litster, K., Ashby, J., Bullock, E. K., Roxburgh, A. L., Shumway, J. F., Speed, E., Covington, B., Hartmann, C., Clarke-Midura, J., Skaria, J., Westenskow, A., MacDonald, B., Symanzik, J., & Jordan, K. (2019). How design features in digital math games support learning and mathematics connections. *Computers in Human Behavior*, *91*, 316–332. https://doi.org/10.1016/j.chb.2018.09.036

Müller-Kalthoff, H., Helm, F., & Möller, J. (2017). The big three of comparative judgment: On the effects of social, temporal, and dimensional comparisons on academic self-concept. *Social Psychology of Education*, *20*(4), 849–873. https://doi.org/10.1007/s11218-017-9395-9

Nakagawa, S., Johnson, P. C., & Schielzeth, H. (2017). The coefficient of determination R 2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, *14*(134), 20170213. https://doi.org/10.1098/rsif.2017.0213

Nash, J. C. (2014). On best practice optimization methods in R. *Journal of Statistical Software*, *60*(2), 1–14. https://doi.org/10.18637/jss.v060.i02

Ostrow, K. S., & Heffernan, N. T. (2018). Testing the validity and reliability of intrinsic motivation inventory subscales within ASSISTments. In *International conference on artificial intelligence in education* (pp. 381–394). Springer.

Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014). Effectiveness of cognitive tutor algebra I at scale. *Educational Evaluation and Policy Analysis*, *36*(2), 127–144. https://doi.org/10.3102/0162373713507480

Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Journal of Educational Technology & Society*, *17*(4), 49–64.

Pardos, Z. A., Baker, R. S., San Pedro, M. O., Gowda, S. M., & Gowda, S. M. (2014). Affective states and state tests: Investigating how affect and engagement during the school year predict end-of-year learning outcomes. *Journal of Learning Analytics*, *1*(1), 107–128. https://doi.org/10.18608/jla.2014.11.6

Parolin, Z., & Lee, E. K. (2021). Large socio-economic, geographic and demographic disparities exist in exposure to school closures. Nature Human Behaviour, 5(4), 522–528.

Peddycord-Liu, Z., Cody, C., Kessler, S. M., Barnes, T., Lynch, C., & Rutherford, T. (2017). Using serious game analytics to inform digital curricular sequencing: What math objective should students play next? In *Proceedings of the ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play (CHI PLAY)*. Amsterdam, Netherlands.

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 *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Glasgow, UK.

Pekrun, R. (2017). Emotion and achievement during adolescence. *Child Development Perspectives*, *11*(3), 215–221. https://doi.org/10.1111/cdep.12237

Pellegrini, M., Lake, C., Neitzel, A., Slavin, R. E. (2020). Effective programs in elementary mathematics: A Meta-Analysis. Best Evidence Encyclopedia. http://www.bestevidence.org/word/elem_math_May_26_2020_full.pdf

Peterson, L., Scharber, C., Thuesen, A., & Baskin, K. (2020). A rapid response to COVID-19: one district's pivot from technology integration to distance learning. *Information and Learning Sciences*, *121*(5/6), 461–469. https://doi.org/10.1108/ILS-04-2020-0131

Putwain, D. W., Nicholson, L. J., Pekrun, R., Becker, S., & Symes, W. (2019). Expectancy of success, attainment value, engagement, and Achievement: A moderated mediation analysis. *Learning and Instruction*, *60*, 117–125. https://doi.org/10.1016/j.learninstruc.2018.11.005

R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

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

Resnick, L. B., Rothman, R., Slattery, J. B., & Vranek, J. L. (2004). Benchmarking and alignment of standards and testing. *Educational Assessment*, *9*(1), 1–27. https://doi.org/10.1207/s15326977ea0901&2_1

Ritter, S., Joshi, A., Fancsali, S., & Nixon, T. (2013). Predicting standardized test scores from Cognitive Tutor interactions. *In S. K. D'Mello, R. A. Calvo, & A. Olney (Eds.) Educational Data Mining, 2013* (pp. 169–176). International Educational Data Mining Society.

Roschelle, J., Feng, M., Murphy, R. F., & Mason, C. A. (2016). Online mathematics homework increases student achievement. *AERA Open*, *2*(4), 233285841667396–233285841667312. https://doi.org/10.1177/2332858416673968

Rutherford, T., Liu, A. S., Lam, A. S., & Schenke, K. (2020). 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., Farkas, G., Duncan, G., Burchinal, M., Kibrick, M., Graham, J., Richland, L., Tran, N., Schneider, S., Duran, L., & Martinez, M. E. (2014). A randomized trial of an elementary school mathematics software intervention: Spatial-temporal math. *Journal of Research on Educational Effectiveness*, *7*(4), 358–383.

Ruzek, E. A., Domina, T., Conley, A. M., Duncan, G. J., & Karabenick, S. A. (2015). Using value-added models to measure teacher effects on students' motivation and achievement. *The Journal of Early Adolescence*, *35*(5–6), 852–882. https://doi.org/10.1177/0272431614525260

San Pedro, M. O. Z., Snow, E. L., Baker, R. S., McNamara, D. S., Heffernan, N. T. (2015). Exploring dynamical assessments of affect, behavior, and cognition and math state test achievement. In *Proceedings of the 8th International Educational Data Mining Society* (pp. 85–92).

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.

Schiefele, U., & Schaffner, E. (2015). Teacher interests, mastery goals, and self-efficacy as predictors of instructional practices and student motivation. *Contemporary Educational Psychology*, *42*, 159–171. https://doi.org/10.1016/j.cedpsych.2015.06.005

Schunk, D. H., Meece, J. R., & Pintrich, P. R. (2012). *Motivation in education: Theory, research, and applications*. Pearson Higher Ed.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Simpkins, S. D., Davis-Kean, P. E., & Eccles, J. S. (2006). Math and science motivation: A longitudinal examination of the links between choices and beliefs. *Developmental Psychology*, *42*(1), 70–83. https://doi.org/10.1037/0012-1649.42.1.70

Smith, J., Guimond, F. A., Bergeron, J., St-Amand, J., Fitzpatrick, C., & Gagnon, M. (2021). Changes in students' achievement motivation in the context of the COVID-19 pandemic: A function of extraversion/introversion?*Education Sciences*, *11*(1), 30–38. https://doi.org/10.3390/educsci11010030

Steinmayr, R., Lazarides, R., Weidinger, A. F., & Christiansen, H. (2021). Teaching and learning during the first COVID-19 school lockdown: Realization and associations with parent-perceived students' academic outcomes. *Zeitschrift für Pädagogische Psychologie*, 35(2–3), 85–106.

Sugarman, J., Lazarín, M. (2020). Educating English Learners during the COVID-19 Pandemic (9-2020 Policy Brief). Migration Policy Institute. https://www.migrationpolicy.org/sites/default/files/publications/mpi-english-learners-covid-19-final.pdf

Syal, S., & Nietfeld, J. L. (2020). The impact of trace data and motivational self-reports in a game-based learning environment. *Computers & Education*, *157*, 103978. https://doi.org/10.1016/j.compedu.2020.103978

Teasley, S. D. (2019). Learning analytics: where information science and the learning sciences meet. *Information and Learning Sciences*, *120*(1/2), 59–73. https://doi.org/10.1108/ILS-06-2018-0045

Trautwein, U., Marsh, H. W., Nagengast, B., Lüdtke, O., Nagy, G., & Jonkmann, K. (2012). Probing for the multiplicative term in modern expectancy–value theory: A latent interaction modeling study. *Journal of Educational Psychology*, *104*(3), 763–777. https://doi.org/10.1037/a0027470

Trust, T., & Whalen, J. (2020). Should teachers be trained in emergency remote teaching? Lessons learned from the COVID-19 pandemic. *Journal of Technology and Teacher Education*, *28*(2), 189–199.

Umer, R., Mathrani, A., Susnjak, T., & Lim, S. (2019). Mining activity log data to predict student's outcome in a course. In *Proceedings of the 2019 International Conference on Big Data and Education* (pp. 52–58). Association for Computing Machinery.

Urdan, T., & Schoenfelder, E. (2006). Classroom effects on student motivation: Goal structures, social relationships, and competence beliefs. *Journal of School Psychology*, *44*(5), 331–349. https://doi.org/10.1016/j.jsp.2006.04.003

Vierhaus, M., Lohaus, A., & Wild, E. (2016). The development of achievement emotions and coping/emotion regulation from primary to secondary school. *Learning and Instruction*, *42*, 12–21. https://doi.org/10.1016/j.learninstruc.2015.11.002

Walters, A. (2020). Inequities in access to education: Lessons from the COVID-19 pandemic. *The Brown University Child and Adolescent Behavior Letter*, *36*(8), 8. https://doi.org/10.1002/cbl.30483

Wang, H., & Woodworth, K. (2011). Evaluation of rocketship education's use of dreambox learning's online mathematics program. *Center for Education Policy*. SRI International.

Warschauer, M., Knobel, M., & Stone, L. (2004). Technology and equity in schooling: Deconstructing the digital divide. *Educational Policy*, *18*(4), 562–588. https://doi.org/10.1177/0895904804266469

Watson, W., & Yang, S. (2016). Games in schools: Teachers' perceptions of barriers to game-based learning. *Journal of Interactive Learning Research*, *27*(2), 153–170.

Wendt, S., Rice, J., Nakamoto, J. (2019). A Cross-State Evaluation of MIND Research Institute's ST Math Program and Math Performance. *WestEd*. https://files.eric.ed.gov/fulltext/ED601907.pdf

West, B. T., Welch, K. B., & Galecki, A. T. (2014). *Linear mixed models: a practical guide using statistical software*. CRC Press.

Wigfield, A., & Eccles, J. S. (1994). Children's competence beliefs, achievement values, and general self-esteem: Change across elementary and middle school. *The Journal of Early Adolescence*, *14*(2), 107–138. https://doi.org/10.1177/027243169401400203

Wigfield, A., & Eccles, J. S. (2002). The development of competence beliefs, expectancies for success, and achievement values from childhood through adolescence. In *Development of achievement motivation* (pp. 91–120). Academic Press.

Wing, C., Simon, K., & Bello-Gomez, R. A. (2018). Designing difference in difference studies: best practices for public health policy research. *Annual Review of Public Health*, *39*, 453–469. https://doi.org/10.1146/annurev-publhealth-040617-013507

Xie, K., Di Tosto, G., Chen, S. B., & Vongkulluksn, V. W. (2018). A systematic review of design and technology components of educational digital resources. *Computers & Education*, *127*, 90–106. https://doi.org/10.1016/j.compedu.2018.08.011

Xu, D., & Jaggars, S. S. (2014). Performance gaps between online and face-to-face courses: Differences across types of students and academic subject areas. *The Journal of Higher Education*, *85*(5), 633–659. https://doi.org/10.1353/jhe.2014.0028

Yuan, T., Chae, H. S., Nantwi, G., & Garg, P. (2010). An examination of 300 iPhone math applications: Implications for the development of an educational math game. In *EdMedia + innovate learning* (pp. 4083–4088). Association for the Advancement of Computing in Education (AACE).

Appendix A

Moderator Results

A. Free/Reduced Priced Lunch Eligibility Moderator Models, Engagement and Performance

	(1)			(2)			(3)			(4)			(5)		
	Minutes Played			Level Attempts			Content Progress			Level Score Ave			Posttest Ave		
Predictors	B	SE	*p*	B	SE	*p*	B	SE	*p*	B	SE	*p*	B	SE	*p*
Grade 4	**−337.55**	54.10	**<0.001**	**−75.38**	7.14	**<0.001**	**−0.03**	0.01	**0.023**	**0.02**	0.01	**0.009**	<0.01	0.01	0.625
Pandemic Cohort	**−229.36**	35.98	**<0.001**	**−40.78**	5.53	**<0.001**	**−0.04**	0.01	**0.001**	<0.01	0.01	0.583	<0.01	0.01	0.597
Gr.4 BY Pandemic	**−219.61**	47.95	**<0.001**	9.40	7.95	0.237	**−0.03**	0.01	**0.009**	0.01	0.01	0.126	0.02	0.01	0.083
FRL	−46.70	31.71	0.141	5.89	4.88	0.228	**−0.03**	0.01	**0.003**	<0.01	0.01	0.521	**−0.02**	0.01	**0.001**
FRL BY Gr. 4	42.96	40.37	0.287	5.18	6.54	0.428	0.01	0.01	0.495	−0.01	0.01	0.303	<0.01	0.01	0.968
FRL BY Pandemic	**106.41**	42.89	**0.013**	5.09	6.62	0.442	0.02	0.01	0.101	−0.01	0.01	0.533	<0.01	0.01	0.956
FRL BY Gr. 4 BY Pandemic	**−275.86**	55.00	**<0.001**	**−35.18**	9.32	**<0.001**	**−0.07**	0.01	**<0.001**	0.01	0.01	0.609	<0.01	0.01	0.846
Hispanic	−31.97	23.06	0.166	4.41	3.58	0.218	**−0.03**	0.01	**0.001**	**−0.01**	<0.01	**0.017**	−0.01	0.01	0.073
Black/African Amer.	31.21	22.15	0.159	**9.74**	3.43	**0.004**	**−0.04**	0.01	**<0.001**	**−0.03**	<0.01	**<0.001**	**−0.03**	0.01	**<0.001**
Two or More Races	−25.06	31.38	0.424	−4.96	4.86	0.307	−0.02	0.01	0.136	0.01	0.01	0.433	<0.01	0.01	0.905
Other Race	13.25	38.45	0.73	−8.46	5.98	0.157	**0.03**	0.02	**0.037**	−0.01	0.01	0.214	0.01	0.01	0.335
Boy	**−118.22**	15.33	**<0.001**	−3.40	2.39	0.153	**0.06**	0.01	**<0.001**	0.01	<0.01	0.056	**−0.01**	<0.01	**<0.001**
Eng. Lang. Learner	−28.42	30.12	0.345	**12.89**	4.54	**0.004**	−0.01	0.01	0.612	<0.01	0.01	0.953	**−0.02**	0.01	**0.001**
Special Education	**−186.86**	24.85	**<0.001**	**16.13**	3.83	**<0.001**	**−0.15**	0.01	**<0.001**	**−0.08**	0.01	**<0.001**	**−0.10**	0.01	**<0.001**
Talented/Gifted	**−137.67**	35.76	**<0.001**	**−40.92**	5.66	**<0.001**	**0.13**	0.01	**<0.001**	**0.08**	0.01	**<0.001**	**0.12**	0.01	**<0.001**
Constant	**3058.67**	44.00	**<0.001**	**223.40**	5.70	**<0.001**	**0.77**	0.01	**<0.001**	**0.72**	0.01	**<0.001**	**0.78**	0.01	**<0.001**
Random Effects															
σ2		413,614.14			10,795.14			0.03			0.02			0.02	
τ00Student		99,113.29			1,363.01			0.03			<0.01			0.01	
Teacher		366,370.72			3,135.66			0.02			<0.01			<0.01	
NsTeacher		642			636			642			636			628	
Student		5,451			5,420			5,451			5,368			5,301	
Obs.		10,902			9,854			10,902			9,593			9,038	
Marg R2/Cond R2		0.120 / 0.586			0.120 / 0.379			0.090 / 0.671			0.069 / 0.327			0.083 / 0.405	

B. Free/Reduced Priced Lunch Eligibility Moderator Models, Motivation

	(6)			(7)			(8)			(9)			(10)		
	Expectancy			Enjoyment			Utility			Importance			Cost		
Predictors	B	SE	*p*	B	SE	*p*	B	SE	*p*	B	SE	*p*	B	SE	*p*
Grade 4	**−0.09**	0.04	**0.031**	**−0.25**	0.11	**0.02**	−0.05	0.04	0.232	−0.06	0.04	0.158	**0.13**	0.05	**0.004**
Pandemic Cohort	−0.03	0.05	0.553	0.10	0.13	0.426	−0.02	0.04	0.62	−0.03	0.04	0.559	−0.02	0.05	0.744
Gr.4 BY Pandemic	**−0.16**	0.06	**0.012**	−0.14	0.16	0.379	−0.05	0.06	0.461	−0.04	0.06	0.487	−0.05	0.07	0.504
FRL	<0.01	0.04	0.919	0.11	0.11	0.332	−0.02	0.04	0.693	0.03	0.04	0.457	−0.05	0.05	0.239
FRL BY Gr. 4	−0.04	0.05	0.438	0.01	0.12	0.906	0.02	0.05	0.619	−0.01	0.05	0.802	0.05	0.05	0.354
FRL BY Pandemic	0.03	0.06	0.653	0.01	0.16	0.946	<0.01	0.05	0.975	0.01	0.05	0.890	0.03	0.06	0.673
FRL BY Gr. 4 BY Pandemic	0.01	0.08	0.912	0.16	0.19	0.398	0.04	0.08	0.571	0.05	0.07	0.509	−0.08	0.08	0.366
Hispanic	−0.06	0.03	0.074	0.08	0.10	0.44	−0.01	0.03	0.765	<0.01	0.03	0.880	−0.01	0.04	0.876
Black/African Amer.	0.01	0.03	0.768	0.12	0.09	0.194	0.04	0.03	0.144	**0.06**	0.03	**0.046**	<0.01	0.04	0.914
Two or More Races	0.01	0.05	0.887	−0.11	0.13	0.43	0.06	0.04	0.151	0.04	0.04	0.324	0.01	0.05	0.891
Other Race	<0.01	0.06	0.966	0.11	0.16	0.495	0.14	0.05	**0.01**	0.11	0.05	**0.035**	−0.07	0.06	0.275
Boy	**0.05**	0.02	**0.015**	**0.64**	0.07	**<0.001**	0.01	0.02	0.687	−0.04	0.02	0.097	**−0.19**	0.03	**<0.001**
Eng. Lang. Learner	0.03	0.04	0.494	**0.23**	0.11	**0.038**	0.04	0.04	0.252	0.03	0.04	0.432	**−0.11**	0.04	**0.010**
Special Education	−0.06	0.03	0.106	**0.20**	0.10	**0.046**	**−0.21**	0.03	**<0.001**	**−0.12**	0.03	**<0.001**	**−0.14**	0.04	**<0.001**
Talented/Gifted	**0.30**	0.05	**<0.001**	**0.57**	0.15	**<0.001**	0.02	0.05	0.745	−0.03	0.05	0.569	**−0.33**	0.06	**<0.001**
Constant	**4.16**	0.04	**<0.001**	**4.83**	0.12	**<0.001**	**4.37**	0.04	**<0.001**	**4.37**	0.04	**<0.001**	**1.20**	0.05	**<0.001**
Random Effects															
σ2		0.53			3.22			0.55			0.52			0.64	
τ00Student		0.30			3.27			0.24			0.26			0.42	
Teacher		0.03			0.17			0.01			0.01			0.02	
NsTeacher		602			602			602			602			602	
Student		5,095			5,095			5,095			5,095			5,095	
Obs.		8,385			8,385			8,385			8,385			8,385	
Marg R2/Cond R2		0.017 / 0.394			0.024 / 0.528			0.009 / 0.314			0.006 / 0.347			0.022 / 0.421	

Note. Unstandardized coefficients; bolded Bs are statistically significant at the $p < .05$ threshold. Reference group is students who are White, girls, not eligible for free/reduced lunch, and not designated as English Language Learners, special education, or talented/gifted. Ns for each model vary depending on the number of students who completed each outcome. Each student contributes one or two years of data. Random effects for student and teacher are estimated. Marginal (for only the fixed effects) and conditional (for both the fixed and random effects) R^2 values for mixed models are calculated based on Nakagawa et al. (2017) via the performance R package (Lüdecke, 2020).

C. English Language Learner Moderator Models, Engagement and Performance

	(1)			(2)			(3)			(4)			(5)		
	Minutes Played			Level Attempts			Content Progress			Level Score Ave			Posttest Ave		
Predictors	B	SE	*p*	B	SE	*p*	B	SE	*p*	B	SE	*p*	B	SE	*p*
Grade 4	**−318.89**	47.61	**<0.001**	**−74.59**	5.80	**<0.001**	**−0.03**	0.01	**0.011**	**0.01**	0.01	**0.011**	<0.01	0.01	0.479
Pandemic Cohort	**−158.73**	24.93	**<0.001**	**−37.47**	3.76	**<0.001**	**−0.02**	0.01	**0.003**	−0.01	<0.01	0.187	0.01	0.01	0.335
Gr.4 BY Pandemic	**−402.12**	34.11	**<0.001**	−9.31	5.55	0.094	**−0.08**	0.01	**<0.001**	0.02	0.01	**0.004**	**0.02**	0.01	**0.002**
ELL	−22.95	51.20	0.654	11.63	7.69	0.130	−0.01	0.02	0.551	**0.01**	0.01	0.584	−0.02	0.01	0.090
ELL BY Gr. 4	−4.22	64.94	0.948	15.76	10.22	0.123	0.01	0.02	0.524	<0.01	0.01	0.966	<0.01	0.01	0.724
ELL BY Pandemic	−31.54	67.74	0.642	−0.49	10.21	0.962	<0.01	0.02	0.855	−0.01	0.01	0.433	<0.01	0.01	0.775
ELL BY Gr. 4 BY Pandemic	54.77	89.19	0.539	**−30.23**	14.85	**0.042**	−0.01	0.02	0.650	<0.01	0.02	0.997	−0.02	0.02	0.254
Hispanic	−32.17	23.06	0.163	4.32	3.58	0.227	−0.03	0.01	0.001	**−0.01**	<0.01	**0.017**	−0.01	0.01	0.074
Black/African Amer.	31.10	22.16	0.160	**9.72**	3.43	**0.005**	**−0.04**	0.01	**<0.001**	**−0.03**	<0.01	**<0.001**	**−0.03**	0.01	**<0.001**
Two or More Races	−23.35	31.37	0.457	−4.76	4.86	0.327	−0.02	0.01	0.147	0.01	0.01	0.429	<0.01	0.01	0.911
Other Race	13.60	38.49	0.724	−7.96	5.98	0.183	**0.03**	0.02	**0.036**	−0.01	0.01	0.227	0.01	0.01	0.314
Boy	**−118.52**	15.34	**<0.001**	−3.41	2.39	0.153	**0.06**	0.01	**<0.001**	0.01	<0.01	0.054	**−0.01**	<0.01	**<0.001**
Free/Reduced Lunch	**−42.72**	17.42	**0.014**	3.40	2.68	0.205	**−0.03**	0.01	**<0.001**	**-0.01**	<0.01	**0.010**	**−0.02**	<0.01	**<0.001**
Special Education	**−185.90**	24.86	**<0.001**	**16.29**	3.83	**<0.001**	**−0.15**	0.01	**<0.001**	**-0.08**	0.01	**<0.001**	**−0.10**	0.01	**<0.001**
Talented/Gifted	**−137.64**	35.77	**<0.001**	**−40.86**	5.66	**<0.001**	**0.13**	0.01	**<0.001**	**0.08**	0.01	**<0.001**	**0.12**	0.01	**<0.001**
Constant	**3061.20**	40.88	**<0.001**	**225.54**	5.11	**<0.001**	**0.77**	0.01	**<0.001**	**0.73**	0.01	**<0.001**	**0.78**	0.01	**<0.001**
Random Effects															
σ2	416,378.91			10,843.63			0.03			0.02			0.02		
τ00Student	97,899.04			13,35.63			0.03			<0.01			0.01		
Teacher	366,886.12			3,121.96			0.02			<0.01			<0.01		
NsTeacher	642			636			642			636			628		
Student	5,451			5,420			5,451			5,368			5,301		
Obs.	10,902			9,854			10,902			9,593			9,038		
Marg R2/Cond R2	0.120 / 0.584			0.120 / 0.376			0.090 / 0.669			0.069 / 0.326			0.084 / 0.406		

D. English Language Learner Moderator Models, Motivation

	(6)			(7)			(8)			(9)			(10)		
	Expectancy			Enjoyment			Utility			Importance			Cost		
Predictors	B	SE	*p*	B	SE	*p*	B	SE	*p*	B	SE	*p*	B	SE	*p*
Grade 4	**−0.11**	0.03	**<0.001**	**−0.22**	0.07	**0.002**	−0.04	0.03	0.146	**−0.06**	0.03	**0.014**	**0.14**	0.03	**<0.001**
Pandemic Cohort	0.01	0.03	0.765	0.11	0.08	0.173	−0.02	0.03	0.473	−0.01	0.03	0.705	<0.01	0.03	0.882
Gr.4 BY Pandemic	**−0.14**	0.04	**0.001**	−0.07	0.10	0.491	−0.02	0.04	0.653	−0.02	0.04	0.558	−0.07	0.04	0.138
ELL	**0.12**	0.06	**0.044**	0.24	0.16	0.141	0.04	0.06	0.511	0.06	0.06	0.315	**−0.17**	0.07	**0.012**
ELL BY Gr. 4	−0.04	0.07	0.523	−0.10	0.17	0.558	0.04	0.07	0.532	−0.01	0.06	0.890	**0.15**	0.07	**0.037**
ELL BY Pandemic	−0.16	0.08	0.054	−0.02	0.23	0.940	−0.03	0.08	0.721	−0.09	0.08	0.265	0.04	0.09	0.640
ELL BY Gr. 4 BY Pandemic	−0.06	0.11	0.574	0.26	0.28	0.349	−0.01	0.11	0.948	0.10	0.11	0.359	−0.22	0.12	0.075
Hispanic	−0.06	0.03	0.078	0.08	0.10	0.437	−0.01	0.03	0.772	0.01	0.03	0.874	−0.01	0.04	0.877
Black/African Amer.	0.01	0.03	0.768	0.12	0.09	0.191	0.04	0.03	0.143	0.06	0.03	0.046	<0.01	0.04	0.915
Two or More Races	0.01	0.05	0.905	−0.11	0.13	0.422	0.06	0.04	0.154	0.04	0.04	0.331	0.01	0.05	0.892
Other Race	0.01	0.06	0.892	0.11	0.16	0.506	**0.14**	0.05	**0.010**	**0.12**	0.05	**0.034**	−0.07	0.06	0.286
Boy	0.06	0.02	0.013	**0.64**	0.07	**<0.001**	0.01	0.02	0.678	−0.04	0.02	0.102	**−0.19**	0.03	**<0.001**
Free/Reduced Lunch	<0.01	0.03	0.961	0.14	0.07	0.051	<0.01	0.02	0.953	0.03	0.02	0.155	−0.03	0.03	0.250
Special Education	−0.06	0.03	0.101	**0.20**	0.10	**0.046**	**-0.21**	0.03	**<0.001**	**−0.13**	0.03	**<0.001**	**−0.14**	0.04	**<0.001**
Talented/Gifted	**0.30**	0.05	**<0.001**	**0.56**	0.15	**<0.001**	0.02	0.05	0.752	−0.03	0.05	0.563	**−0.33**	0.06	**<0.001**
Constant	**4.15**	0.04	**<0.001**	**4.81**	0.10	**<0.001**	**4.36**	0.03	**<0.001**	**4.37**	0.03	**<0.001**	**1.19**	0.04	**<0.001**
Random Effects															
σ2	0.53			3.22			0.55			0.52			0.64		
τ00Student	0.30			3.27			0.24			0.26			0.42		
Teacher	0.03			0.17			0.01			0.01			0.02		
NsTeacher	602			602			602			602			602		
Student	5,095			5,095			5,095			5,095			5,095		
Obs.	8,385			8,385			8,385			8,385			8,385		
Marg R2/Cond R2	0.018 / 0.394			0.024 / 0.529			0.009 / 0.314			0.006 / 0.347			0.022 / 0.422		

Note. Unstandardized coefficients; bolded Bs are statistically significant at the $p < .05$ threshold. Reference group is students who are White, girls, not eligible for free/reduced lunch, and not designated as English Language Learners, special education, or talented/gifted. Ns for each model vary depending on the number of students who completed each outcome. Each student contributes one or two years of data. Random effects for student and teacher are estimated. Marginal (for only the fixed effects) and conditional (for both the fixed and random effects) R^2 values for mixed models are calculated based on Nakagawa et al. (2017) via the performance R package (Lüdecke, 2020).

Appendix B

Residualized Change Regressions

A. Engagement and Performance Outcomes

	(1)			(2)			(3)			(4)			(5)		
	Minutes Played			Level Attempts			Progress			Level Score Ave			Posttest Ave		
Predictors	B	SE	*p*	B	SE	*p*	B	SE	*p*	B	SE	*p*	B	SE	*p*
Prior	0.18	0.01	<0.001	0.10	0.01	<0.001	0.55	0.01	<0.001	0.30	0.02	<0.001	0.35	0.02	<0.001
Pandemic Cohort	−556.72	25.66	<0.001	−53.3	3.83	<0.001	−0.09	0.01	<0.001	0.01	0.01	0.014	0.02	0.01	0.005
Latinx	−17.41	33.01	0.598	3.41	5.30	0.520	−0.01	0.01	0.35	−0.02	0.01	0.013	−0.02	0.01	0.015
Black/African Amer.	55.81	31.59	0.077	7.32	5.04	0.147	−0.01	0.01	0.439	−0.04	0.01	<0.001	−0.03	0.01	<0.001
Two or More Races	24.18	44.87	0.590	−5.48	7.18	0.446	< 0.01	0.01	0.994	−0.01	0.01	0.472	−0.02	0.01	0.052
Other Race	39.84	54.96	0.468	−12.12	8.88	0.172	0.02	0.02	0.191	< 0.01	0.01	0.750	0.01	0.01	0.453
Boy	−59.16	22.07	0.007	−5.53	3.55	0.119	0.03	0.01	<0.001	0.01	< 0.01	0.066	−0.01	0.01	0.221
Eng. Lang. Learner	120.19	42.29	0.004	21.61	6.25	0.001	0.03	0.01	0.011	< 0.01	0.01	0.606	−0.01	0.01	0.168
Free/Reduced Lunch	−84.71	24.88	0.001	0.27	3.89	0.945	−0.04	0.01	<0.001	−0.01	0.01	0.291	−0.01	0.01	0.051
Special Education	−70.46	35.13	0.045	22.57	5.57	<0.001	−0.04	0.01	<0.001	−0.07	0.01	<0.001	−0.06	0.01	<0.001
Talented/Gifted	−138.1	51.46	0.007	−21.8	8.50	0.010	0.06	0.02	<0.001	0.07	0.01	<0.001	0.09	0.01	<0.001
Constant	2164.38	58.99	<0.001	130.37	6.23	<0.001	0.32	0.02	<0.001	0.53	0.02	<0.001	0.51	0.02	<0.001
Random Effects															
σ2	624,558.36			13,496.68			0.06			0.03			0.03		
τ00Teacher	153,679.71			605.22			0.01			< 0.01			< 0.01		
NsTeacher	375			351			375			351			334		
Student	5,451			4,434			5,451			4,225			3,737		
Marg R2/Cond R2	0.135 / 0.306			0.078 / 0.118			0.317 / 0.393			0.123 / 0.145			0.166 / 0.195		

Note. Unstandardized coefficients; bolded Bs are statistically significant at the $p < .05$ threshold. Reference group is students who are White, girls, not eligible for free/reduced lunch, and not designated as English Language Learners, special education, or talented/ gifted. Ns for each model vary depending on the number of students who completed each outcome. Prior represents the third grade value for the relevant outcome. All students had minutes and progress, level attempts includes all attempts, even those that were replays of previously-passed levels, level score only includes plays of levels not previously passed. Each student contributes one year of data. Random effects for teacher are estimated. Marginal (for only the fixed effects) and conditional (for both the fixed and random effects) R2 values for mixed models are calculated based on Nakagawa et al. (2017) via the performance R package (Lüdecke, 2020).

B. Motivation Outcomes

	(6)			(7)			(8)			(9)			(10)		
	Expectancy			Enjoyment			Utility			Importance			Cost		
Predictors	B	SE	*p*	B	SE	*p*	B	SE	*p*	B	SE	*p*	B	SE	*p*
Prior	0.37	0.02	<0.001	0.49	0.02	<0.001	0.29	0.02	<0.001	0.35	0.02	<0.001	0.38	0.02	<0.001
Pandemic Cohort	−0.15	0.03	<0.001	0.03	0.09	0.76	−0.03	0.03	0.396	−0.02	0.03	0.624	−0.10	0.04	0.004
Latinx	−0.05	0.05	0.242	−0.07	0.12	0.558	0.02	0.04	0.648	0.07	0.04	0.137	0.03	0.05	0.551
Black/African Amer.	0.04	0.04	0.353	0.17	0.11	0.137	0.11	0.04	0.007	0.15	0.04	0.001	−0.02	0.05	0.642
Two or More Races	−0.04	0.06	0.571	−0.01	0.16	0.962	0.07	0.06	0.256	0.05	0.06	0.419	0.01	0.07	0.906
Other Race	0.07	0.08	0.386	0.44	0.20	0.028	0.09	0.08	0.229	0.15	0.08	0.057	−0.13	0.08	0.135
Boy	0.10	0.03	0.001	0.29	0.08	<0.001	< 0.01	0.03	0.969	0.01	0.03	0.860	−0.09	0.03	0.005
Eng. Lang. Learner	0.03	0.05	0.618	0.18	0.13	0.170	0.08	0.05	0.104	0.04	0.05	0.380	−0.05	0.05	0.352
Free/Reduced Lunch	−0.02	0.03	0.636	0.12	0.09	0.154	0.02	0.03	0.550	0.02	0.03	0.489	−0.02	0.04	0.582
Special Education	−0.07	0.05	0.122	0.16	0.12	0.180	−0.10	0.05	0.025	−0.04	0.05	0.414	−0.03	0.05	0.535
Talented/Gifted	0.23	0.08	0.002	0.39	0.19	0.043	0.06	0.07	0.423	0.03	0.07	0.721	−0.28	0.08	0.001
Constant	2.47	0.08	<0.001	2.16	0.13	<0.001	2.98	0.08	<0.001	2.70	0.09	<0.001	0.88	0.05	<0.001
Random Effects															
σ2	0.76			4.93			0.72			0.73			0.88		
τ00Teacher	0.01			0.11			< 0.01			< 0.01			0.01		
NsTeacher	328			328			328			328			328		
Student	3,290			3,290			3,290			3,290			3,290		
Marg R2/Cond R2	0.148 / 0.159			0.261 / 0.276			0.094 / 0.100			0.119 / 0.121			0.165 / 0.174		

Note. Unstandardized coefficients; bolded Bs are statistically significant at the $p < .05$ threshold. Reference group is students who are White, girls, not eligible for free/reduced lunch, and not designated as English Language Learners, special education, or talented/ gifted. Ns for each model vary depending on the number of students who completed each outcome. Prior represents the third grade value for the relevant outcome. All students had minutes and progress, level attempts includes all attempts, even those that were replays of previously-passed levels, level score only includes plays of levels not previously passed. Each student contributes one year of data. Random effects for teacher are estimated. Marginal (for only the fixed effects) and conditional (for both the fixed and random effects) R2 values for mixed models are calculated based on Nakagawa et al. (2017) via the performance R package (Lüdecke, 2020).

Appendix C

Percentage Play of Target Objectives Regressed on Week and Whether Matched with Non-ST Math Curriculum

	Coefficient	Standard Error
Obj/Content Aligned	0.0001	(0.0036)
Objective:		
2	0.1226***	(0.0210)
3	0.1111***	(0.0162)
4	0.0092	(0.0116)
5	0.2062***	(0.0150)
6	0.0349***	(0.0069)
7	−0.1557***	(0.0051)
8	−0.0717***	(0.0076)
9	−0.1662***	(0.0068)
10	−0.0302***	(0.0062)
11	−0.0126	(0.0066)
12	−0.035***	(0.0077)
13	−0.0022	(0.0074)
14	−0.0543***	(0.0072)
15	−0.0438***	(0.0073)
16	0.0278**	(0.0082)
17	−0.0351***	(0.0058)
18	0.0019	(0.0069)
19	0.1452***	(0.0103)
20	−0.2164***	(0.0082)
21	−0.077***	(0.0085)
22	−0.044***	(0.0079)
30	−0.0098	(0.0071)
31	−0.0243**	(0.0073)
32	−0.0065	(0.0087)
40	−0.114***	(0.0089)
50	0.0788***	(0.0080)
52	0.0878***	(0.0081)
54	0.095***	(0.0083)
56	−0.0016	(0.0086)
58	−0.0772***	(0.0093)
60	0.0592***	(0.0107)
Constant	0.7699***	(0.0049)
N (students)	1,891	
N (plays)	174,639	
R2	.2366	

Note. Standard errors in parentheses. Regression uses student-level fixed effects and therefore only examines variance within student.

*$p < .05$; **$p < .01$; ***$p < .001$.

Appendix D

Predicting Level Performance from Alignment between Objective and Instruction

	(1)	(2)	(3)	(4)
	Obj 8	Obj 31	Obj 8	Obj 31
Obj/Content Aligned	0.020	0.011		
	(0.99)	(1.01)		
Week (2-8)	0.017	−0.014		
	(0.79)	(-0.91)		
Week 3			−0.016	0.009
			(-1.29)	(0.66)
Week 4			−0.006	−0.011
			(-0.40)	(-0.82)
Week 5			−0.020	−0.015
			(-1.45)	(-1.15)
Week 6			−0.016	0.007
			(-1.19)	(0.43)
Week 7			−0.013	0.000
			(-0.99)	(0.02)
Week 8			0.006	−0.014
			(0.38)	(-1.13)
Week 9			0.005	−0.002
			(0.32)	(-0.13)
Week 10			−0.015	−0.008
			(-1.23)	(-0.65)
Hispanic	0.003	0.012	0.004	0.012
	(0.20)	(0.60)	(0.20)	(0.61)
Black/African Amer.	0.022	0.018	0.023	0.018
	(1.14)	(0.93)	(1.17)	(0.92)
Two or More Races	−0.009	0.006	−0.009	0.006
	(-0.76)	(0.38)	(-0.74)	(0.38)
Other Race	−0.010	−0.016	−0.010	−0.016
	(-0.78)	(-1.27)	(-0.78)	(-1.27)
Boy	−0.004	0.028	−0.004	0.028
	(-0.28)	(1.84)	(-0.27)	(1.84)
English Lang Learner	0.007	0.006	0.007	0.006
	(0.39)	(0.34)	(0.37)	(0.33)
Free/Reduced Lunch	−0.024	−0.014	−0.024	−0.014
	(-1.52)	(-0.95)	(-1.50)	(-0.96)
Special Education	−0.001	−0.024	−0.001	−0.023
	(-0.05)	(-1.93)	(-0.04)	(-1.93)
Talented/Gifted	0.005	0.012	0.005	0.012
	(0.32)	(0.77)	(0.32)	(0.78)
Total Levels Played	0.098***	0.047***	0.098***	0.046***
	(7.07)	(5.21)	(7.07)	(5.12)
Constant	0.011	0.023***	0.021***	0.021***
	(0.009)	(0.006)	(0.006)	(0.005)
N	8,264	8,264	8,264	8,264
R2	0.011	0.004	0.012	0.005

Note. Standard errors in parentheses. Reference group is White, girl, non-English Language Learner, not eligible for free/reduced lunch, not designated as special education or talented/gifted. Each student provides between 1 and 10 weeks of data. In models 1-2, week is represented as ordinal variables 2 through 8; in models 3-4, week is represented as a series of dummy variables with week 2 as the reference week. Standard errors clustered on student.

$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.