Computers & Education 191 (2022) 104615

Contents lists available at ScienceDirect

Computers & Education

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

The relationship between self-regulated student use of a virtual learning environment for algebra and student achievement: An examination of the role of teacher orchestration

Walter L. Leite *, Huan Kuang, Zeyuan Jing, Wanli Xing, Catherine Cavanaugh, A. Corinne Huggins-Manley

University of Florida, 1215 Norman Hall, Gainesville, FL, 32611, USA

ARTICLE INFO

Keywords:
Student self-regulated learning
Student achievement
Teacher orchestration
Virtual learning environments
System log data
Structural equation modeling

ABSTRACT

The current study examines both student self-regulated learning (SRL) and teacher orchestration in a virtual learning environment (VLE), with respect to student achievement. The study used SRL indicators derived from the log data on how students used the VLE system, survey data on how teachers made use of the VLE for Algebra instruction, as well as formative and summative Algebra assessment scores. The sample included 6,174 students being taught by 93 teachers in 49 schools in the 2018–2019 academic year. Multilevel structural equation modeling (SEM) analysis revealed the relationship between SRL indicators and student achievement. We found a positive relationship between student SRL and student achievements. Specifically, three SRL indicators, "Watched Recommended Videos", "Watched Videos" and "Answered Quizzes after Watching a Video", were significantly associated with student achievement. We did not observe direct associations between teacher orchestration and student achievement, but we found an indirect association between teacher orchestration and student achievement via one SRL indicator, "Watched Recommended Videos". The results show that a positive relationship between EOC scores and student reviewing incorrect questions increased as the use of instructional videos by the teacher increased. This study supports the critical role of SRL and indicates that teachers should have flexibility in adapting learning activities for online learning.

1. Introduction

Virtual learning environments (VLE) have been extensively developed during the last 30 years, as evidenced by 180 studies reviewed across seven meta-analyses of learning using such systems (du Boulay, 2019). Most evaluations of VLE focused on experimental studies comparing the achievement of students using the VLE to the achievement of students in a traditional learning environment (e.g., Ma et al., 2014; Steenbergen-Hu & Cooper, 2013). However, these experimental studies focused on the causal attribution of changes in student achievement when exposed to the VLE, but did not engage in modeling the relationships between students' and teachers' strategies of using the VLE and student achievement. The relationships between specific actions of students in a VLE, such as watching or re-watching videos or taking practice quizzes, and student achievement have been scarcely studied. These actions are crucial to study since they reflect student self-regulated learning (SRL). Furthermore, the interplay between student SRL

* Corresponding author.
E-mail address: walter.leite@coe.ufl.edu (W.L. Leite).

https://doi.org/10.1016/j.compedu.2022.104615
Received 17 September 2021; Received in revised form 4 August 2022; Accepted 7 August 2022
Available online 18 August 2022

and teacher orchestration has not been examined in the literature. The objective of this study is to use secondary data from a large-scale field experiment to investigate the relationship between student SRL in a VLE and student achievement. We also examined whether the relationship between student SRL and student achievement is moderated by teacher orchestration.

SRL strategies employed by students influence learning in a VLE both during and outside of class time (Alavi et al., 2009), and affect their achievement. Self-regulation research in mathematics specifically indicated the relationships among student self-regulation, academic enjoyment, and achievement (Ahmed et al., 2013), as well as among goal setting, self-evaluation, and performance (Fuchs et al., 2003). Researchers found that SRL positively affected achievement (Ahmed et al., 2013; Fuchs et al., 2003). These SRL strategies, taken together, represent a student's intrinsic willingness to learn, and include practices such as environment structuring, goal setting, self-consequences (self-rewarding and self-punishment), self-evaluating, seeking social assistance, and reviewing class notes (Wang, 1983; Zimmerman, 1983; Zimmerman & Pons, 1986). SRL with digital technology is so important that in its absence, learners are likely to struggle with complex topics and could fail to achieve conceptual understanding (Azevedo & Hadwin, 2005).

When a VLE is used with a whole class, the teacher has a critical role in developing strategies to integrate the VLE and other technologies as well as activities into coherent pedagogical practices (Prieto et al., 2011). Prieto et al. (2011) provided a taxonomy of mechanisms by which a teacher orchestrates student use of technology. du Boulay (2019) emphasized this point by differentiating the pedagogy of the VLE system from the pedagogy for teacher orchestration of the use of the system. Instrumental orchestration theory explains the types of orchestrations that teachers develop for managing complexity when using technology, and the relationship of teachers' views of mathematics education and their orchestration preferences (Drijvers et al., 2010). In addition to established practices and beliefs influencing teacher orchestration with technology in mathematics teaching, the feedback provided by the technology can affect the ways teachers orchestrate learning (Holstein et al., 2019).

Although there have been several studies on student SRL with technology in mathematics (Ahmed et al., 2013; Fuchs et al., 2003; Rienties et al., 2019), and teacher orchestration in mathematics VLEs (Drijvers et al., 2010, 2020; Holstein et al., 2019), there is a scarcity of studies examining whether the relationship between student SRL in a VLE and student achievement is moderated by teacher orchestration. This scarcity can, in part, be explained by the predominant focus in VLEs on students as learners independent of teachers (Hamalainen & Osakanen, 2012). However, researchers recognized the potential relationships between student SRL and teacher orchestration in teacher-facilitated learning in VLEs (Hamalainen & Oksanen, 2021; Göbel et al., 2012). This is particularly important because, with large-scale adoption of VLE in multiple schools, students and teachers may have a great deal of flexibility in when and how to use the VLE, and different choices may lead to diverse outcomes (Bartelet et al., 2016). Given these research gaps, the current study addresses the following research questions:

- What is the relationship between student SRL in an Algebra VLE and student achievement in formative assessments and a summative high-stakes test?
- Does the relationship between student SRL in an Algebra VLE and their achievement depend on teacher orchestration in the VLE?

In terms of the overall relationship between student SRL and their achievement based on the first research question, we hypothesized that students with a higher level of SRL strategies will be associated with higher scores on the formative and summative assessments. In terms of teacher orchestration as a moderator based on the second research question, we hypothesized that teacher orchestration in the VLE will enhance the relationship between student SRL and student achievement. We suspected that teacher orchestration will affect the strength but not the direction of the relation between student SRL and their achievement. These hypotheses are supported by previous research showing that teacher orchestration has a positive relationship with student productive knowledge construction (Hamalainen & Osakanen, 2012).

1.1. *Student self-regulated learning*

Research indicates that self-regulated students are aware of their learning process, and able to take an active role in adjusting to different learning environments (McCardle & Hadwin, 2015). Self-regulated students could effectively use SRL strategies including "actions directed at acquiring information or skill that involve agency, purpose (goals), and instrumentality self-perceptions by a learner" (Zimmerman & Pons, 1986). SRL strategies consist of a list of categories representing students' intrinsic willingness during learning, such as environment structuring, goal setting, self-consequences (self-rewarding and self-punishment), self-evaluating, seeking social assistance, and reviewing class notes (Wang, 1983; Zimmerman & Pons, 1986).

The significant increase in online learning has brought much research attention to SRL in VLEs. SRL is assumed to be essential for success and efficiency in VLEs as students have agency regarding when, what, and how to learn as well as with whom to learn as learning partners with limited presence of an instructor (Lehmann et al., 2014; Rienties et al., 2019). Studies have repeatedly reported the positive relationship between SRL and students' achievement. For example, Fuchs et al. (2003) found that goal setting and self-evaluation are positively associated with students' achievement in a third-grade math classroom. Ahmed et al. (2013) studied the relationship between SRL, academic emotions, and academic achievement in 7th-grade mathematics classes. They found that changes in academic enjoyment towards the course and pride towards oneself consistently led to changes in students' SRL strategies and achievement in a positive way. The positive relationship sustains for higher education and online settings, as well. With first-year college students, Rienties et al. (2019) studied the temporal relationships among SRL strategies, students' engagement, emotions and academic performance in a blended learning environment where face-to-face lectures and online learning materials are available. Their research indicated students' dispositions of SRL strategies are closely connected with their learning processes and performance. Meanwhile, learners have a hard time regulating their learning on complex topics and may not achieve conceptual understanding in the absence of SRL in online learning settings (Azevedo & Hadwin, 2005). Broadbent and Poon (2015) conducted a systematic review and identified nine strategies that have been examined by existing research: time management, peer learning, elaboration, effort regulation, metacognition, critical thinking, organization, rehearsal, and help seeking. They found that time management, effort regulation, metacognition and critical thinking are statistically significantly related to academic achievement. Wong et al., 2019 systematically reviewed 35 empirical studies and concluded several commonly used strategies to support SRL in online learning: prompts, integrated support systems, feedback, a combination of prompts and feedback, and others. According to their research, prompts are the most commonly used strategy. From an instructional perspective, researchers recommended promoting SRL behaviors in online learning environments for improving students' performance purpose (Wang et al., 2013).

1.2. Teacher orchestration

The effectiveness of technology for learning is partially dependent on how teachers orchestrate student learning. Orchestration in the context of technology-enhanced learning is defined as "coordinating a teaching/learning situation, from the point of view of the teacher" (Prieto et al., 2011), with the teacher as a discerning mediator of learning who is reflecting in action on the "affordances and constraints of situation and interaction" (Joyce-Gibbons, 2014, p. 52). Orchestration functionalities include actions taken by teachers as they enable learning activities with technology, monitor student learning activities, and adapt activities in response to the dynamic classroom context (Dillenbourg, 2013). These actions express a teacher's pedagogical empowerment and typically attempt to minimize orchestration load through selective use of technology. Instrumental orchestration theory explains the types of orchestrations teachers develop for managing complexity when using technology (Drijvers et al., 2010). Based on the instrumental orchestration theory, orchestration types include discuss-the-screen, explain-the-screen, technical-demo, Sherpa-at-work, and link-screen-board, which indicate a teacher's student-centeredness and level of technology adaptation. Instrumental orchestration accounts for the people in the teaching-learning scenario, the learning objectives, the configuration of the learning activity, VLE exploitation, and a didactical performance (Drijvers et al., 2020).

Orchestration of technology-enhanced learning has been observed to be a developmental practice among teachers (Artigue, 2002). After an exploratory phase in which new digital tools are used ad hoc in ways similar to established practices, teachers adapt their theoretical discourse to accompany productive uses of the tools. Moreover, the teachers used these new digital tools in a more structured and efficient way. In addition to established practices and beliefs influencing teacher orchestration with technology in teaching, the feedback provided by the technology can affect the ways teachers orchestrate learning (Holstein et al., 2019). Specifically, teachers who receive real-time analytics from the technology used by students shift to whom teachers give individual attention and reduce the time between receiving feedback and providing guidance to students.

Educational technology often provides feedback to students, including options for getting help from the application, from peers, and from teachers. An area for supporting teacher orchestration is offering the hand-off between students and teachers by the technology (Fancsali et al., 2018). Design recommendations include providing the teacher with relevant details of the student's specific need, so teachers can orchestrate by grouping students or supporting students in timely ways. Further, a VLE could proactively involve a teacher when students need a teacher's help, thus removing some of the orchestration load from teachers in real time and leveraging their "superpowers" (Holstein et al., 2019). Specific indicators such as which students are not using their learning applications productively, which students are struggling with learning and stalled in their learning progress, which students are making repeated errors, which students have been inactive for a period of time, and which students have made improvements, are useful to teachers. With such indications, teachers were able to more strategically allocate their time to help their students to learn (Holstein et al., 2019). There has been increasing research regarding the types of feedback from teachers that correspond to student success, as well as which students require teacher support and the specific student needs. For example, Joyce-Gibbons (2014) found that students experienced greater rates of success when teachers used problematization utterances with students engaged in group problem solving tasks. Ultimately, the outcomes realized from teacher orchestration in a learning experience using technology are dependent on the teacher's beliefs, technological pedagogical content knowledge, and her knowledge of the students (Looi & Song, 2013). Efforts to foster the skills underlying effective orchestration have included overlay technology that stitches together components of a learning environment to provide increased awareness and decision support to the teacher, with the recognition that novice educators are likely to benefit from more supports than experienced teachers (Muñoz-Cristóbal et al., 2015). Orchestration skill development has also been aided by tools that offer post-lesson feedback to teachers with opportunities to reflect and incorporate learning into their design of future lessons (Rodríguez-Triana et al., 2015).

1.3. Relationship between SRL and teacher orchestration

A recently proposed bridge in VLEs between SRL and teacher orchestration is teacher regulation, in which teachers use data about student activity in the VLE that indicate the lack of self-regulation to regulate their orchestration among whole-group, small-group, and individual facilitation of learning (Dillenbourg, 2021). In VLE, teachers enable learning activities with technology, monitor student learning activities, provide feedback, and/or adapt the course to the dynamic context via the use of instructional videos, workbooks, and other available materials. Meanwhile, students seek information, and develop their knowledge and skills via watching videos, reviewing materials, doing exercises, asking for help, and reviewing solutions. Essentially, effective teacher orchestration of learning in VLEs is directly dependent on timely indicators of issues in individual SRL that prompt teacher intervention. Therefore, there is potential for interactions between teacher orchestration and students SRL strategies. Fig. 1 shows this conceptualization where the direct effects of student self-regulated learning (e.g., information seeking, reviewing) on both formative and summative assessment performance depend on the orchestration implemented by teachers. Yet, limited work has investigated these interactions because of the lack of research in teacher-led VLEs.

2. Methods

In this section, we will first introduce the setting of the VLE that was used to obtain the data. Then we discuss the sample and variables used in this paper. We also present the analyses step by step from forming the latent variables for teacher orchestration, to modeling the relationship between student SRL and their achievement, and to evaluating the influence of teacher orchestration on the relationship between student SRL and student achievement.

2.1. Setting

This study obtained data from Math Nation, formerly Algebra Nation, an Algebra VLE that is extensively available in school districts in Florida, with some districts adopting it as the main curriculum (Leite et al., 2019). It is integrated with the school districts' student information system so that every student and teacher can log in with their school username and password. Teachers typically use the VLE in conjunction with the paper workbook (Mitten et al., 2021). This VLE has a series of instructional videos and formative assessments organized into ten domains (e.g., linear equations, quadratic equations, exponential functions, etc.). These domains are aligned with the state's Algebra standards. Within each domain, there are several topics (ranging from 6 to 12) with a total of 93 topics across all ten domains. Each topic is associated with a main instructional video, and there are five versions of each video that were taught by different tutors who are ethnically diverse and have different instructional approaches. Students can choose which tutor to follow for these videos. In each video, the tutor goes over an example Algebra 1 question. The students have these questions available to them in a printed workbook, which is provided to all students whose teachers request it. There is a fixed 3-question formative quiz at the topic level, called a Check Your Understanding (CYU) quiz, with varying question formats that are similar to the state's Algebra 1 End-of-Course (EOC) assessment. Secondly, there are 10-item formative assessments at the domain level, called Test Yourself (TYS) assessments. Each of the 10 domains has a large item bank, from which random sets of TYS assessments can be drawn on demand. There is a solution video for each TYS question, which is available to the student once they complete the 10-item TYS assessment. The VLE also has a monitored discussion forum.

2.2. Sample

Secondary data for this study was obtained from a large-scale field experiment, which was designed to evaluate the effect of a reinforcement learning algorithm for displaying video recommendations to students after they complete CYU quizzes. Large scale field experiments using VLE are rare (e.g., Roschelle et al., 2016; Phillips et al., 2020), because they are difficult to implement, require large scale adoption of the VLE, and access to student achievement data. However, field experiments are stronger than lab experiments with

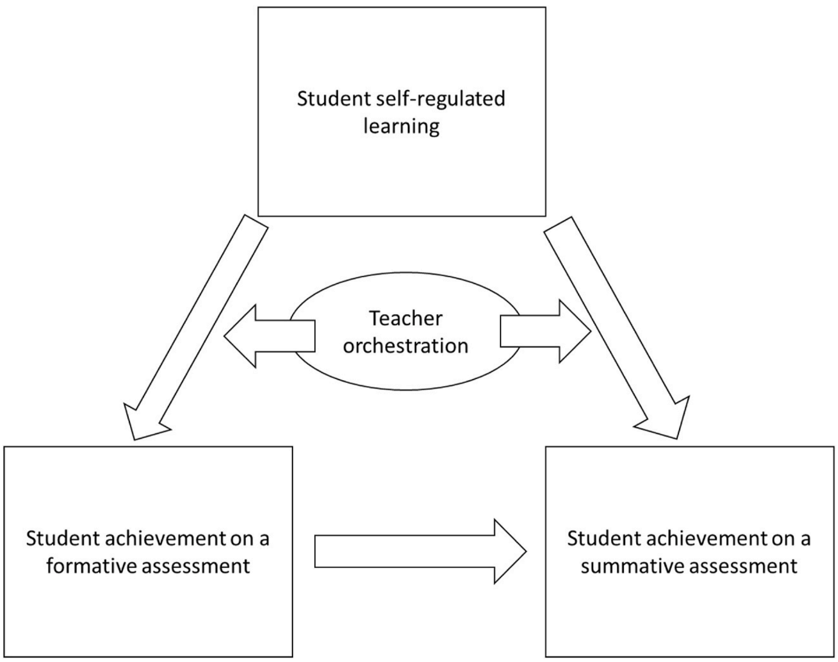

Fig. 1. Conceptual model of the relationship between student SRL and achievement moderated by teacher orchestration.

respect to external validity (Shadish et al., 2002), because they produce results that may generalize to other students, teachers, schools, and educational settings.

The field experiment lasted one semester (Spring 2019). Students in the sample were randomly assigned to see two types of video recommendations. The treatment group received video recommendations obtained from a Markov decision process (Shen et al., 2018), while the control group saw recommended videos that were the next video in the curriculum's sequence, with the assignment being blind to students and teachers (Chakraborty et al., 2021, for details of the system's structure). The results of the field experiment showed that exposure to the recommendation system without accounting for the amount or type of usage had no effect on student achievement, but a high total number of recommended videos viewed by a student was associated with improved scores on a state-mandated high-stakes test. However, this previous study focused on evaluating the average treatment effect of student exposure to the video recommendation algorithm and did not examine student SRL or include any teacher data to investigate teacher orchestration.

The sample for the current study consisted of middle and high school students who enrolled in Algebra 1 or Algebra 1 Honors courses in 2018–2019 academic year. The sample included 6,174 students being taught by 93 teachers in 49 schools from a large school district in Florida. Among these students, 49.4% were male, 33.2% were Hispanic, and 65.2% were eligible for free or reduced lunch.

2.3. Measures

The current study used survey data on how teachers made use of the VLE for Algebra instruction, as well as log data from the VLE on how students used this VLE, and formative and summative Algebra assessment scores. In this measures section, we first introduce the six SRL indicators. Secondly, we explain how we measured teacher orchestration. Lastly, we discuss how we measure student achievement based on formative assessments in the VLE and one summative high-stakes test.

2.3.1. Measures of student self-regulated learning

Previous research has found that student SRL was related to student achievement (Broadbent et al., 2021; Broadbent & Poon, 2015). However, the majority of these studies used student self-reported responses to the Motivated Strategies for Learning Questionnaire (MSLQ) or the Learning and Study Strategies Inventory (LASSI) as an overall measure of SRL (Broadbent & Poon, 2015). One caveat of collecting information through a self-reported measure is that respondents may be biased when they report their own experiences (Rosenman et al., 2011). To overcome this shortcoming, a number of studies recently explored using log data to investigate student SRL strategies in online learning environments (e.g., Kim et al., 2018; Wong, Baars, et al., 2019; Li et al., 2020). A series of log variables have been used to measure SRL, such as time spent on online courses/Q&A board, login intervals (Kim et al., 2018), the number of times a learner views prompts/watches videos (Wong, Baars, et al., 2019), and the frequency of notes-makings/reflections (Li et al., 2020). To summarize, the frequency of observed actions can be used as one explicit indicator of SRL (Winne, 2010).

In this study, six indicators were obtained from the system logs of the VLE to present student SRL after two experts independently selected which logs were consistent with existing theories of student SRL (see Table 1). These indicators were observed variables, namely, 1) "*Reviewed Incorrect Questions*", 2) "*Watched Solution Videos*", 3) "*Watched Recommended Videos*", 4) "*Watched Videos*", 5) "*Answered Quizzes after Watching a Video*", and 6) "*Answered Quizzes without Watching a Video*". The video refers to the instructional/content video in the last three variables. All the student SRL indicators were log-transformed to reduce non-normality. Moreover, we chose to obtain two indicators ("*Reviewed Incorrect Questions*" and "*Watched Solution Videos*") before the experiment and the other four indicators during the experiment to reduce the collinearity between these indicators.

"*Reviewed Incorrect Questions*" and "*Watched Solution Videos*" before the experiment were selected as SRL indicators because they are examples of meta-cognition in the form of monitoring (Chang, 2007). Specifically, "*Reviewed Incorrect Questions*" and "*Watched Solution Videos*" showed that students were interested in understanding why their answer was incorrect and how they can work toward a correct solution. These two indicators were created by taking the log-transformation of the frequency (counts) regarding how many incorrect questions a student reviewed and how many solution videos they watched before the experiment (from August 2018 to January 2019).

"*Watched Recommended Videos*" and "*Watched Videos*" were selected as SRL indicators because when students choose to watch content videos and to watch them completely, they are enacting the effort regulation SRL strategy (Carson, 2011; ChanLin, 2012; Cho & Shen, 2013). While students watch videos in an online course, they are highly likely to regulate themselves in the process of learning. "*Watched Recommended Videos*" was created by taking the log-transformation of the frequency (counts) regarding how many

Table 1
Descriptive analyses of the student SRL indicators.

Measure	Min.	Max.	Mean (SD)
Watched Solution Videos*	0.000	3.401	0.072 (0.348)
Reviewed Incorrect Questions*	0.000	5.493	0.296 (0.905)
Watched Recommended Videos	0.000	4.771	0.147 (0.536)
Watched Videos	0.000	6.100	2.319 (1.634)
Answered Quizzes after Watching a Video	0.000	6.948	0.483 (1.276)
Answered Quizzes without Watching a Video	0.000	7.363	1.516 (2.127)

Note. * indicate the variables were measured before the experiment.

recommended videos a student watched during the experiment (from February 2019 to May 2019). "*Watched Recommended Videos*" was only available after the field experiment started in Spring 2019 (Chakraborty et al., 2021). Likewise, "*Watched Videos*" was created by taking the log-transformation of the frequency (counts) regarding how many content/instructional videos a student watched during the experiment (from February 2019 to May 2019).

"*Answered Quizzes after Watching a Video*" and "*Answered Quizzes without Watching a Video*" were selected as SRL indicators because responding to quizzes after watching videos is meta-cognition as self-testing (Carson, 2011). These two indicators were created by taking the log-transformation of the frequency (counts) regarding how many CYU quizzes a student answered after watching a video and without watching a video during the experiment (from February 2019 to May 2019).

2.3.2. Measures of teacher orchestration

We used one observed and three latent variables for measurement of teacher orchestration, because the existing literature has not presented any general scale to measure teacher orchestration behaviors. The data about teacher orchestration were collected from a teacher survey, which was administered at the end of the experimental study (in June 2019). Participating teachers responded to the survey about how they used the VLE in their classroom during Spring 2019. Specifically, the survey included a series of questions about how teachers orchestrated the instructional videos (content videos) and workbooks in the VLE to their classroom teaching, as well as how they prepared their students for the state-mandated high-stakes Algebra test.

The observed variable was "*Use VLE as Remediation Tool*", which was derived from a yes-or-no survey question about whether teachers used the VLE for remediation. Specific strategies used for remediation with this VLE include having individual students work on the pre-Algebra adaptive remediation system or review previous videos related to Algebra skills that the student may have not mastered. Moreover, three latent variables that reflect teacher orchestration were proposed. We included these three latent variables in this study because they account for teacher responses to 17 survey questions, as shown in Table 3. The first latent variable was termed "*Use of Instructional Videos*", and it was measured by six survey questions such as "I use all or a portion of an instructional video as a bell ringer/warm up activity". The second latent variable was termed "*Use of Workbooks*", and it was measured by five survey questions such as "I assign workbooks to groups/centers". The first and second latent variables are important because both instructional videos and the workbook are the core components of the VLE. The third latent variable was termed "*Preparation for High-stakes Tests*", and it was measured by six survey questions such as "I taught students to use test-taking strategies". Each survey question was measured through a 5-point frequency scale: 1) not used, 2) less than once in every five classes, 3) one or two times in every five classes, 4) three or four times in every five classes, and 5) in every class. This third latent variable is important because the VLE was developed to prepare students to pass the state's Algebra 1 End-of-Course (EOC) exam (Florida Department of Education, 2021).

We obtained internal structure validity evidence (American Educational Research Association et al., 2014) for the three latent variables using confirmatory factor analysis (CFA; Bandalos, 2018). The specific CFA models and related results can be found in section 2.4.1, section 3.1, and Table 3. It is worth noticing that although the SRL and teacher orchestration measures were specifically used in this study, they aligned with previous studies of this VLE (Leite et al., 2022) where focus groups were used to identify strategies that teachers use with the VLE, and a previous survey of teacher use, providing content validity evidence (American Educational Research Association et al., 2014).

2.3.3. Measures of student achievement

Student achievement was measured by a proximal outcome based on the formative assessments in this VLE, and one distal outcome based on a summative assessment outside this VLE. The proximal outcome was "*Total Quiz Scores after Watching Videos*", and this variable was created for each student by summing their CYU quiz scores across all CYU quizzes that were answered during the experiment. Due to the nature of the experimental design, we only used the CYU quizzes that were completed after watching a video. As mentioned before, there were 93 CYU quizzes and each had 3-items, covering 93 topics that are aligned to the state's adopted curriculum. Psychometric properties (i.e., discrimination and difficulty) for the CYU items were obtained using item response theory (IRT) in Xue et al. (2021). Furthermore, two SRL indicators, "*Answered Quizzes after Watching a Video*" and "*Answered Quizzes without Watching a Video*", also served as covariates to control for variation in the number of quizzes taken across students.

This study's distal summative outcome was the students' scores on the Algebra 1 EOC exam, which is a high-stakes test mandated by the state for high school graduation. The Algebra I EOC assessment is a computer-based and criterion-referenced assessment that measures the Florida State Standards for Algebra I (Florida Department of Education, 2021). Reliability and validity information about this assessment is provided by the Florida Department of Education (2018). A passing Algebra I EOC assessment score (hereafter referred to as "*EOC Score(s)*") was required for high school graduation in Florida prior to the COVID-19 pandemic. The passing score for Spring 2016 and beyond was 497. The average EOC score in our sample was 500.26 ($SD = 31.64$, $Min = 425$, $Max = 575$). We also included the students' scores on the Mathematics Florida Standards Assessment (FSA) from the previous year (2017–2018 academic year) as a covariate. This is an important covariate to control for because the mathematics FSA scores in the previous year (hereafter referred to as "*Previous FSA Scores*") correlate strongly with the EOC scores (Pearson $r = 0.745$). Therefore, we treated the previous FSA as a pre-test for Algebra I EOC exam, and the "*Previous FSA Scores*" as a proxy of students' previous Algebra achievement. The "*EOC Scores*" in May 2019 and "*Previous FSA Scores*" in May 2018 were obtained from the school district. These two scores were converted to z scores in this study.

2.4. Analyses

In the analyses section, we first introduce the CFA models which were used to form the three latent teacher orchestration variables.

We discuss the standards (model fit indices) to assess how well the latent teacher orchestration variables were measured by the corresponding survey questions. Secondly, we present the multilevel SEM model which was used to answer the two research questions.

2.4.1. Confirmatory factor analysis

We fitted a single-factor CFA for each of the three teacher orchestration latent variables (i.e., "*Use of Instructional Videos*", "*Use of Workbooks*", and "*Preparation for High-stakes Tests*") to obtain validity evidence of adequate internal structure of the data underlying each latent variable (American Educational Research Association et al., 2014; Bandalos, 2018). We fitted three separate one-factor CFA models rather than a single three-factor CFA because the teacher-level sample size (i.e., n = 93) did not support fitting a larger model, given that we targeted having at least five observations per parameter estimated (Jackson, 2003). Each CFA was fitted by using maximum likelihood estimation with the "*lavaan*" package (Rosseel, 2012; version 0.6–7) in the R statistical software (version 4.0).[1] After evaluating the quality of each latent variable, factor scores were estimated with the regression method.

There were four evidence sources obtained in this study to assess the statistical quality of the latent teacher orchestration variables, namely, 1) adequate fit of the hypothesized one-factor models, 2) appropriate size of factor loadings, 3) high composite reliability, and 4) high average variance extracted. Firstly, we evaluated both exact and close model fit for each CFA. To determine the exact model fit, we looked at whether the chi-square (χ^2) test was significant. However, because the exact fit is rarely attainable, we examined multiple close model fit indices. The standards for close model fit were whether the Comparative Fit Index (CFI) and Tucker Lewis Index (TLI) exceeded 0.9, whether the standardized root mean squared residual (SRMR) was less than 0.08, and whether the root mean squared of approximation (RMSEA) was less than 0.06 (Marsh et al., 2004). These threshold values have been widely used as acceptable criteria for the close fit in SEM (Byrne, 1994). Secondly, we used the recommended minimum of 0.40 for evaluating the size of factor loadings (Ford et al., 1986). A CFA model with factor loadings higher than 0.40 was considered an acceptable model. Thirdly, we used composite reliability (i.e., coefficient Ω) as reliability estimation, instead of the widely used Cronbach's alpha, which is known to underestimate reliability (Raykov, 1997, 1998). A commonly used standard about Ω was that the values exceeding 0.70 indicate adequate reliability for low-stakes measurement purposes (Bacon et al., 1995; Morera & Stokes, 2016; Raykov, 2009a, 2009b). Lastly, we aimed at having the average variance extracted (AVE) at 0.5 for each CFA model, which means that on average, at least 50% of the variance of the indicators is explained by the latent variable (Fornell & Larcker, 1981), with the remaining variance as random error.

2.4.2. Multilevel structural equation modeling

Multilevel SEM modeling was conducted to simultaneously answer the first and second research questions. Fig. 2 shows the two-level multilevel SEM model (Muthén, 1994) used in this study. To address our first research question, the within-level (i.e., student level) was specified to estimate the relationships between student SRL indicators and student achievement. At the between-level (i.e., teacher level), we modeled the direct relationship between teacher orchestration measures and average student achievement. The interactions between teacher orchestration and two SRL indicators ("*Reviewed Incorrect Questions*" and "*Watched Solution Videos*" before the experiment) were added to the model (cross-level) to address the second research question. We estimated the multilevel SEM with the *Mplus* 8.1 software (Muthén & Muthén, 2017) using robust maximum likelihood estimation.[2] We evaluated model fit of the multilevel SEM using the same criteria presented earlier for the CFAs.

To facilitate understanding, we explain the model at each level with respect to three endogenous variables (i.e., "*Watched Recommended Videos*", "*Total Quiz Scores after Watching Videos*", and "*EOC Scores*"). "*Total Quiz Scores after Watching Videos*" and "*EOC Scores*" represented student achievement. "*Watched Recommendation Videos*" was an endogenous variable related to SRL.

2.4.2.1. "Watched recommended videos". We added four student level variables (i.e., "*Treatment Status*", "*Previous FSA Scores*", "*Watched Videos*" and "*Answered Quizzes after Watching a Video*") as predictors of *Watched Recommended Videos*" to control for confounders due to students self-selecting to watch or reject video recommendations. We added "*Treatment Status*" as a covariate to control for the variance in "*Watched Recommended Videos*" that were due to random assignment to treatment or control group in the experiment. Although the current study was not specifically interested in estimating the effects of random assignment to the video recommendation conditions (see Chakraborty et al., 2021; Leite et al., 2022; Leite et al., 2022 for estimates of the treatment effect of the video recommendation system), including the binary indicator of random assignment to treatment or control group in the model is helpful, because it is an instrumental variable (Bollen, 2012) for the number of watched recommended videos. In other words, regressing the "*Watched Recommended Videos*" on the binary indicator of treatment assignment removes confounding due to students self-selecting to watch recommended videos, given that the required assumptions for instrumental variables were met (see Angrist, 2006, for details on the assumptions). As mentioned in section 2.3.3, "*Previous FSA Scores*" was treated as a proxy of students' previous Algebra achievement. Therefore, we included it in the model to control for variance in "*Watched Recommended Videos*" that was due to students' previous achievement. We added student SRL indicators, "*Watched Videos*" and "*Answered Quizzes after Watching a Video*", as covariates of "*Watched Recommended Videos*", because the VLE only displays a recommended video after a student completes a 3-item CYU quiz, and quizzes are frequently completed after watching a video assigned by the teacher. Therefore, it is important to control for the differences in the frequencies when students take quizzes and watch videos.

At the teacher level, we added four teacher orchestration variables (i.e., "*Use of Instructional Videos*", "*Use of Workbooks*", "*Use VLE*

[1] The R code for the CFAs is available in the Open Science Framework site https://osf.io/8gbwf/.

[2] The Mplus code and output for the multilevel SEM is available in the Open Science Framework site https://osf.io/8gbwf/.

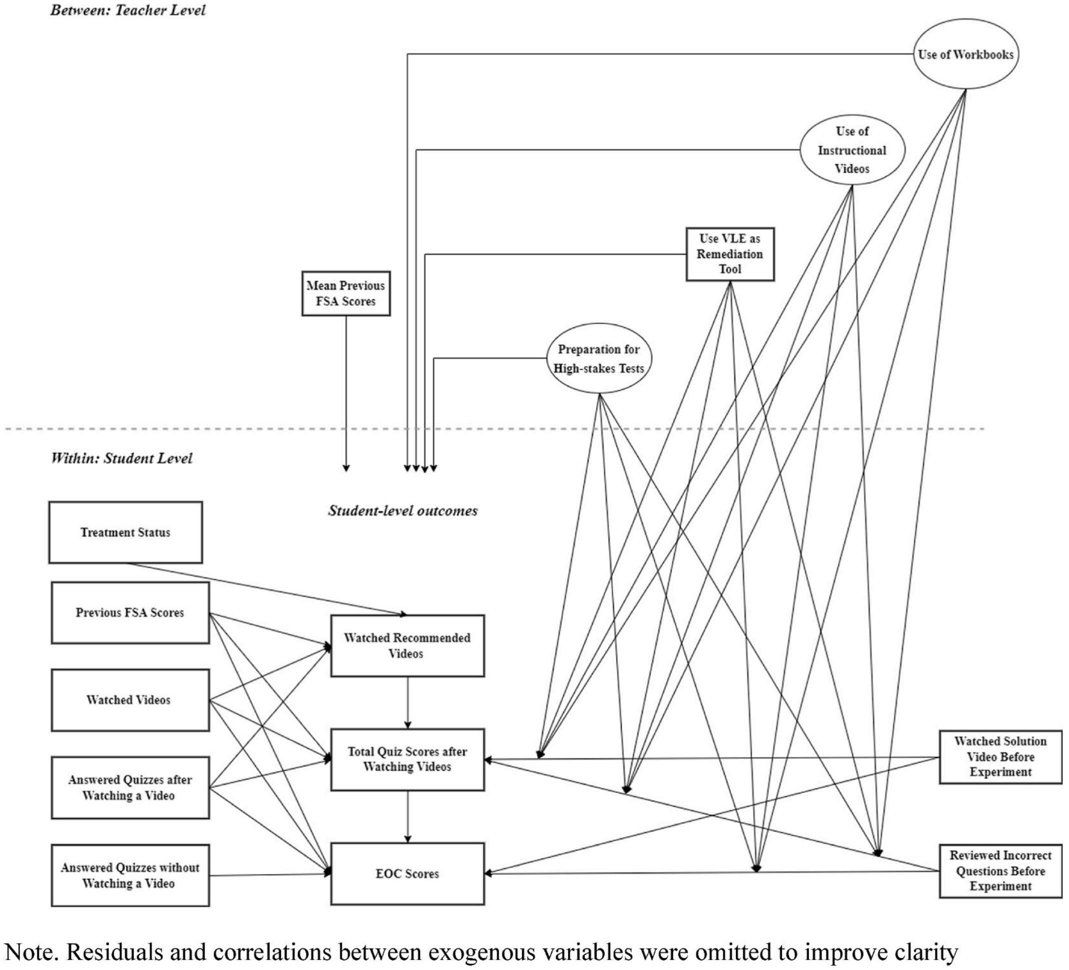

Fig. 2. Multilevel Structural Equation Model
Note. Residuals and correlations between exogenous variables were omitted to improve clarity.

as Remediation Tool" and "*Preparation for High-stakes Tests*") to control for the extent that teacher orchestration contributes to the differences in the frequency that students watched recommended videos. We added "*Mean Previous FSA Scores*" as a covariate to control for the aggregate previous achievement of all students in the classroom.

2.4.2.2. "Total quiz Scores after watching videos". Based on the first research question, the hypothesis evaluated with this part of the multilevel SEM model was that students with a higher level of SRL strategies will be associated with higher "*Total Quiz Scores after Watching Videos*". We examined the relationship between "*Total Quiz Scores after Watching Videos*" and six student level variables. Among these six variables, five were student SRL indicators and one was "*Previous FSA Scores*" that served as a covariate variable to control for students' previous Algebra achievement. One student SRL indicator, namely, "*Answered Quizzes Without Watching a Video*", was excluded from predicting "*Total Quiz Scores after Watching Video*" since it did not directly contribute to this outcome. At the teacher level, we examined the direct relationship between "*Total Quiz Scores after Watching Videos*" and the four teacher orchestration variables. Additionally, "*Mean Previous FSA scores*" and "*Mean Watched Recommended Videos*" were added as covariates to control for students' previous achievement and aggregated frequency that students watched recommended videos in the classroom.

2.4.2.3. "EOC scores". Based on the first research question, the hypothesis for this part of the multilevel SEM model was that students with a higher level of SRL strategies will be associated with higher "*EOC Scores*". We examined the relationship between "*EOC Scores*" and seven student-level variables. Among these seven variables, five were student SRL indicators, one was "*Previous FSA Scores*", and the last one was "*Total Quiz Scores after Watching Video*". One student SRL indicator, namely, "*Watched Recommended Videos*", was excluded from regressing "*EOC scores*" since the effect of "*Watched Recommended Videos*" was explained through "*Total Quiz Scores after Watching Video*". At the teacher level, we examined the direct relationship between "*EOC Scores*" and the four teacher orchestration variables. Additionally, "*Mean Previous FSA scores*" and "*Mean Total Quiz Scores after Watching Videos*" were added as covariates to control for students' previous achievement.

2.4.2.4. Interactions between SRL and teacher orchestration. We included two-way interactions between four teacher orchestration variables and the two SRL indicators (i.e., "*Watched Solution Videos*", and "*Reviewed Incorrect Questions*") in the multilevel SEM. These two SRL indicators were selected because they were collected before the experiment, such that they were relatively independent of the teacher orchestration during the experiment. We represented the interactions using product terms between each student-level variable and teacher-level variable. Student-level variables were group-mean centered and teacher-level variables were grand-mean centered before calculating the product terms (Enders & Tofighi, 2007). We regressed eight interaction variables on student level "*Total Quiz Scores after Watching Videos*" to investigate the second research question. The hypothesis was that teacher orchestration would affect the relation between student SRL and their "*Total Quiz Scores after Watching Videos*".

We also regressed eight interaction variables on student level "*EOC Scores*" to investigate the second research question. These interaction variables were created between the four teacher orchestration variables and the two SRL indicators (i.e., "*Watched Solution Videos*", and "*Reviewed Incorrect Questions*"). The hypothesis was that teacher orchestration would affect the relation between student SRL and their "*EOC Scores*".

Moderation was not evaluated for the other four SRL indicators that were measured after the experiment because these can be affected by teacher orchestration strategies, so they may be mediators. Although it is possible to model moderated mediation (Preacher et al., 2007), the resulting model has too many parameters to estimate with the current sample size. A more complex model that also included random slopes of the student-level variables (Aguinis et al., 2013) was also attempted but did not converge.

3. Results

In the results section, we first show the result of the three CFA models. Overall, the fit indices suggested that the proposed three latent teacher orchestration variables had close model fit. Secondly, we display the result of the multilevel SEM model concerning the first and second research questions.

3.1. Result of confirmatory factor analysis

CFA models were built to obtain validity evidence of the adequate internal structure of the teacher orchestration latent variables (i.e., "*Use of Instructional Videos*", "*Use of Workbooks*", and "*Preparation for High-stakes Tests*"). The model fit indices of the three CFAs are shown in Table 2. The significant chi-square tests showed that there was no exact model fit, but the close model fit was supported by three out of the four fit indices evaluated. We found that for all three latent variables, the close fit was supported by the CFI, TLI, and SRMR, but not supported by the RMSEA. This result was not surprising since RMSEA is sensitive to the small sample size ((Marsh et al., 2004; West et al., 2012) and small degrees of freedom (Shi et al., 2022), and we had only 93 teachers for estimating the teacher-level CFAs and the degrees of freedom were either 5 or 9.

Table 3 summarizes the factor loading, composite reliability (coefficient Ω), and average variance extracted (AVE) of the three CFA models. Firstly, the standardized factor loadings were all above 0.667 for the latent variable of "*Use of Instructional Videos*", which were higher than the recommended minimum in the social sciences, namely, 0.40 (Ford et al., 1986). The coefficient Ω was 0.924, which was above the recommended minimum of 0.70. The AVE was 0.672, which was higher than the recommended minimum, namely, 0.50 (Fornell & Larcker, 1981). It indicates that the six survey questions were well represented by the latent factors of "*Use of Instructional Videos*". Secondly, the standardized factor loadings are all above 0.580 for the "*Use of Workbooks*" variable. The composite reliability was 0.92, and the AVE was 0.461, which was slightly below but close to the 0.5 target. Thirdly, the standardized factor loadings were all above 0.570 for the "*Preparation for High-stakes Tests*" variable. The composite reliability was 0.866, and the AVE was 0.524. Given that the vast majority of our criteria established in the method section for evaluating dimensionality were met, we proceeded with estimating factor scores for each of the three latent variables using the regression method and inserting these factor scores into the dataset for use in the subsequent multilevel SEM analysis.

3.2. Result of multilevel structural equation model

The fit for the multilevel SEM model was excellent, with a non-significant chi-square test indicating exact fit ($\chi2$ (16) = 17.882, p = 0.331). Also, all the fit indices supported close fit with CFI equal to 1, TLI equal to 0.999, within-level SRMR equal to 0.012, between-level SRMR equal to 0.007, and RMSEA equal to 0.004. These fit indices indicated that the model matched the data well.

The intraclass correlation coefficient (ICC) is used in multilevel modeling to reflect the degree of clustering within groups (i.e., teachers in this study) and the degree of variability between groups (i.e., teachers). It is a statistic ranging from zero to one, and 0.5 was treated as a high ICC value, which indicated the need for multilevel analysis (Musca et al., 2011). In this study, the ICC for "*Total Quiz Scores after Watching Videos*" was 0.51, which means that 51% of the variance of students' total quiz scores was related to their

Table 2
Fit indices for the CFA models (n = 93).

Model	χ^2	*df*	*p value*	CFI	TLI	SRMR	RMSEA
Use of Instructional Videos	48.856	9	<0.001	0.912	0.913	0.050	0.215
Use of Workbook	12.173	5	0.032	0.947	0.949	0.049	0.124
Preparation for High-stakes Tests	26.398	9	0.002	0.932	0.934	0.059	0.142

Table 3
Standardized loadings for CFA models (n = 93).

Factors	Items	Standardized Loadings	*p* value
Use of Instructional Videos (Ω = 0.924, AVE = 0.672)			
	Use all or a portion of an instructional video as a bell ringer/warm up activity	0.817	<.001
	Present an instructional video to the whole class prior to your lesson instruction	0.894	<.001
	Present an instructional video after the lesson to reinforce concepts	0.876	<.001
	Use instructional videos for test review	0.745	<.001
	After showing a video, students try a workbook practice problem	0.893	<.001
	Students first try a workbook practice problem, then watch the corresponding video	0.667	<.001
Use of Workbook (Ω = 0.809, AVE = 0.461)			
	Assign workbook as a bell ringer/warm up activity	0.719	<.001
	Assign workbook to groups/centers	0.580	<.001
	Use workbook as a diagnostic tool during class	0.768	<.001
	Use workbook for an opening for a lesson	0.660	<.001
	Use workbook as an assessment after a lesson	0.652	<.001
Preparation for High-Stakes Tests (Ω = 0.866, AVE = 0.524)			
	I communicated the state standards of EOC with students	0.707	<.001
	I taught students to use test-taking strategies	0.824	<.001
	I had students review questions from previous years' EOC assessments	0.634	<.001
	I let students practice the type of questions that they are likely to see on the EOC	0.849	<.001
	I asked students to review and correct all practice problems	0.570	<.001
	I had students review incorrect answers to questions	0.717	<.001

Note. Ω is the coefficient omega, which is an estimate of composite reliability, and AVE refers to average variance extracted.

teachers. Likewise, the ICC value for "*EOC Scores*" was 0.59. It showed that 59% of the variability in students' EOC scores was also related to their teachers. In other words, both the "*Total Quiz Scores after Watching Videos*" and "*EOC Scores*" were similar for students who shared the teachers but varied for students with different teachers. The ICC showed that using multilevel SEM was appropriate.

3.2.1. Relationship between student SRL and student achievement

Overall, there were positive relationships between student SRL and student achievement. Table 4 shows the standardized coefficients for regressing "*Total Quiz Scores after Watching Video*", which corresponded to a proximal outcome based on the formative assessment. Table 5 shows the standardized coefficients for regressing "*EOC Scores*", which is the distal outcome based on the summative assessment. Table 6 shows the standardized coefficients for regressing "*Watched Recommended Videos*" on predictors.

The result from the within-level (student level) model regarding "*Total Quiz Scores after Watching Video*" (Table 4) showed that three out of five SRL indicators were positively related to this proximal student achievement: 1) "*Watched Recommended Videos*" ($\boldsymbol{\beta} = 0.036$, $p = 0.004$), 2) "*Answered Quizzes after Watching a Video*" ($\boldsymbol{\beta} = 0.935$, $p < 0.001$), and 3) "*Reviewed Incorrect Questions*" ($\boldsymbol{\beta} = 0.034$, $p = 0.002$). Specifically, the students who showed a high degree of SRL skills in watching a larger number of recommended videos, answering more quizzes after watching a video, and reviewing more incorrect answers were associated with having higher "*Total Quiz*

Table 4
Standardized coefficients for regressing "total quiz scores after watching video".

Levels	Explanatory Variables	Estimate	SE	*p* value
Within ($R^2 = 0.925$)	**Previous FSA Scores**	0.018*	0.005	<0.001
	Watched Recommended Videos	0.036*	0.013	0.004
	Answered Quizzes after Watching a Video	0.935*	0.012	<0.001
	Watched Videos	−0.006*	0.003	0.034
	Watched Solution Videos	−0.004	0.006	0.516
	Reviewed Incorrect Questions	0.034*	0.011	0.002
	Cross-level interactions:			
	Watched Solution Videos × Use of Workbook	−0.004	0.006	0.507
	Watched Solution Videos × Use of Instructional Videos	−0.004	0.007	0.599
	Watched Solution Videos × Preparation for High-stakes Tests	−0.003	0.006	0.666
	Watched Solution Videos × Use VLE as Remediation Tool	−0.004	0.006	0.442
	Reviewed Incorrect Questions × Use of Workbook	0.009	0.018	0.604
	Reviewed Incorrect Questions × Use of Instructional Videos	−0.023	0.016	0.150
	Reviewed Incorrect Questions × Preparation for High-stakes Tests	0.008	0.013	0.530
	Reviewed Incorrect questions × Use VLE as Remediation Tool	0.008	0.015	0.603
Between ($R^2 = 0.879$)	Mean Previous FSA scores	−0.029	0.032	0.363
	Mean Watched Recommended Videos	0.921*	0.037	<0.001
	Use of Workbook	0.023	0.048	0.632
	Use of Instructional Videos	0.000	0.032	0.992
	Preparation for High-stakes Tests	0.071	0.039	0.069
	Use VLE as Remediation Tool	0.031	0.041	0.458

Table 5
Standardized Coefficients for Regressing "EOC Scores".

Levels	Explanatory Variables	Estimate	SE	*p* value
Within	**Previous FSA scores**	0.584*	0.015	<0.001
($R^2 = 0.382$)	Total Quiz Scores after Watching Videos	0.053	0.038	0.158
	Answered Quizzes after Watching a Video	−0.039	0.039	0.314
	Answered Quizzes without Watching a Video	0.075*	0.015	<0.001
	Watched Videos	0.084*	0.013	<0.001
	Watched Solution Videos	0.034*	0.008	<0.001
	Reviewed Incorrect Questions	0.028*	0.010	0.005
	Cross-level interactions:			
	Watched Solution Videos × Use of Workbook	0.002	0.009	0.804
	Watched Solution Videos × Use of Instructional Videos	0.001	0.012	0.917
	Watched Solution Videos × Preparation for High-stakes Tests	−0.007	0.008	0.370
	Watched Solution Videos × Use VLE as Remediation Tool	0.011	0.008	0.205
	Reviewed Incorrect Questions × Use of Workbook	−0.008	0.011	0.482
	Reviewed Incorrect Questions × Use of Instructional Videos	0.037*	0.015	0.017
	Reviewed Incorrect Questions × Preparation for High-stakes Tests	−0.010	0.011	0.352
	Reviewed Incorrect Questions × Use VLE as Remediation Tool	−0.008	0.012	0.487
Between	**Mean Previous FSA Scores**	0.953*	0.016	<0.001
($R^2 = 0.899$)	**Mean Total Quiz Scores after Watching Videos**	0.093*	0.037	0.012
	Use of Workbook	0.030	0.045	0.508
	Use of Instructional Video	−0.037	0.053	0.479
	Preparation for High-stakes Tests	−0.005	0.034	0.883
	Use as Remediation Tool	0.043	0.033	0.187

Table 6
Standardized coefficients for regressing "watched recommended videos".

Levels	Predictors	Estimate	SE	*p* value
Within	**Treatment Status**	−0.057*	0.016	0.001
($R^2 = 0.316$)	Previous FSA Scores	0.002	0.010	0.834
	Answered Quizzes after Watching a Video	0.541*	0.038	<0.001
	Watched Videos	0.063*	0.014	<0.001
Between	Mean Previous FSA Scores	−0.033	0.083	0.691
($R^2 = 0.073$)	Use of Workbook	0.027	0.175	0.879
	Use of Instructional Video	0.183	0.109	0.093
	Preparation for High-stakes Tests	0.158*	0.076	0.037
	Use VLE as Remediation Tool	−0.044	0.073	0.544

Scores after Watching Video" after controlling for the previous achievement. The "*Previous FSA scores*" were positively related to "*Total Quiz Scores after Watching Video*" ($\boldsymbol{\beta} = 0.018$, $p < 0.001$) but the association was weak. No statistically significant relationship was found between "*Watched Solution Videos*" and "*Total Quiz Scores after Watching Video*". However, one student SRL indicator, "*Watched Videos*" ($\boldsymbol{\beta} = -0.006$, $p = 0.034$) was negatively related to "*Total Quiz Scores after Watching Video*", indicating that students who watched more videos performed worse on the formative CYU assessments. One potential explanation could be that the lower achievement students tended to watch more videos or were required to watch more videos by their teachers.

Likewise, the result from the within-level model regarding "*EOC Scores*" (Table 5) showed that four out of five SRL indicators were positively associated with "*EOC Scores*": 1) "*Watched Videos*" ($\boldsymbol{\beta} = 0.084$, $p < 0.001$), 2) "*Watched Solution Videos*" ($\boldsymbol{\beta} = 0.034$, $p < 0.001$), 3) "*Reviewed Incorrect Questions*" ($\boldsymbol{\beta} = 0.028$, $p = 0.005$), and 4) "*Answered Quizzes without Watching a Video*" ($\boldsymbol{\beta} = 0.075$, $p < 0.001$). In other words, students who showed a high degree of SRL skills by watching more videos and solution videos, reviewing more incorrect questions, and answering more quizzes without watching a video tended to have higher Algebra 1 EOC scores. There was no statistically significant relationship between "*Answered Quizzes after Watching a Video*" and "*EOC Scores*". As expected, the findings show that obtaining higher "Previous FSA scores" was strongly associated with higher "*EOC Scores*" ($\boldsymbol{\beta} = 0.584$, $p < 0.001$), because these are both measures of mathematics ability aligned to Florida state standards. However, no statistically significant relationship was found between "*Total Quiz Scores after Watching Video*" and "*EOC scores*". One potential explanation is that the variance in "*Total Quiz Scores after Watching Video*" was explained by the student SRL indicators.

The result from the within-level model regarding "*Watched Recommendation Video*" (Table 6) suggested both SRL indicators could explain the differences regarding "*Watched Recommendation Video*": "*Watched Videos*" ($\boldsymbol{\beta} = 0.541$, $p < 0.001$), and "*Answered Quizzes after Watching a Video*" ($\boldsymbol{\beta} = 0.063$, $p < 0.001$). In other words, students who watched a larger number of videos and took a larger number of quizzes after watching videos were likely to watch more recommended videos. We also found that students in the treatment group were less likely to watch recommended videos ($\boldsymbol{\beta} = -0.057$, $p = 0.001$) than students in the control group. This can be explained by students perceived larger effort to watch a treatment video, which could send students back to previous topics, than a control video, which is always the next topic. Lastly, no statistically significant relationship was found between "*Previous FSA Scores*" and "*Watched*

Recommendation Video", which indicated that watching recommended videos did not depend on students' previous achievement.

3.2.2. The role of teacher orchestration

The between-level (teacher level) was built to investigate the direct association between teacher orchestration and student achievement, and the interactions terms explore whether the relationship between student SRL and their achievement depends on teacher orchestration in the VLE. Overall, we did not observe a direct association between teacher orchestration and student achievement in this study. However, we found some evidence that teacher orchestration moderated the relationship between student SRL and their achievement.

The result from the between-level model regarding "*Total Quiz Scores after Watching Video*" (Table 4) showed that none of the teacher orchestration variables was directly related to "*Mean Total Quiz Scores after Watching Videos*", nor the "*Mean Previous FSA scores*". However, the "*Mean Watched Recommended Videos*" was strongly positively related to "*Total Quiz Scores after Watching Videos*" ($\boldsymbol{\beta} = 0.921$, $p = 0.037$), which showed that the teachers whose students watched more recommended videos on average obtained higher mean total quiz scores.

Similarly, the result from the between-level model showed that none of the teacher orchestration variables was directly related to "*EOC Scores*" (Table 5). However, both "*Mean Previous FSA scores*" ($\boldsymbol{\beta} = 0.953$, $p < 0.001$) and "*Mean Total Quiz Scores after Watching Video*" ($\boldsymbol{\beta} = 0.093$, $p = 0.012$) were significantly positively associated with "*EOC scores*". It meant that the teachers whose students had higher previous achievement and scored higher on formative assessments on average, were associated with higher mean EOC scores.

The result from the between-level model regarding "*Watched Recommendation Video*" showed there was a positive direct association between one teacher orchestration variable, "*Preparation for High-stakes Tests*" and on SRL indicator, "*Watched Recommendation Video*" ($\boldsymbol{\beta} = 0.037$, $p = 0.017$). It indicated that teachers who used the VLE as a tool for preparing their students to take the EOC test positively influenced the mean usage of the video recommendation system in the classroom. This finding also revealed a statistically significant indirect relationship between "*Preparation for High-stakes Tests*" and "*Total Quiz Scores after Watching Video*" through the mediation of "*Watched Recommendation Video*" (indirect relationship = 0.145, SE = 0.071, $p = 0.040$), since there was a significant association between "*Watched Recommendation Video*" and "*Total Quiz Scores after Watching Video*" at within-level. We did not observe statistically signification results with the rest three teacher orchestration variables and "Previous FSA Scores" in terms of their direct associations with "*Watched Recommendation Videos*".

The results from testing the interaction terms (Table 4) indicated that none of the interactions between SRL indicators and teacher orchestration were statistically significant for "*Total Quiz Scores after Watching Videos*". It suggested that the relationship between student SRL and "*Total Quiz Scores after Watching Videos*" did not depend on teacher orchestration. However, the interaction terms for "*EOC Scores*" (Table 5) showed one of the eight cross-level interactions between SRL indicators and teacher orchestration was significant. Specifically, the relationship between "*Reviewed Incorrect Questions*" and "*EOC Scores*" was moderated by "*Use of Instructional Videos*" ($\boldsymbol{\beta} = 0.037$, $p = 0.017$). It reflected that the positive relationship between EOC scores and student reviewing incorrect questions increased as the use of instructional videos by the teacher increased.

4. Discussion and conclusion

To the best of our knowledge, this study is among the first to investigate the interaction between teacher orchestration and student SRL in online learning environments. We used a large data set of 6,174 students taught by 93 teachers from 49 schools and applied multilevel SEM to examine the interactions. This work adds to the literature by advancing our understanding of how teacher orchestration may moderate the relationship between student SRL and student achievement in online learning environments.

4.1. Self-regulated learning

While a number of previous studies focused on either student SRL in mathematics learning with technology or teacher orchestration with mathematics VLE, this study investigated the interplay between student SRL and teacher orchestration in algebra classrooms. The results have extended our understanding of student SRL use of videos and quizzes in a mathematics VLE. This study carefully examined several SRL indicators regarding the use of videos and quizzes, and showed that "*Watched Recommended Videos*", "*Watched Videos*" and "*Answered Quizzes after Watching a Video*", were significantly associated with student achievement. Specially, students who have demonstrated a high degree of SRL strategies on a VLE are more likely to have higher achievements. Watching videos and answering quizzes explicitly represent students' intrinsic willingness to learn (Zimmerman, 1983). The strong relationship between answering quizzes and the formative measure of student achievement is also in line with previous research that practice matters (Bartelet et al., 2016; Rienties et al., 2019).

Meanwhile, the role of student SRL indicators is different regarding their relationship with formative and summative measures of student achievement. If a student answers quizzes after watching a video, they are more likely to obtain a higher score because they obtained information about the concepts addressed by the quiz immediately prior to completing the quiz. However, students are less likely to answer a quiz after watching a video if a student feels confident about the topic or if they are re-watching a video. Interestingly, "*Answered Quizzes after Watching a Video*" was significantly associated with "*Total Quiz Scores after Watching a Video*", yet insignificantly associated with "*EOC Scores*". One potential explanation could be that we included "*Answered Quizzes after Watching a Video*" as a covariate when regressing "*EOC Scores*", thus that the "*Answered Quizzes after Watching a Video*" served as a mediator between "*Total Quiz Scores after Watching a Video*" and "*EOC Scores*".

"*Watched Recommended Videos*" and "*Reviewed Incorrect Questions*" are considered to be helpful SRL actions, which contribute to student achievement in a VLE. This is also in line with previous studies on the critical role of SRL in online learning (Azevedo & Hadwin, 2005; Lehmann et al., 2014; Rienties et al., 2019). The fact that the positive relationship was identified within a large-scale study has practical implications. The positive relationship between student SRL and student achievement suggests the importance of SRL in VLE. Instructors shall be encouraged to adapt various strategies to prompt student SRL skills. The learning environment design is also expected to incorporate such components to encourage student SRL strategies. Future research could continue to explore the impact of teacher orchestration regarding promoting students' SRL skills on student outcomes.

4.2. Teacher orchestration

This study identified two key components that connect teacher orchestration to student achievement, namely, "*Watched Recommended Videos*" and "*Reviewed Incorrect Questions*", with the first connected through mediation and the second through moderation. Considering the mediation through "*Watched Recommended Videos*", although we did not find direct associations between teacher orchestration and formative and summative measures of student achievement in this study, teacher orchestration with respect to "*Preparing for High-stakes tests*" was positively associated with "*Watched Recommended Videos*". Therefore, there was a statistically significant indirect relationship between "*Preparing for High-stakes tests*" and "*Total Quiz Scores after Watching a Video*" through the mediation of "*Watched Recommended Videos*". In other words, when the teachers use the VLE to prepare their students for a high-stake test, their students tended to watch more recommended videos and had higher scores on the quizzes.

Considering the moderation through "*Reviewed Incorrect Questions*", the results showed that teacher orchestration as "*Use of Instructional Videos*" moderates the relationship between "*Reviewed Incorrect Questions*" and "EOC Scores". The Algebra VLE is extensively available in school districts in Florida, however, to what extent teachers use the platform varies among schools. Our analysis results indicated that when teachers use the instructional videos more often, the relationship between SRL use and student achievement is stronger. These results support the conclusion that, when teachers enable and adapt learning activities to a dynamic classroom context (Dillenbourg, 2013), there is a stronger relationship between the SRL use and student achievement. As the teachers enacts a variety of strategies is likely to come with some overt reflection on learning and concepts, which in turn may encourage students to practice some meta-cognition and reflection in the form of reviewing incorrect responses. Future research could investigate the impact of teachers prompting students' SRL use and reflections like reviewing the incorrect answers, as well as expressing learning goals while enabling and adapting learning activities in a VLE.

Meanwhile, the relationship we observed between "*Preparing for High-stakes tests*" and "*Watched Recommended Videos*" supports the conclusion that learning-objectives-related instrumental orchestration (Drijvers et al., 2020) is potentially helpful to improve student achievement. Teachers may be encouraged to incorporate such strategies into their practice as these strategies are manipulable by the teacher more than SRL. Previous research focused on experimental studies comparing achievement of students using a VLE versus students in traditional learning environments (Pardos et al., 2011; Steenbergen-Hu & Cooper, 2013). This study adds to the literature by showing moderated and mediated influence of teacher orchestration on student achievement using secondary data from a large-scale field experiment. The field setting allows the study to have strong external validity (Shadish et al., 2002), and the findings and implications may generalize to other VLE that have wide adoption in schools. Further studies using other VLE adopted by entire school systems would provide additional evidence of generalizability.

4.3. Limitations and future research

There are a few limitations for this study: First, it is a correlational study because student SRL and teacher orchestration were measured rather than manipulated. Therefore, there is potential for confounding due to variables not included in the model. The internal validity (Shadish et al., 2002) of the current results are strengthened by the inclusion of the previous FSA mathematics assessment as a covariate, which served as a proxy of a pre-test for both formative and summative outcomes. The very large proportion of variance extracted also serves as internal validity evidence, because small residuals indicate that omitted confounders are less likely. Also, the internal validity of the regression of the formative outcome on watched recommended video was strengthened by the inclusion of the treatment indicator as a covariate, which is related to watched recommended video but only indirectly related to the formative outcome, thus serving as a proper instrumental variable (Angrist, 2006; Bollen, 2012). A second limitation of the study is that we focused on student use of videos and quizzes only when measuring their SRL behaviors in the Algebra VLE, excluding other SRL strategies such as help seeking or peer learning. Future research may include more SRL strategies and investigate whether teacher instrumental orchestration continues to moderate the relationship between student SRL and their achievement. A third limitation is that the measurement of teacher instrumental orchestration was based on indicators taken from a survey, but not on a psychometric scale specifically designed to measure teacher orchestration. Although we provided evidence of content validity and internal structure validity of the measures using CFA, there is still a need for the development of a general measurement of teacher orchestration.

Author statement

Walter L. Leite: Conceptualization, Supervision, Methodology, Formal analysis, Writing – original draft, Writing – review & editing Huan Kuang: Methodology, Formal analysis, Writing – original draft, Writing – review & editing Zeyuan Jing: Formal analysis Wanli Xing: Writing – original draft, Writing – review & editing Catherine Cavanaugh: Writing – original draft, Writing – review & editing, A. Corinne Huggins-Manley: Writing – original draft, Writing – review & editing

Acknowledgements

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305C160004to University of Florida. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education.

References

Aguinis, H., Gottfredson, R. K., & Culpepper, S. A. (2013). Best-practice recommendations for estimating cross-level interaction effects using multilevel modeling. *Journal of Management, 39*(6), 1490–1528. https://doi.org/10.1177/0149206313478188
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.
Alavi, H. S., Dillenbourg, P., & Kaplan, F. (2009). Distributed awareness for class orchestration. In *European Conference on technology enhanced learning* (pp. 211–225). Berlin, Heidelberg: Springer.
American Educational Research Association. (2014). American Psychological association, & national council on measurement in education. In *Standards for educational and psychological testing*. American Educational Research Association.
Angrist, J. D. (2006). Instrumental variables methods in experimental criminological research: what, why and how. *Journal of Experimental Criminology, 2*(1), 23–44. https://doi.org/10.1007/s11292-005-5126-x
Artigue, M. (2002). Learning mathematics in a CAS environment: The genesis of a reflection about instrumentation and the dialectics between technical and conceptual work. *International Journal of Computers for Mathematical Learning, 7*(3), 245–274.
Azevedo, R., & Hadwin, A. F. (2005). Scaffolding self-regulated learning and metacognition: Implications for the design of computer-based scaffolds. *Instructional Science, 33*(5/6), 367–379.
Bacon, D. R., Sauer, P. L., & Young, M. (1995). Composite reliability in structural equations modeling. *Educational and Psychological Measurement, 55*(3), 394–406.
Bandalos, D. L. (2018). *Measurement theory and applications for the social sciences*. The Guilford Press.
Bartelet, D., Ghysels, J., Groot, W., Haelermans, C., & Maassen van den Brink, H. (2016). The differential effect of basic mathematics skills homework via a web-based intelligent tutoring system across achievement subgroups and mathematics domains: A randomized field experiment. *Journal of Educational Psychology, 108*(1), 1–20. https://doi.org/10.1037/edu0000051
Bollen, K. A. (2012). Instrumental variables in sociology and the social sciences. *Annual Review of Sociology, 38*(1), 37–72. https://doi.org/10.1146/annurev-soc-081309-150141
du Boulay, B. (2019). Escape from the Skinner Box: The case for contemporary intelligent learning environments. *British Journal of Educational Technology, 50*(6), 2902–2919. https://doi.org/10.1111/bjet.12860
Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education, 27*, 1–13.
Broadbent, J., Sharman, S., Panadero, E., & Fuller-Tyszkiewicz, M. (2021). How does self-regulated learning influence formative assessment and summative grade? Comparing online and blended learners. *The Internet and Higher Education, 50*, Article 100805. https://doi.org/10.1016/j.iheduc.2021.100805
Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows*. Sage Publications.
Carson, A. D. (2011). Predicting student success from the LASSI for learning online (LLO). *Journal of Educational Computing Research, 45*(4), 399–414.
Chakraborty, N., Roy, S., Leite, W. L., Faradonbeh, M. K. S., & Michailidis, G.. The effects of a personalized recommendation system on students' high-stakes achievement scores: A field experiment. https://educationaldatamining.org/EDM2021/virtual/static/pdf/EDM21_paper_87.pdf.
Chang, M. M. (2007). Enhancing web-based language learning through self-monitoring. *Journal of Computer Assisted Learning, 23*(3), 187–196.
ChanLin, L. J. (2012). Learning strategies in web-supported collaborative project. *Innovations in Education & Teaching International, 49*(3), 319–331.
Cho, M. H., & Shen, D. (2013). Self-regulation in online learning. *Distance Education, 34*(3), 290–301.
Dillenbourg, P. (2013). Design for classroom orchestration. *Computers & Education, 69*, 485–492. https://doi.org/10.1016/j.compedu.2013.04.013
Dillenbourg, P. (2021). Classroom analytics: Zooming out from a pupil to a classroom. In *OECD digital education Outlook 2021: Pushing the Frontiers with Artificial Intelligence* (pp. 105–122). Paris: Blockchain and Robots, OECD Publishing. https://doi.org/10.1787/589b283f-en.
Drijvers, P., Doorman, M., Boon, P., Reed, H., & Gravemeijer, K. (2010). The teacher and the tool: Instrumental orchestrations in the technology-rich mathematics classroom. *Educational Studies in Mathematics, 75*(2), 213–234. https://doi.org/10.1007/s10649-010-9254-5
Drijvers, P., Grauwin, S., & Trouche, L. (2020). When bibliometrics met mathematics education research: The case of instrumental orchestration. *Zdm, 52*(7), 1455–1469. https://doi.org/10.1007/s11858-020-01169-3
Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*(2), 121–138.
Fancsali, S. E., Yudelson, M. V., Berman, S. R., & Steven, R. (2018). Intelligent instructional hand offs. In *Proceedings of 11th International Conference on educational data mining*. New York: Templeton Landing in Buffalo.
Florida Department of Education. (2021). *Statewide assessment guide 2020-2021*. Florida Department of Education. https://www.fldoe.org/core/fileparse.php/5663/urlt/swapig.pdf.
Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology, 39*(2), 291–314.
Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39–50.
Fuchs, L. S., Fuchs, D., Prentice, K., Burch, M., Hamlett, C. L., Owen, R., & Schroeter, K. (2003). Enhancing third-grade student'mathematical problem solving with self-regulated learning strategies. *Journal of Educational Psychology, 95*(2), 306.
E-learning and games for training, education, health and sports. In Göbel, S., Müller, W., Urban, B., & Wiemeyer, J. (Eds.), *7th International Conference, Edutainment 2012, and 3rd International Conference, GameDays 2012, Darmstadt, Germany, September 18-20, 2012, Proceedings*, (2012). Berlin: Springer Berlin Heidelberg.
Hamalainen, R., & Oksanen, K. (2021). Challenge of supporting vocational learning: Empowering collaboration in a scripted 3D game – how does teachers' real-time orchestration make a difference? *Computers & Education, 59*(2), 281–293. https://doi.org/10.1016/j.compedu.2012.01.002
Holstein, K., McLaren, B. M., & Aleven, V. (2019). Co-designing a real-time classroom orchestration tool to support teacher–AI complementarity. *Journal of Learning Analytics, 6*(2), 27–52. https://doi.org/10.18608/jla.2019.62.3
Jackson, D. L. (2003). Revisiting sample size and number of parameter estimates: Some support for the N:q hypothesis. *Structural Equation Modeling, 10*(1), 128–141. https://doi.org/10.1207/S15328007SEM1001_6
Joyce-Gibbons, A. (2014). *An exploratory study of teacher orchestration of collaborative Mathematics tasks in relation to learning and interaction in primary schools, Doctoral thesis*. Durham University. http://etheses.dur.ac.uk/9464/.
Kim, D., Yoon, M., Jo, I. H., & Branch, R. M. (2018). Learning analytics to support self-regulated learning in asynchronous online courses: A case study at a women's university in South Korea. *Computers & Education, 127*, 233–251.
Lehmann, T., Hähnlein, I., & Ifenthaler, D. (2014). Cognitive, metacognitive and motivational perspectives on preflection in self-regulated online learning. *Computers in Human Behavior, 32*, 313–323.
Leite, W., L., Roy, S., Chakraborty, N., Michailidis, G., Huggins-Manley, A., D'Mello, S., K., … Jing, Z. (2022). A novel video recommendation system for algebra: An effectiveness evaluation study. *LAK22: 12th International Learning Analytics and Knowledge Conference (LAK22)*. Online, USA: Society of Learning Analytics.

Leite, W. L., Xing, W., Fish, G., & Li, C. (2022). Teacher strategies to use virtual learning environments to facilitate algebra learning during school closures. *Journal of Research on Technology in Education*. https://doi.org/10.1080/15391523.2022.2110335

Leite, Walter, L., Cetin-Berber, Dee, D., Collier, Zachary, K., & Beal, Carole, R. (2019). The relationship between Algebra Nation usage and high-stakes test performance for struggling students. *Journal of Computer Assisted Learning, 35*(5), 569–581. https://doi.org/10.1111/jcal.12360

Leite, W. L., Kuang, H., Shen, Z., Chakraborty, N., Michailidis, M., D'Mello, S., & Xing, W. (2022). Heterogeneity of Treatment Effects of a Video Recommendation System for Algebra. *Ninth ACM Conference on Learning @Scale, (L@S'22),*. https://doi.org/10.1145/3491140.3528275

Li, S., Du, H., Xing, W., Zheng, J., Chen, G., & Xie, C. (2020). Examining temporal dynamics of self-regulated learning behaviors in STEM learning: A network approach. *Computers & Education, 158*. https://doi.org/10.1016/j.compedu.2020.103987

Looi, C.-K., & Song, Y. (2013). Orchestration in a networked classroom: Where the teacher's real-time enactment matters. *Computers & Education, 69*, 510–513.

Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology, 106*(4), 901–918.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2

McCardle, L., & Hadwin, A. F. (2015). Using multiple, contextualized data sources to measure learners' perceptions of their self-regulated learning. *Metacognition and Learning, 10*(1), 43–75.

Mitten, C., Collier, Z., K., & Leite, W., L. (2021). Online resources for mathematics: Exploring the relationship between teacher use and student performance. *Investigations in Mathematics Learning, 13*(3), 249–266. https://doi.org/10.1080/19477503.2021.1906041

Morera, O. F., & Stokes, S. M. (2016). Coefficient α as a measure of test score reliability: Review of 3 popular misconceptions. *American Journal of Public Health, 106*(3), 458–461.

Muñoz-Cristóbal, J. A., Jorrín-Abellán, I. M., Asensio-Pérez, J. I., Martínez-Monés, A., Prieto, L. P., & Dimitriadis, Y. (2015). Supporting teacher orchestration in ubiquitous learning environments: A study in Primary education. *IEEE Transactions on Learning Technologies, 8*(1), 83–97.

Musca, S. C., Kamiejski, R., Nugier, A., Méot, A., Er-Rafiy, A., & Brauer, M. (2011). Data with hierarchical structure: Impact of intraclass correlation and sample size on type-I error. *Frontiers in Psychology, 2*, 74.

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research, 22*(3), 376–398. https://doi.org/10.1177/0049124194022003006

Muthén, & Muthén. (2017). *Mplus*. Muthén & Muthén Version 8.1. .

Pardos, Z. A., Dailey, M. D., & Heffernan, N. T. (2011). Learning what works in ITS from non-traditional randomized controlled trial data. *International Journal of Artificial Intelligence in Education, 21*, 47–63.

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and Prescriptions. *Multivariate Behavioral Research, 42*(1), 185–227. https://doi.org/10.3389/fpsyg.2011.00074

Prieto, L. P., Dlab, M. H., Gutiérrez, I., Abdulwahed, M., & Balid, W. (2011). Orchestrating technology enhanced learning: A literature review and a conceptual framework. *International Journal of Technology Enhanced Learning, 3*(6), 583–598.

Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement, 21*(2), 173–184. https://doi.org/10.1177/01466216970212006

Raykov, T. (1998). Coefficient Alpha and composite reliability with interrelated nonhomogeneous items. *Applied Psychological Measurement, 22*(4), 375–385. https://doi.org/10.1177/014662169802200407

Raykov, T. (2009a). Interval estimation of revision effect on scale reliability via covariance structure modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 16*(3), 539–555.

Raykov, T. (2009b). Evaluation of scale reliability for unidimensional measures using latent variable modeling. *Measurement and Evaluation in Counseling and Development, 42*(3), 223–232.

Rienties, B., Tempelaar, D., Nguyen, Q., & Littlejohn, A. (2019). Unpacking the intertemporal impact of self-regulation in a blended mathematics environment. *Computers in Human Behavior, 100*, 345–357.

Rodríguez-Triana, M. J., Martínez-Monés, A., Asensio-Pérez, J. I., & Dimitriadis, Y. (2015). Scripting and monitoring meet each other: Aligning learning analytics and learning design to support teachers in orchestrating CSCL situations. *British Journal of Educational Technology, 46*(2), 330–343.

Roschelle, J., Feng, M., Murphy, R. F., & Mason, C. A. (2016). Online mathematics homework increases student achievement. *AERA Open, 2*(4), 1–12. https://doi.org/10.1177/2332858416673968

Rosenman, R., Tennekoon, V., & Hill, L. G. (2011). Measuring bias in self-reported data. *International Journal of Behavioural and Healthcare Research, 2*(4), 320–332.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

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

Shen, S., Mostafavi, B., Barnes, T., & Chi, M. (2018). Exploring induced pedagogical strategies through a Markov decision process framework: Lessons learned. *Journal of Educational Data Mining, 10*(3), 27–68.

Shi, D., DiStefano, C., Maydeu-Olivares, A., & Lee, T. (2022). Evaluating SEM model fit with small degrees of freedom. *Multivariate Behavioral Research, 57*(2–3), 179–207. https://doi.org/10.1080/00273171.2020.1868965

Steenbergen-Hu, S., & Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. *Journal of Educational Psychology, 105*(4), 970–987. https://doi.org/10.1037/a0032447

Wang, M. C. (1983). Development and consequences of students' sense of personal control. In J. M. Levine, & M. C. Wang (Eds.), *Teacher and student perceptions: Implications for learning* (pp. 213–247). Hillsdale, NJ: Erlbaum.

Wang, C. H., Shannon, D. M., & Ross, M. E. (2013). Students' characteristics, self-regulated learning, technology self-efficacy, and course outcomes in online learning. *Distance Education, 34*(3), 302–323.

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 209–231). The Guilford Press.

Winne, P. H. (2010). Improving measurements of self-regulated learning. *Educational Psychologist, 45*(4), 267–276.

Wong, J., Baars, M., Davis, D., Van Der Zee, T., Houben, G. J., & Paas, F. (2019). Supporting self-regulated learning in online learning environments and MOOCs: A systematic review. *International Journal of Human-Computer Interaction, 35*, 356–373.

Wong, J., Khalil, M., Baars, M., de Koning, B. B., & Paas, F. (2019). Exploring sequences of learner activities in relation to self-regulated learning in a massive open online course. *Computers & Education, 140*, Article 103595. https://doi.org/10.1016/j.compedu.2019.103595

Xue, K., Huggins-Manley, A., C., & Leite, W., L. (2021). Semisupervised learning method to adjust biased item difficulty estimates caused by nonignorable missingness in a virtual learning environment. *Educational and Psychological Measurement*. https://doi.org/10.1177/00131644211020494

Zimmerman, B. J. (1983). Social learning theory: A contextualist account of cognitive functioning. In C. J. Brainerd (Ed.), *Recent advances in cognitive-developmental theory* (pp. 1–50). Springer.

Zimmerman, B. J., & Pons, M. M. (1986). Development of a structured interview for assessing student use of self-regulated learning strategies. *American Educational Research Journal, 23*(4), 614–628.