Simulating the dynamics of self-regulation, emotion, grit, and student performance in cyber-learning environments

Janice W. Kooken¹ · Raafat Zaini¹ · Ivon Arroyo²

Received: 28 April 2020 / Accepted: 9 November 2020 / Published online: 5 February 2021
© Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

This research presents the results of development and validation of the Cyclical Self-Regulated Learning (SRL) Simulation Model, a model of student cognitive and metacognitive experiences learning mathematics within an intelligent tutoring system (ITS). Patterned after Zimmerman and Moylan's (2009) Cyclical SRL Model, the Simulation Model depicts a feedback cycle connecting forethought, performance and self-reflection, with emotion hypothesized as a key determinant of student learning. A mathematical model was developed in steps, using data collected from students during their sessions within the ITS, developing solutions using structural equation modeling, and using these coefficients to calibrate a System Dynamics (SD) Simulation model. Results provide validation of the Cyclical SRL Model, confirming the interplay of grit, emotion, and performance in the ITS. The Simulation Model enables mathematical simulations depicting a variety of student background types and intervention styles and supporting deeper future explorations of dimensions of student learning.

Keywords Grit · Intelligent tutoring system · System dynamics · Structural equation modeling · Simulation science · Self-regulated learning

✉ Janice W. Kooken
jwkooken@gmail.com

Raafat Zaini
rzaini@wpi.edu

Ivon Arroyo
ivon@cs.umass.edu

¹ Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609, USA

² University of Massachusetts Amherst, 140 Governors Dr, Amherst, MA 01003, USA

Introduction

Students in personalized learning environments including intelligent tutoring systems (ITS) receive experiences custom designed to connect precisely with where the student sits on a continuum of states and traits including ability level, content knowledge, help seeking, and self-regulation (Arroyo et al. 2014). The Zimmerman and Moylan (2009) Cyclical self-regulated learning (SRL) model represents a personal feedback loop comprising three phases, the forethought phase, the performance phase and the self-reflection phases which provides the foundation in learning sciences for MathSpring (Arroyo et al. 2014), an adaptive tutoring system for K-12 students learning mathematics. While interacting with the tutoring system, students also experience emotional states from confidence and enjoyment to frustration and hopelessness. Research has shown that how students feel during instructional activities influences achievement outcomes (Arroyo et al. 2009, 2011; Woolf et al. 2010). ITS are carefully calibrated to support student growth, progressing through more advanced content and more effective academic functioning.

Many theories of learning and emotions converge in the intelligent tutoring environment with each intervention, nudge, and hint carefully designed and supported by research to efficiently and effectively support learning. The extant literature provides evidence and theory on cognitive and metacognitive constructs and their effect on performance (e.g., Muenks et al. 2017; Sottolare et al. 2013), but the challenge of understanding how they all work together - all within a single student during a learning experience - has not been explored. In particular, little is known about what takes place systemically in the student as they participate in their “personal feedback loop” (Zimmerman and Moylan 2009) responding to prompts and supports from the tutoring system. Moreover, it is difficult to predict with accuracy a student’s response to a nudge in one node of the ITS with changes in other cognitive and metacognitive states. These questions provide the motivation for a mathematical model that utilizes the log data from an ITS, interconnected theory, and an automated simulation model.

System Dynamics (SD) represents a set of theory and techniques for modeling complex interconnected components and activities within a living dynamic entity to analyze how changes are felt across the system (Forrester 1968; Richardson 1991). SD was originally developed to model economic theories and logistics in manufacturing and industry, but it has since been shown very effective in advancing knowledge in business, management, higher education and organizational research (Zaini et al. 2016, 2017). We propose opening a window into a student’s experience within an ITS through simulations within a SD model using mathematical models built with path modeling, supporting the following goals:

1. We seek a deeper understanding of the interrelationships among learning processes by designing a mathematical model of the Cyclical SRL model, testing the validity of this model with real data, testing the sensitivity of the system to change, and using this model in simulations to bolster good decision making regarding improvements and interventions within the system.
2. We seek to use SD modeling to investigate the relationship between the Cyclical SRL model and emotions as measured by frustration and confidence.

This article contributes to the learning sciences in that no prior research has examined so many aspects of learning theory with real data from an ITS and simulated these outcomes within an SD model. The central theory conceptualizes a student’s experience in the ITS using

the Cyclical SRL Model (Zimmerman 2013), reflecting interactions of these phases with students' emotions experienced during learning activities, and motivational factors such as a student's grit. This research describes the process, results, and innovations of building a SD model using what we have learned from the 10 year use of an intelligent math tutoring system, the MathSpring cyberlearning environment.

Background

MathSpring: An intelligent tutoring platform

MathSpring (Arroyo et al. 2014) is designed for student use to provide “cognitive, metacognitive, and affective” content and supports to enhance learning of mathematics in an online platform. ITS platforms such as MathSpring are designed to personalize the learning experience utilizing measures of student learning behaviors to tailor ITS feedback and actions to each student's needs. The mathematical content of MathSpring aligns with the Common Core State Standards (National Governors Association Center for Best Practices and Council of Chief State School Officers 2010) for students in grades K-12. The Common Core State Standards were developed in the United States beginning in 2009 to codify one national set of content level standards for what students should be able to do by content area and grade level, first for Math and ELA/Literacy, and then for additional content areas. These content standards have been adopted in many states, and the state accountability policy and assessments are now aligned to these standards. To support student growth in these standards and the teacher's ability to track a student's progress towards these goals, the mathematical content of MathSpring includes items directly from the Common Core aligned state standardized tests. Items may also be generated by teachers and researchers. These problems are coded with measures representing content and difficulty level as identified with item response theory (Hambleton et al. 1991) within the ITS, resulting in an effective artificial intelligence based platform. MathSpring assesses both student cognitive skills and student emotions. As students solve mathematics problems, they receive support, feedback, worked out examples, tutorial videos and socioemotional support. Animated characters, called “learning companions” (Fig. 1) provide significant beneficial effects for most students and were shown more effective for lower achieving students and for female students in general (Arroyo 2009, 2011, 2013).

Behind the scenes, students are assessed at the mouse click level, as the system gathers extensive log data about them. Murray and Arroyo (2002) describe the use of effort-based tutoring (EBT) towards mastery learning, meaning the student engages cognitive processes to activate prior learning, applies this to current problems, and combines prior learning with newly gained knowledge to show evidence of ability to solve problems at a distinct difficulty level and content area. In addition to hints and videos that build learning of content, MathSpring also boosts motivation, effort, and effective use of hints through the EBT algorithm which provides affective and metacognitive prompts to encourage student persistence.

In MathSpring, mastery is an estimation of how much a student knows within a content domain. In general mastery is calculated as an average value of student performance on a problem set. MathSpring uses EBT models to optimize problem selection within content areas and to match student cognitive and metacognitive states, supporting the goal of mastery

The net of a triangular prism and some of its dimensions are shown in the diagram.

What is the total surface area, in square inches, of the triangular prism?

Submit Answer!

Step 1 Step 2

Let's start by finding the areas of the rectangular pieces of the prism's surface. All of the pieces are 12 inches long. One piece is 8 inches wide while the other two are 5 inches wide. We can add these widths together to create one large rectangle to make finding the area simpler.

Good job!

See how taking your time to work through these questions can make you get the right answer?

Fig. 1 MathSpring animated companion screen provides animated companions (bottom, right), that reflect students' emotion, along with hints, animated problems, audio help, worked-out examples, and video tutorials to aid students

defined as 95% probability of knowing the content (Arroyo et al. 2014). Mastery is calculated using Bayesian estimation and updated based upon a student's performance on each problem in a content area taking into account student engagement (Corbett and Anderson 1994). Higher mastery results from greater success, and mistakes lead to reduced mastery. Mastery is not updated when students seem to be disengaged, for instance when they guess quickly, meaning they answer quickly and incorrectly, or when they skip the problem without answering. This is done because such behavior creates the illusion that the student does not know when in fact they are disengaged (Arroyo et al. 2014). In addition, when students ask for hints before answering correctly, mastery increases by a smaller fraction than when they answer correctly without hints, indicating a likely increase in knowledge but incrementally less than if they did not ask for hints. Further, mastery and success work through a feedback mechanism, where a student's ability to succeed on a problem changes based upon level of mastery, holding other things constant (Corbett & Bhannagar 1997).

Log data from MathSpring are ideal to support research needed on integrated theories of learning, motivation, and emotions and to improve decision making on enhancing the adaptive technologies. Even though the use of an ITS flexibly supports research in content, problem types, scaffolds, interventions, and calibration, research is complex and costly to administer and analyze. There is much more to learn regarding how students learn, and which features represent optimal design. The rate at which we can learn from the prior discoveries can be accelerated through the use of simulations using an SD model. SD models can apply the vast warehouse of what we know and have learned to a mathematical model to simulate changes and decisions that will allow us to explore new theories. These contributions can support improvement in adaptive technologies for the design loop, the task loop, and the step loop, as described in Aleven, McLaughlin, Glenn, and Koedinger (2003).

Foundational to MathSpring lie the theories of how students learn mathematics which includes knowledge of the interrelationships between student attributes such as self-regulation, emotion, grit, expectation of success, and how these attributes accelerate or decelerate a student's progress towards mastery of a content area in mathematics (Arroyo et al. 2014). Each of these foundational theories are discussed briefly in the sections that follow.

Self-regulated learning

Self-regulation refers to an individual's self-directed ability to set and track progress towards goals through planning, implementing, strategizing, monitoring, and evaluating results. In education settings, students who exhibit self-regulation in learning behaviors are able to direct their efforts toward achieving academic goals (Zimmerman and Moylan 2009). They operate at the intersection of motivational beliefs and metacognitive processes, examining goals, reflecting upon current skills, processing feedback, and beginning again. The Cyclical SRL model is comprised of three interconnected phases: the forethought phase, the performance phase and the self-reflection phase (Zimmerman and Moylan 2009). Until recently, few empirical studies have tested the Cyclical SRL model and theory of Zimmerman and Moylan (2009) on academic processes. Zimmerman and Martinez-Pons (1988) first sought to validate the construct using structured interviews and teachers' ratings of 80 students finding evidence to support SRL as a single dimension construct. In Panadero's (2017) review of theory, empirical evidence, and published meta-analysis establishing six SRL models, he refers to empirical evidence collected by Schmitz, Klug and Schmidt (2011) using learning diaries, but these inform the use of Schmidt's model which is variant from Zimmerman's. Similar to Zimmerman and Moylan (2009), Callan and Cleary (2019) used microanalytic interviews to investigate the strength of relationships of the cyclical phase model of SRL, forethought, performance, and self-reflection in the context of mathematics problem solving. Callan and Cleary (2019) found strong correlations between measures from the forethought and performance phases, specifically goal-setting and strategic planning predicted strategy use as measured with either microanalytic interviewing or trace analysis. Callan and Cleary did not find evidence that processes during the forethought phase were predictive of metacognitive processes in the self-reflection phase.

Two recent active areas of research in the field of SRL have addressed SRL in an ITS environment. Rovers et al. (2019) recommend the use of what they call "behavioral measures of SRL", including measures such as the log data collected in MathSpring, as better at predicting achievement than data from student self-reports. Further, Raaijmakers et al. (2019) had mixed results in their investigation of the beneficial effect of feedback on self-assessment accuracy.

MathSpring was carefully and intentionally designed by learning scientists to engage students using best practices for learning mathematics in an adaptive tutoring system (Arroyo et al. 2014). Utilizing the science-based structure, we can represent the student experience in a model which aligns very closely to the Zimmerman and Moylan (2009) Cyclical SRL model. Students enter with their background knowledge, and background states and traits in metacognitive and affective domains, representing the forethought phase. Students bring these attributes to the performance phase where they interact within the system, attempting problems, requesting scaffolding such as hints and videos, receiving affective and motivational prompts, taking time to strive towards successful completion of the problem. Students receive feedback in the form of correct/incorrect answers and the quality of hints and other supportive interventions. Studies have shown significant interactions among hint

seeking, error making (degree of success), and knowledge level/mastery (Aleven et al. 2003; Wood and Wood 1999). Wood and Wood (1999) found that time, number of errors, and number of help requests were all negatively correlated with a pre-test measure representing prior knowledge. Aleven et al. (2003) reviewed empirical work to create a model and theoretical framework on help-seeking and found evidence that learners in ITS do not use help such as hints as frequently or as effectively as was expected. Further, synthesizing results from a variety of studies, Aleven et al. (2003) concluded that learners with low prior knowledge who seek help have greater learning gains.

When the problem is completed, students can then view how this success or mistake has affected their mastery, by viewing a dashboard containing a progress report and based upon mastery calculations embedded in the tutoring system. The progress report includes a growing plant to depict the growth in their knowledge of the topic. Further, mastery and success work through a feedback mechanism, where a student's ability to succeed on a problem changes based upon level of mastery, holding other things constant (Corbett and Bhannagar 1997). Reflection on progress made in the evaluation phase can then influence the states in the forethought phase, changing motivation and self-efficacy. The student then once again enters the performance phase as they make choices and attempt a new problem. Based upon how well they did on the prior problems, mastery is recalculated, and the problem difficulty changes to select problems within the zone of proximal development (Vygotsky 1980). The zone of proximal development refers to learning experiences in an optimal range of difficulty to support growth, stimulate interest, and reduce frustration. Work by Arroyo et al. (2017) contributed to the development of a theoretical framework presented in Fig. 2.

Although our understanding of the role of emotions in student learning in an ITS has advanced (Arroyo et al. [under review](#); Arroyo et al. 2009, 2011). However, the role of a

Fig. 2 A hypothetical self-regulation cycle that combines emotions with the Cyclical SRL model

student's emotional state as they are learning with digital technologies has not been addressed. In particular we seek to understand how emotions affect cognitive and metacognitive functioning as modeled within the Cyclical SRL model. We present here a clear theoretical framework, summarized in Fig. 2 and detailed in the following section.

Emotion

The theoretical framework for this research is based on the control-value (CV) theory of achievement emotion in education (Pekrun et al. 2007; 2009; 2010; Zeidner 2007), which establishes that students experience specific achievement emotions when they feel in/out of control of activities and outcomes that are subjectively important to them. The control-value theory models emotions as epistemic or resulting from cognitive assessments of one's own control to influence an outcome and the degree to which one values a particular outcome (Pekrun et al. 2007). This theory posits that when students value the task/outcome and have a high degree of control then they are likely to experience a positively valenced emotion that may take many forms, such as enjoyment, pride and hopefulness. Emotions vary depending on whether they are in relation to a retrospective-focused event, looking forward to a future-focused event, or during the process of academic work, within-work-focused. Considering negative emotions, if students have failed at a task they may experience a negatively-valenced anxiety (associated with a future-focus lack of confidence in their ability to perform), or frustration (resentment towards their ability to perform the task at hand, within-work-focused), and also shame or anger (when considering their past performance, or in relation to a retrospective-focused event). Eventually, this could turn into anxiety when facing the idea of working in future problem solving (associated with a future-focus lack of confidence in their ability to perform). Baker et al. (2008) found in an ITS environment that gaming the system was associated with frustration. Baker also found a relationship between frustration and belief that time in the ITS was not helpful.

While Pekrun did not mention 'confidence' as an emotion, in many ways, feelings of confidence can be considered the opposite-valenced construct of anxiety. In this article, we refer to 'confidence' as an emotion, for a variety of reasons, despite the fact that there are obviously some metacognitive components to this construct. Culturally, it is common language among students and teachers to talk about their confidence in problem solving, how confident they are, as an emotional experience, a positive future-focused affective experience when facing the idea of working in future problem solving in this case. What we call 'confidence' also quite overlaps with Pekrun's future-focused emotion of 'hope', although talking with students about their confidence seems more common language to students than their hope.

This research focuses predominantly on students' frustration and confidence as they are solving math problems in the MathSpring tutoring system. While many other emotions could have been assessed, these are two emotions (one positive valence, one negative valence) that result from ongoing academic activities, and are typical and familiar to students. Emotions are experienced in all of the self-regulatory cycle, in each of the phases, however, they are assessed in between pairs of learning activities (math problems). We consider it is now time to scientifically understand how emotions relate to each other and how they interact with larger motivational factors such as grit and sustained initiative, to move beyond failure and disenchantment, and address more complicated yet thoroughly motivational, affective, and self-regulatory learning scenarios.

Grit

Researchers, teachers, policymakers, and parents recognize that there are certain traits and dimensions of character other than intelligence that are strong determinants of a person's unique path towards success despite setbacks. Duckworth et al. (2007) identified grit as a trait of anyone who works hard and displays perseverance and passion for long-term goals. Grit is not just a catchy word and a pop-culture phenomenon evidenced by mainstream publications and over 19 million views of her Ted Talk (Duckworth 2020). Supported by good science, educational researchers have turned their attention towards examining grit, how it changes, and how it is related to other learning constructs such as effort, interest, and engagement (e.g. Duckworth and Quinn 2009; Von Culin et al. 2014; Wolters and Hussain 2015), motivating the inclusion of grit in our study of the process of student learning in a cyberlearning environment. Operationalized using an 8 or 12 item scale representing two constructs: consistency of interests and perseverance of effort (Duckworth and Quinn 2009), grit is positively correlated with education and age, Conscientiousness in the Big Five Inventory, GPA in an elite university, and predictive of completion of the training program for West Point Cadets (Duckworth et al. 2007). Further, those students in a national spelling bee with higher grit score were more likely to advance to the final rounds than students with lower grit (Duckworth and Quinn 2009). The long-term nature of grit is what differentiates it from similar constructs such as self-control and conscientiousness (Sumpter 2017).

Grit has been conceived as a very stable, domain-general inclination to maintain consistent, focused interest and effort for personally important and challenging goals. A body of active research, including current research on students in cyberlearning environments, is exploring how grit may change over time (Arroyo et al. [under review](#); Park et al. 2018; Tang et al. 2019). Nonetheless, recent research has questioned the use of grit as a two-factor construct, suggesting that consistency of interest and perseverance of effort may be more accurately analyzed as separate factors (Crede et al. 2017). In their meta-analysis, Credé, Tynan and Harms (2017) found an average correlation between the two factors of $\rho=0.66$ for the short grit scale. Steinmayr, Weidinger and Wigfield (2018) found that perseverance of effort was a stronger predictor of academic achievement than consistency of interest. Crede (2018) also questioned the use of grit as a predictor of academic achievement. This body of research represents valid concerns for grit scale, but none of these reflect the use of the measure with students in intelligent tutoring systems, an area that has become more urgent during the time of Covid-19 and the scaling up of use of online tools.

Grit and its relationship to emotion

Achievement emotions have been studied for decades in relation to learning situations, from anxiety to interest and confusion or frustration, and particularly in the context of digital learning environments. Considering grit in addition to emotion provides a larger context for a student emotion. Solely knowing student emotion is less useful when providing interventions in digital learning environments, without also knowing the larger context of when/how the emotion occurred, and the general initiative of the student to follow through. This research moves beyond the understanding and/or addressing basic emotions in a learning situation (frustration, confidence, and boredom) and attempts to recognize and see the interaction with more stable motivational constructs such as grit, help-seeking, effort and self-efficacy, within a learning situation, in the context of a digital learning environment such as MathSpring.

Expectation of success

One of the areas of great interest for researchers and practitioners in the intersection of cognitive and affective dimensions is in the area of motivation for success. Research has supported theories by Eccles and Wigfield (1999) and others that a student's expectation of success affects their goals, their choices, their perceptions, and ultimately their achievement in cognitive tasks. "Achievement motivation refers to motivation in situations in which (an) individual's competence is at issue" (Wigfield and Eccles 2002, p.1). Expectation of success changes over a student's experiences in school and can increase or decrease based upon student achievement, attributable to peer influences, grading, and other factors related to how an individual perceives their level of success.

Research has shown that emotion is predictive of achievement, (Arroyo et al. 2009, 2011; du Boulay 2018; Pekrun et al. 2002), but it has proven more difficult to show the cause-effect relationship between achievement and emotion. As the measure of mastery increases, this leads to incremental increases in problem difficulty—raising the bar—carefully gauged to support continued cognitive growth in that content area. As problem difficulty increases, students face challenges to success and mastery at that problem difficulty level, leading to increased need for hints and non-cognitive supports to improve their chances of success.

The cyclical SRL simulation model

This current research presents the results of our development of the Cyclical SRL Simulation Model (the Simulation Model), a model of student experiences within an ITS operationalized in terms of cognitive, metacognitive, and emotion constructs simulated within a SD model..

Using the two areas of inquiry, listed again below, we present our research questions.

1. We investigate the interrelationships among learning processes by exploring the validity of the Cyclical SRL model, testing the sensitivity of the system to change, and bolstering good decision-making regarding improvements and interventions within the system.

RQ1: Do log data from MathSpring along with additional self-report measures of grit and expectation of success support an interconnected mathematical model of self-regulated learning with connections to the Zimmerman and Moylan (2009) model?

RQ2: What learnings do we gain from significant regression paths in the hypothesized Cyclical SRL model?

RQ3: Using the coefficients to inform the Simulation Model, how can results from simulations of interventions using the Cyclical SRL model be used to support good decision making on live interventions?

2. We use SD modeling to investigate the relationship between the Cyclical SRL model and emotions as measured by frustration and confidence.

RQ4: In what way do the emotions of frustration and confidence predict measures in the SRL model phases of Forethought, Performance and Self-regulation?

RQ5: Do changes in mastery performance predict consequent changes in emotion within MathSpring?

Utilizing real data from student experiences within an ITS, we develop a statistical model to represent the theoretical framework, use those results to calibrate the Simulation Model, and demonstrate multiple simulation scenarios and their interpretations. Finally, we discuss the implications for further research and practice.

Method

Overview

This research was designed to validate Zimmerman's Cyclical SRL Model and test theories on the role of emotion and grit in SRL. Table 1 displays the connection between the Cyclical SRL model, the conceptual framework displayed in Fig. 2, and the measures collected. Models were designed to test the strength of linear relationships based upon the research described above between variables measured within each phase of the Cyclical SRL model. Grit and expectation of success are measured during the forethought phase. We hypothesize that grit predicts effort, as measured by time spent on the problem, and help seeking, as measured by hints. Mastery is initially set at 0.10 and then is recalculated after each problem. Success is measured during the performance phase, and we hypothesize success as a function of hints, time, problem difficulty and mastery.

Participants

Performance and affective measures were collected using log data on student work within MathSpring for two student groups. The first group consisted of measures from student performance within MathSpring for 305 middle school students (6th–7th grade) in several public schools in Massachusetts across the state, spanning districts with different student socio-economic and math achievement profiles, representing learning experiences from February, 2017 to April, 2018. The second group consisted of 153 middle school students (6th graders) in three schools in Argentina who worked within MathSpring weekly for approximately five weeks in the fall, 2019. Some students experienced the ITS in Spanish, and some experienced it in both English and Spanish. After listwise deletion of student records from ITS sessions that were identified as invalid, e.g. student was in demonstration mode, our final sample consisted of 296 students from Massachusetts and 153 from Argentina.

Table 1 Comparison of SRL Model with the current model constructs and measures

Cyclical SRL Model	Current Model Constructs	Measures
Forethought Phase	Perceived Competence	Expectation of Success Grit
Performance Phase	Help Seeking Time management Self-consequences	Hints Time Success
Evaluation Phase Emotion	Self-evaluation Frustration Confidence	Mastery Emotion composite

Table 2 Description of Variables

Variable	Description	Standardization method	Argentina		Massachusetts	
			<i>M (SD)</i>	<i>N</i>	<i>M (SD)</i>	<i>N</i>
Mastery	A measure of student knowledge of content based upon a combination of cognitive, affective, and metacognitive states of students working in the ITS.	Measures of Mastery are on a scale [0,1].	0.235 (0.145)	153	0.204 (0.107)	296
Problem Difficulty	Different from traditional item difficulty within item response theory, problem difficulty reflects student knowledge taking into account different student behaviors related to engagement.	Measures of Problem Difficulty are on a scale [0,1].	0.568 (0.055)	153	0.688 (0.074)	296
Mistakes	While working on a problem, this represents the number of attempts made by the student that were evaluated as a mistake.	Mistakes were capped at 5 per problem and then standardized within a problem in a two-step process to reflect the distribution of the student and the problem.	0.042 (0.568)	153	-0.211 (0.494)	296
Success on the problem	Success represents the opposite of mistakes.	Success reflects the negative of the z-score for mistakes.	-0.042 (0.568)	153	0.211 (0.494)	296
Hints	While struggling with a problem, students may request hints to help support learning and success. These hints come in the form of spoken words, animation, and videos. Hints represent the total number of hints received by the student.	The number of hints was capped at 4. Z-scores were calculated for hints following the same procedures as for mistakes.	-0.190 (0.685)	153	-0.159 (0.412)	296
Time	Time represents the amount of time in seconds the student spends working on the problem, and it is the maximum of several different time measures.	Z-scores for time were calculated following the same procedures as for mistakes.	0.052 (0.609)	153	-0.043 (0.591)	296
Emotion	Within MathSpring, students are asked about their emotional state, either frustration or confidence, randomly every 5–8 problems. This question asks their level of the emotion on an interval of 1 = not confident/frustrated to 5 = very confident/frustrated.	Emotion is recoded/reverse coded on a scale from 1 to 5 to reflect higher values for positive emotions, lower values for negative emotions. Emotion was then rescaled to [0,1] to support model stability and interpretation.	0.649 (0.197)	153	0.599 (0.238)	296

Table 2 (continued)

Variable	Description	Standardization method	Argentina		Massachusetts	
			<i>M (SD)</i>	N	<i>M (SD)</i>	N
Grit ^a	Using the 8 item Grit-S scale (Duckworth and Quinn 2009), measures were collected for 126 students prior to beginning work in MathSpring. Items were translated into Spanish.	The grit score represents the average of the 8 items, rescaled to [0,1] to support model stability and interpretation.	0.466 (0.208)	126	NA	NA
Expectation of Success ^a	Using the theory of Wigfield and Eccles (2000), we selected 4 items measuring expectation of success from Butler (2016). Items were rewritten to reflect the ITS context and translated to Spanish.		0.704 (0.197)	61	NA	NA

^a Grit and Expectation of Success measures collected in the Argentina sample only

Measures

A uniform set of measures was collected for all students in both data collections while in the MathSpring representing the result of the student's experience during their learning session. Data from log files were cleaned and re-coded following uniform criteria across the two samples. For example, demonstration problems were omitted, data outliers were truncated such as in situations where students asked for more than four hints, and records were aggregated by summarizing measures at the student-topic level using median values. Descriptive statistics along with a brief description of variables and data preparation for each measure are provided in Table 2. Specifically, the table displays the variable name, description, and transformations used to standardize the data for the analysis completed in this research. The initial data extract from the Argentina sample contained log data of 1279 problems in MathSpring reflecting the work of students on multiple problems and topics over the five weeks. Cleaning, recoding, and summarizing to the student-topic level using median measure values resulting in 524 observations representing the work of 153 students. The Massachusetts sample contained log data of 17,673 problems completed by students in MathSpring. Cleaning, recoding, and summarizing to the student-topic level resulted in a dataset containing 1691 records on 296 students.

For the Argentina sample we also used scales to measure grit and expectation of success (Table 3). The eight-item grit scale has favorable psychometric properties and is multidimensional representing two constructs: consistency of interest and perseverance of effort (Duckworth and Quinn 2009). Using four unique samples, Duckworth and Quinn (2009) found strong internal consistency reliability for both the construct Consistency of Interest ($\alpha = 0.73$ to 0.79) and the construct Perseverance of Effort ($\alpha = 0.60$ to 0.78). Consistency of interest and perseverance of effort had a low correlation ($\rho = 0.264$, $p < .05$) in the Argentina sample. In all cases where grit is referenced in this research, we refer to the average of 8 item grit scale, following the methodological recommendation of Duckworth and Quinn (2009). A four-item expectation of success scale was used based upon Wigfield and Eccles (2000) and Butler (2016), within the context of MathSpring and translated to Spanish. An example is "I think I will do really well today in MathSpring". The expectation of success scale used a 5-point Likert Scale, and we found strong internal consistency reliability scale ($\alpha = 0.85$). The scale average for expectation of success was used in this research.

When large and complex models use measures employing different scales, the varying metrics can introduce regression artifacts due to the effect of scaling across measures that are difficult to interpret. In machine learning, standard practice suggests scaling all variables to similar metric size (Jaitley 2019). In this research, we sought efficient scaling of variables to support accurate estimation and ease of interpretation of results, so all metrics within the regressions and SD models were scaled to be within similar ranges. Mastery and problem difficulty measures were maintained on a $[0,1]$ scale matching their scale within the ITS. Measures of the number of hints, time, and success were normalized to create measures on a standard normal distribution scale balancing between per student and per problem. Emotion, grit, and expectation of success were rescaled to the range from $[0,1]$. In all cases, higher scores - and in the models higher coefficients - represent constructs that are positive and more favorable towards learning. Correlations among measures are provided in Table 4. Please see Arroyo et al. (2014) for a greater level of detail on the ITS and the calculation of these variables.

Procedures

SD modeling

We use SD modeling and simulation methodology (Forrester 1968; Richardson 1991), a mathematical approach to modeling complex problems from a holistic perspective focusing on the endogenous factors that explain behavior. SD uses visual representations to explicitly demonstrate cause, effect, and feedback (Sterman 2000). SD was originally developed to model economic theories and logistics in manufacturing and industry, but it has since been shown very effective in advancing knowledge in business, management, higher education and organizational research (Zaini et al. 2016, 2017). The outcomes of these interactions can be tested both logically and empirically, allowing researchers to use the models as tools for theory building and testing through experiments (Zaini et al. 2019). SD models support the solution of many problems and provide mathematical models to represent and test theory and associations that underlie complex processes, in an understandable visualization that supports “thinking, communicating, and learning in an interdependent system” (Richmond 2013, p. 4).

SD models are founded on three building blocks: stocks, flows, and feedback loops (See Fig. 3 below). A stock represents quantities that are tangible (e.g., water) or intangible phenomena (e.g., motivation) that change over time. A flow is a rate of change that influences the level of a stock through filling (inflow or act of motivation), or draining (outflow or act of demotivation) over time in what is known as the bathtub analogy (Richmond 2013). A feedback loop is a circular information path connecting flows to stocks that can create endogenously driven changes in the system. Table 5 shows the basic icons used in SD models.

Table 3 Grit and expectation of success items

Grit Item Number ^a	Item	Reverse Score
1	New ideas and projects sometimes distract me from previous ones.	yes
2	Setbacks don't discourage me.	no
3	I have been obsessed with a certain idea or project for a short time but later lost interest.	yes
4	I am a hard worker.	no
5	I often set a goal but later choose to pursue a different one.	yes
6	I have difficulty maintaining my focus on projects that take more than a few months to complete.	yes
7	I finish whatever I begin.	no
8	I am diligent.	no
Expectation of Success Item Number ^b	Item	
1	I think that I will do really well today in MathSpring.	
2	I am sure that I can understand the math problems that are being taught in MathSpring.	
3	I know that I can solve the problems that appear in MathSpring.	
4	I am sure that I can make the plants grow and master the topics in MathSpring.	

^a A 5-point Likert scale was used from “Very much like me” to “Not like me at all”

^b A 5-point Likert scale was used from “This is definitely true” to “This is not true at all”

Fig. 3 Basic stock and flow representation

The SD model is built and simulated using Stella Architect software 1.9.4¹ environment. Typically, the modeler defines initial values of stocks, constant parameters, algebraic functions, and graphical relationships that characterize each stock, flow, and converter in the model.

Our current application which examines the interplay of cognitive and metacognitive factors in play as a student has an experience in an ITS is similar to the economic systems that inspired the development of SD modeling. Comparing economic models to learning models, students experience many inputs, outputs, stocks and flows when entering and participating in an ITS. They come in with background knowledge, states, and traits in psychological and learning constructs including motivation, grit, and emotional makeup. The student then receives the stimuli of the problems, hints, and nudges within the tutoring system, all of which are aimed at changing the student: improving knowledge, guiding motivation, enhancing grit, and stabilizing emotions, to name a few. SD models provide a set of theory and tools to create a single, customizable, and upgradable mathematical model of learning in an ITS. Knowledge gained has the potential to go beyond informing the fine tuning process for an ITS, to contributing to the knowledge base of how student states and traits interact in the learning process, thus having the potential to impact education policy.

Model formulation and behavior analysis are important steps in the SD modeling process during which we develop a model structure based upon research literature and fine tune it to reflect the structure of the ITS we are simulating. As outlined by Oladokun and Aigbavboa (2018), Forrester criticized economic models in the following areas: they did not reflect feedback loops, did not include many interrelated components in a single model, failed to incorporate the ability to reflect changing attitudes, relied mainly on linear ordinary least squares models to define relationships, and lacked reflection on assumptions. Mindful of these concerns, our approach involved a five step process to model construction, equation formulation, as described in Table 6. We used contextually relevant data collected from the tutoring system incorporated within the implied mathematical model within a SD model structure.

The theoretical system of coupled integral equations in the SD model are solved iteratively using numerical techniques like Runge–Kutta fourth-order or Euler at consecutive time steps over the time period of interest to produce the simulation results (Authors 2019). The solution moves through the simulation duration at a predetermined time step selected to ensure the solution's stability and avoid integration errors. In the current model, instead of time, the steps represent the completion of a problem set by a single student.

Analysis

A non-recursive path model approach was used to analyze the data to test for significant linear relationships between variables, to provide evidence to validate the Zimmerman Cyclical

¹ Stella Architect <https://www.iseesystems.com/>

Table 4 Pearson correlation matrix (n = 524)

—	1	2	3	4	5	6	7
1. Success	—	—	—	—	—	—	—
2. Mastery	0.212**	—	—	—	—	—	—
3. Problem Difficulty	0.039	0.130**	—	—	—	—	—
4. Hints	-0.249**	-0.136**	-0.026	—	—	—	—
5. Time	-0.049	0.017	0.061	0.448**	—	—	—
6. Grit	-0.025	0.059	0.033	0.101*	0.007	—	—
7. Expectation of Success	0.162*	0.103	0.132*	0.040	-0.050	0.457**	—
8. Emotion	0.049	0.203**	0.091*	-0.060	-0.027	0.106*	0.284**

Correlations from Argentina data sample using data summarized by student by topic

* $p < .05$

** $p < .01$

Model of SRL, and to estimate the beta coefficients for intercepts and slopes necessary to inform the SD model equations. Path modeling represents a statistical analysis technique within Structural Equation Modeling (SEM) (Kline 2011). With SEM, model parameters and measurement error can be estimated simultaneously for complex multi-equation models with feedback loops. For the models in this research, a non-recursive technique was used to support the feedback - cyclical model hypothesized by Zimmerman and Moylan. Estimation was completed using Mplus (Muthén and Muthén 1998–2017). These criteria provided the basis of the method used to build and estimate the path model. Model fit in SEM was assessed by giving consideration to significance of coefficients within the theoretical model, the Comparative Fit Index (CFI), and the Root Mean Square Error of Approximation (RMSEA) (Kline 2011). Model fit in SEM is not a measure of absolute fit but relative fit among competing models. CFI values greater than 0.90 and RMSEA values less than 0.06. were set as an indication of acceptable model fit (Hu and Bentler 1999; Netemeyer et al. 2003).

Table 5 Basic icons used in SD visual models

Process	Icon	Explanation
Stock	name	Accumulation or integration of flows linked to the icon
Flow	name	A rate of change or a derivative of a stock. Empty arrowhead indicates normal direction of flow. Normally connected to a stock. Cloud at one end represents unlimited source or sink
Converter	name	Algebraic function of stocks, other converters and constants
Graphical function	name	Graphically represented function of another variable in the system
Causal link		Information relationship between two variables

Data from the Argentina collection and the Massachusetts collection were analyzed using Mplus (Muthén and Muthén 1998-2017) using the maximum likelihood estimator with robust standard errors to correct for lack of normality in some of the measures. Further, results using the maximum likelihood estimator have been found to exhibit less bias than the two stage least squares estimator in models containing direct and indirect feedback loops in structural equation modeling (Price et al. 2019). Results of the Argentina collection provide the most thorough model due to the completeness of the set of variables, but the Massachusetts data collection was also analyzed to validate the results. Starting with testing linear relationships between groups of variables supported by our theoretical model, paths were analyzed and trimmed where not significant, with competing models compared using model fit indices. Models and procedures mirrored each other where possible between the Argentina data collection and the Massachusetts data collection. Feedback loops were included with caution to ensure the models were identified and coefficient changes made sense.

Results

Step 1: Construct the SD model structure

Preparation of a structural model involved combining the theoretical framework outlined earlier into the SD model structure. That is, we provided a system map which combined the cyclical Model of SRL, the interrelationships with emotion and grit, and the functions and response feedback built into the ITS, representing all these in the SD model structure. Within the SD literature, this process is referred to as problem articulation and formulation of dynamic hypothesis (Houghton et al. 2014).

Model description

Figure 4 below shows a conceptual structure of the Cyclical SRL Model with emotions/affect interacting with the three phases in directional and bidirectional relationships. This figure represents the dynamic hypothesis, representing the current knowledge of interrelationships supported by prior research and theory. This section details the process used to build the hypothesized SD model.

The SD model of Cyclical SRL shown in Fig. 5 is designed to simulate the following situation. A student begins working in the ITS by receiving problems at a predescribed average level of difficulty. The student is successful or not on this problem depending upon their level of mastery relative to the problem difficulty and the number and quality of hints and supports sought. The number of hints sought is a function of student emotion and grit. Student emotion changes dependent upon the relative comparison of their mastery with their expectation of success. This is measured by mastery as a predictor of expectation of success, and when students are successful, knowledge mastery goes up, resulting in improved emotion and grit, and ultimately in success and mastery (Success-Mastery- Emotion- Grit combined feedback loops). The level of mastery has a causal connection with student success, and as mastery increases, it is followed by an increase in problem difficulty (raising the bar feedback loop). At some point, the problem difficulty increases to the point where success on the problem and mastery become asymptotic exhibiting equilibrium. Emotion may change quickly, mastery

Fig. 4 The aggregate-level feedback structure combines the three-phase Cyclical SLR model and emotion resulting in richer dynamics relationship between the three phases

more slowly depending on their adjustment times. If emotion drops due to a decline in mastery and expectation of success, an emotional support algorithm can be activated to intervene to improve student emotion with a certain effectiveness (ITS feedback loop).

Because MathSpring was inspired by and theoretically designed after the cyclical phase model of Zimmerman and Moylan (2009), we used real data from student sessions in MathSpring to support equation formulation, calibration, and validation of the model functionality and accuracy.

The Simulation Model was designed to depict the Cyclical SRL model including forethought, performance, and self-evaluation sections. Performance is represented by a measure of problem success, self-evaluation by a measure of mastery, and forethought by measures of grit and expectation of success. Emotion is depicted in the model as outside of and connected with the phases of the Cyclical SRL model. The results of depicting each phase containing stocks and flows in the SD model are described in the following sections using the theoretical Simulation Model (Fig. 5).

Mastery. Feedback between success and mastery align with the feedback loop between the performance and self-evaluation phases advanced in the Cyclical SRL model (Fig. 5). We depict this in the system of equations as follows:

$$\text{Success} = \beta_{10} + \beta_{11} (\text{Mastery}) + \varepsilon_1 \quad (\text{holding other variables constant}) \quad (1)$$

$$\text{Mastery} = \beta_{20} + \beta_{21} (\text{Success}) + \varepsilon_2 \quad (2)$$

Based upon our synthesis of the theories, the SD model depicts Mastery as a stock computed using a moving average according to the following integral equation:

$$\text{Mastery}(t) = \text{Mastery}(t-dt) + (\text{change_in_mastery}) * dt \quad (3)$$

Theory suggests the dynamics of how mastery changes as follows. With each experience in MathSpring, the student receives tutoring and supports to promote growth, contributing to changes to instantaneous mastery, which are modeled as taking place over a hypothesized

$$\text{Change in mastery} = (\text{Instantaneous_mastery} - \text{Mastery}) / \text{Mastery_adjustment_time} \quad (4)$$
 Springer

also increases, making the student less likely to succeed on these harder problems. To reflect these interrelationships, the model was designed to predict success on a problem as a function of four variables: mastery, number of hints, problem difficulty, and time spent on the problem. Therefore, the regression equation mentioned above (1) was revised as follows:

$$\text{Success} = \beta_{10} + \beta_{11} (\text{Mastery}) + \beta_{12} (\text{Hints}) + \beta_{13} (\text{Problem difficulty}) + \beta_{14} (\text{Time}) + \varepsilon_1 \quad (5)$$

Our hypothesized formulation includes, in addition to problem difficulty, a dynamic form of mastery and hint seeking connected with perseverance, measured here as grit. These relationships are depicted in the following equations:

$$\text{Problem difficulty} = \beta_{30} + \beta_{31} (\text{Mastery}) + \varepsilon_3 \quad (6)$$

$$\text{Hints} = \beta_{40} + \beta_{41} (\text{Grit}) + \beta_{42} (\text{Emotion}) + \varepsilon_4 \quad (7)$$

$$\text{Time} = \beta_{50} + \beta_{51} (\text{Grit}) + \beta_{52} (\text{Emotion}) + \varepsilon_5 \quad (8)$$

This feedback between hints, time, grit and emotion align with the feedback loop between the performance and the forethought phases advanced in the Cyclical SRL model (see Fig. 5).

Emotion. Emotions change as a result of a student's experiences in the ITS that increase frustration or increase confidence. Prior theory suggests that emotional levels can vary based upon the experience when a student compares the ITS evaluation of their mastery to their own expectation of success. The SD model reflects the effect of expectation of success on the relationship between mastery and emotion. The connected pathway from mastery to expectation of success to emotion aligns with the feedback loop between the forethought and evaluation phases advanced in the Cyclical SRL model (see Fig. 5).

To simulate support features within the ITS, the SD model contains a mechanism that simulates an intervention to help sustain more positive levels of emotions. Depending on the emotion level, the model simulates a situation where the ITS intervenes to provide emotional support. Depending upon the results of the simulation, a decision can then be made on what the intervention might look like. Similar to Mastery, emotion is operationalized as a stock using the following equation, reflecting the opportunity to improve emotion through the ITS intervention, and predicted as a function of expectation of success.

$$\text{Emotion}(t) = \text{Emotion}(t-dt) + (\text{change_in_student_emotion} + \text{ITS_intervention}) * dt \quad (9)$$

$$\text{ITS_intervention} = (\text{desired_emotion} - \text{Emotion}) / \text{ITS_intervention_delay} \quad (10)$$

$$\text{Emotion} = \beta_{60} + \beta_{61} (\text{Expectation of Success}) + \beta_{62} (\text{Success on the Problem}) + \varepsilon_6 \quad (11)$$

$$\text{change_in_student_emotion} = (\text{Instantaneous_emotion} - \text{Emotion}) / \text{emotion_adjustment_time} \quad (12)$$

Grit. The SD model is designed to test the effect of emotion and expectation of success on grit. In the SD model, grit is also conceptualized as a stock that changes as a function of emotion and expectation of success (See Fig. 5).

$$\text{Grit} = \beta_{70} + \beta_{71} (\text{Expectation of Success}) + \beta_{72} (\text{Emotion}) + \varepsilon_7 \quad (13)$$

$$\text{Expectation of Success} = \beta_{80} + \beta_{81} (\text{Mastery}) + \varepsilon_8 \quad (14)$$

$$\text{Grit}(t) = \text{Grit}(t-dt) + (\text{change_in_grit}) * dt \quad (15)$$

Grit also functions as the driving force, a source of motivation that sustains student work to stay on task and striving forward to complete the goal. In our model, we hypothesize that grit is positively correlated with hints and time, and that grit measured at pre-test predicts help seeking as measured by the request of more hints and taking more time to achieve the goal of success on the problem. These relationships are depicted by the following equations provided earlier as part of the forethought phase.

$$\text{Hints} = \beta_{40} + \beta_{41} (\text{Grit}) + \beta_{42} (\text{Emotion}) + \varepsilon_4 \quad (7)$$

$$\text{Time} = \beta_{50} + \beta_{51} (\text{Grit}) + \beta_{52} (\text{Emotion}) + \varepsilon_5 \quad (8)$$

This theoretical SD model represents the Cyclical SRL model connecting forethought, performance, and self-evaluation with level of emotion using measures collected through student experiences in MathSpring, along with measures of expectation of success and grit. In step 2, we use structural equation modeling to test the theoretical model by evaluating model fit and significance of parameter estimates.

Step 2. Use real data to calibrate the SD model

The resultant variables in the theoretical model for SRL reflect the following system of linear equations which include feedback loops without the stocks that act as buffers.

$$\text{Success} = \beta_{10} + \beta_{11} (\text{Mastery}) + \beta_{12} (\text{Hints}) + \beta_{13} (\text{Problem difficulty}) + \beta_{14} (\text{Time}) + \varepsilon_1$$

$$\text{Mastery} = \beta_{20} + \beta_{21} (\text{Success}) + \varepsilon_2$$

$$\text{Problem difficulty} = \beta_{30} + \beta_{31} (\text{Mastery}) + \varepsilon_3$$

$$\text{Hints} = \beta_{40} + \beta_{41} (\text{Grit}) + \beta_{42} (\text{Emotion}) + \varepsilon_4$$

$$\text{Time} = \beta_{50} + \beta_{51} (\text{Grit}) + \beta_{52} (\text{Emotion}) + \varepsilon_5$$

$$\text{Emotion} = \beta_{60} + \beta_{61} (\text{Expectation of Success}) + \beta_{62} (\text{Success on the Problem}) + \varepsilon_6$$

$$\text{Grit} = \beta_{70} + \beta_{71} (\text{Expectation of Success}) + \beta_{72} (\text{Emotion}) + \varepsilon_7$$

$$\text{Expectation of Success} = \beta_{80} + \beta_{81} (\text{Mastery}) + \varepsilon_8$$

Data from MathSpring collected in Argentina and in Massachusetts were used to estimate the theoretical model. The data collection from Argentina was designed to provide measures for all variables in the theoretical model from the same set of students adding validity to our models. The Massachusetts data collection was used to compare and validate relationships as reflected in MathSpring measures, but this data collection did not include grit and expectation of success measures (see Table 2). Using Mplus 8.3 (Muthén & Muthén, 1998–2017), a series of path models were estimated using full information maximum likelihood estimation with robust standard errors. Tables 7–9 provide the parameter estimates for the model coefficients and residual variances for all three models presented.

The full theoretical model (Table 7) did not provide the desired level of model fit (CFI = 0.539, RMSEA 90% CI = [0.109, 0.148]). Moreover, several coefficients were not statistically significantly different from zero. The model was trimmed to exclude paths that were not significant. Specifically, we saw that although time was a significant predictor of success, the coefficients for the regression predicting time were not significant. Because time is represented as a covariate reflecting the amount of time spent on the problem, it may be acting as a suppressor variable relative to hints in the regression predicting success, as evidenced by the high correlation between time and hints ($\rho = 0.448, p < .01$). The preferred model (Table 8) has superior model fit (CFI = 0.931, RMSEA 90% CI = [0.010, 0.064]), and all coefficients are significant at $p < .10$ level. Table 9 provides the results of testing the model fit for the preferred

Table 6 Development of SD Model of Cyclical SRL

Step	Validation
1. Construct the SD model structure. - to interpret the Cyclical SRL Simulation Model.	Use theory and findings from the development of the SD model and cross validate with literature.
2. Use real data to estimate equations for use in the Cyclical SRL Simulation Model. Using data from the ITS with measures from one set of participants across all constructs, estimate model parameters using path analysis.	Use theory and sound methodological principles to identify useful data sources. All metrics should be standardized to comparable metrics to support interpretation of SD equation results. Assess model fit using CFI and RMSEA. Identify the model with best fit. Select 1 candidate model. Use data from a second set of participants to validate coefficients' relative size.
3. Simulate baseline data. Incorporate parameter estimates into SD model and simulate baseline data. Test SD model sensitivity.	Adjust model as needed to improve interpretability and model fit. Incorporate revised parameter estimates into SD model and repeat.
4. Fine tune SD models. Identify whether models should be simulated with error or without error using residual error variance estimates from path models.	Cross validate by comparing simulated data to real data and reflect upon assumptions.
5. Simulate potential scenarios. Use SD model to simulate possible scenarios within the tutoring system.	Use knowledge gained to inform theory and practice.

model using the Massachusetts MathSpring data. The global fit indices for the preferred model (CFI = 0.942, RMSEA 90% CI = [0.039, 0.081], and the statistical significance of many variable coefficients provided additional evidence in support of the validity of the Argentina preferred model. Specifically, the sign of the coefficients for mastery, hints, and problem difficulty match those from the Argentina models and are within the same size range. Success has a positive coefficient in the equation predicting mastery. The Argentina preferred model was selected to provide the coefficients to inform the Simulation model. Further, coefficients from the Massachusetts model can be used to conduct sensitivity testing of the Simulation model. The validated model in aggregate is provided in Fig. 6, and the detailed final SD Simulation model is provided in Fig. 7. The structural equation model diagram is provided in the appendix (Fig. 12).

Table 7 Path Model Results - Argentina Data - Theoretical Model

Dependent Variable	Parameter	Parameter Estimate	SE
Success on Problem	Intercept	0 ^a	—
	Hints	−0.355*	0.070
	Problem Difficulty	0.614`	0.367
	Mastery	−1.649	0.804
	Time	0.116*	0.057
Mastery	Intercept	0.265*	0.009
	Success on Problem	0.131*	0.038
Problem Difficulty	Intercept	0.556*	0.006
	Mastery	0.043`	0.023
Hints	Intercept	−0.144	0.150
	Grit	0.587*	0.224
	Emotion	−0.225	0.162
Expectation of Success	Intercept	0.688*	0.018
	Mastery	0.189*	0.058
Emotion	Intercept	0.354*	0.047
	Expectation of Success	0.397*	0.059
	Success on Problem	0.002	0.013
Grit	Intercept	0.120*	0.037
	Emotion	−0.083`	0.045
	Expectation of Success	0.548*	0.061
Time	Intercept	0.003	0.162
	Grit	0.094	0.221
	Emotion	−0.118	0.178
Dependent Variable	—	Residual Variance	SE
Success on Problem		0.833*	0.169
Mastery		0.041*	0.005
Problem Difficulty		0.006*	0.001
Hints		0.682*	0.057
Expectation of Success		0.040*	0.004
Emotion		0.046*	0.004
Grit		0.029*	0.003
Time	—	0.665*	0.053

*significant at $p < .05$, `significant at $p < .10$; CFI = 0.539, RMSEA 90% CI = [0.109, 0.148];

^a Model constraint

Table 8 Path model results - Argentina data – preferred model

Dependent Variable	Parameter	Parameter Estimate	SE
Success on Problem	Intercept	0 ^a	—
	Hints	−0.278*	0.053
	Problem Difficulty	0.250 [†]	0.257
	Mastery	−0.838*	0.483
Mastery	Intercept	0.262*	0.009
	Success on Problem	0.093*	0.03
Problem Difficulty	Intercept	0.554*	0.006
	Mastery	0.050*	0.022
Hints	Intercept	−0.257*	0.1
	Grit	0.518*	0.206
Expectation of Success	Intercept	0.689*	0.018
	Mastery	0.187*	0.057
Emotion	Intercept	0.351*	0.048
	Expectation of Success	0.401*	0.059
Grit	Intercept	0.122*	0.037
	Emotion	−0.084 [†]	0.045
	Expectation of Success	0.546*	0.061
Dependent Variable	—	Residual Variance	SE
Success on Problem		0.700 ^a	—
Mastery		0.038*	0.003
Problem Difficulty		0.006*	0.001
Hints		0.687*	0.057
Expectation of Success		0.040*	0.004
Emotion		0.046*	0.004
Grit	—	0.029*	0.003

*significant at $p < .05$, [†]significant at $p < .10$; CFI = 0.931, RMSEA 90% CI = [0.010, 0.064]

^a Model constraint

Step 3. Simulate baseline data

Using the beta coefficients from Step 2, the Cyclical SRL Simulation Model was calibrated and simulations were conducted for sensitivity testing (See Table 9). During sensitivity testing,

Fig. 6 Aggregate Model

results of simulations were compared to real data and known relationships to test model accuracy. For most testing simulations, we set the sample size to 15 representing a single student completing 15 problems within a single topical area, reflective of the Massachusetts and Argentina data samples. During the baseline simulation, the model initializes mastery to reflect the typical initial value used within the tutoring system, mastery = 0.10. All other variables are initialized at their estimated intercept and then updated using the model equations. Criteria for the baseline data included ensuring that the model results for each node reflected expected relationships among the variables. The model as calibrated reflects an asymptotic shape to the problem difficulty graph gradually increasing towards some ceiling representing the highest value that student can complete with success at that time. Problem difficulty, hints and mastery interact to support success on the problem. For example, when students need scaffolding, as indicated by lower success, they are able to request hints, resulting in a negative relationship between hints and success. Higher success supports higher

Fig. 7 Preferred Simulation model

Table 9 Path model results - Massachusetts - preferred model

Dependent Variable	Parameter	Parameter Estimate	SE
Success on Problem	Intercept	-0.693*	0.269
	Hints	-0.189*	0.041
	Problem Difficulty	2.019*	0.374
	Mastery	-2.267*	0.124
Mastery	Intercept	0.194*	0.006
	Success on Problem	0.244*	0.014
Problem Difficulty	Intercept	0.735*	0.006
	Mastery	-0.149*	0.021
Hints	Intercept	-0.047	0.048
	Grit	—	—
	Emotion	-0.130*	0.071
Expectation of Success	Intercept	—	—
	Mastery	—	—
Emotion	Intercept	0.594*	0.008
	Expectation of Success	—	—
	Success on Problem	0.088*	0.013
Grit	Intercept	—	—
	Emotion	—	—
	Expectation of Success	—	—
Dependent Variable	—	Residual Variance	SE
Success on Problem		0.8 ¹	—
Mastery		.036*	0.002
Problem Difficulty		.010*	0
Hints		.359*	0.049
Expectation of Success		—	—
Emotion		.067*	0.002
Grit		—	—

*significant at $p < .05$, †significant at $p < .10$; CFI = 0.942, RMSEA 90% CI = [0.039, 0.081]

¹ Model constraint

mastery, but as mastery increases, success levels off asymptotically indicating the cap in mastery and success for that student at that point in time (Table 9).

We used the preferred model structure in a SEM model to estimate coefficients using the Massachusetts data. A comparison of the relative size and sign of the coefficients provided guidance as to the validity of the model structure. Most noticeable, the coefficient for problem

Fig. 8 Change in Mastery and Problem Difficulty over time in ITS. Average change in mastery and problem difficulty over all students in the Massachusetts sample from first to fifteenth problem

difficulty was higher in the model results using the Massachusetts data ($\beta = 2.019$, $SE = 0.374$) than the model results using the Argentina data ($\beta = 0.250$, $SE = 0.257$). In addition, success on the problem had a lower mean in the Massachusetts data. Further analysis of descriptives indicate that the range of success and problem difficulty in the Massachusetts sample was wider with a higher max, reflecting more variance and more difficult problems, suggesting that the variation in the parameter estimates was within reason.

To further illustrate patterns of mastery and problem difficulty over a student's experiences in the ITS, we examined change in the Massachusetts sample. We provided a proxy for the passage of time by which to measure change using problem sequencing in the order the student completed the problem, i.e., the first problem seen is problem 1, the second problem seen is problem 2, etc. Calculating the means for all students in their first, second, through 15th problems, we then illustrate the change pattern in the data in Fig. 8. As a student completes problems, their mastery increases and then levels off and problem difficulty increases and then fluctuates in a pattern reminiscent of the pattern around the zone of proximal development (Arroyo et al. 2014, p. 399). The shapes of these sample data patterns are similar to the output from the Simulation models, as shown in Figs. 9, 10, and 11 below, providing further validation of the approach and interpretation.

Step 4. Fine-tune the cyclical SRL simulation model

Based on the results of the baseline simulations, the Cyclical SRL Simulation Model was examined and further calibrated to align with the programming within the ITS. For example, mastery has a negative effect on success on the problem, interpretable as the higher the

Fig. 9 Simulation 1: Baseline Model

Fig. 10 Simulation 2. Intervention to improve emotion overlaid on Simulation 1

mastery, the less successful the student is ($\beta = -0.838$, $SE = 0.483$). This may at first seem counterintuitive, but this is an artifact of the feedback loop as follows. As a student works within the system, their work produces learning, and this learning produces successful outcomes and higher mastery. In an infinite system with no boundaries, students would continue to learn, have success on problems of increasing difficulty, and increased mastery. However, neither the ITS, the student experience, or the real data are infinite, such that at some point higher mastery drives higher problem difficulty. Students reach their capacity at that time and begin to make errors on problems, thus explaining how higher mastery can have a causal effect on lower success, due to the changing problem difficulty. The SD model was further calibrated to reflect this same relationship and used to simulate learning scenarios within the ITS as described in the following section. The equations of the system dynamics model and the Mplus syntax are available upon request.

Step 5. Simulate potential scenarios

Simulation 1: Baseline model

The baseline model reflects the experience within MathSpring for an average student. During the baseline simulation, the model initializes mastery to reflect the typical initial value used within the tutoring system, mastery = 0.10. All other stocks are initialized at their estimated intercept and then updated using the model equations. Fig. 9 provides a graph showing how the measures change over the course of attempting 15 problems. As a student attempts a problem, they experience success with easier problems and learning through interaction with supports in the ITS. The mastery measure is incremented upward triggering an increase in

Fig. 11 Simulation 3 Intervention to improve emotion for a student with slower mastery rate of change overlaid on Simulations 1 and 2

problem difficulty. The model reflects an increase in problem difficulty as the model simulates more student learning and increase in mastery. As problem difficulty increases, success decreases, however, the student continues learning until reaching some peak at which time mastery flattens and so does the success decline. During this time, we see emotion and grit also increase, although at a low and steadier rate. In summary, this baseline model illustrates how learning takes place, with a student attempting problems, solving them, selecting correct answers, being challenged by more difficult problems, asking for additional hints, and then ultimately achieving the targeted level of mastery for that topic. Although the baseline model reflects an average learner's experience, we are interested in understanding more about the experience, which is reflected in the next two simulations.

Simulation 2: Intervention to improve emotion

The second simulation reflects the experience when an intervention is used to sustain positive emotion at 80% of the highest score for a learner with average mastery rate. In contrast to simulation 1, an intervention is simulated to provide support for the student's emotional state. The specific type of intervention is not identified at this point, but rather, the model is used to understand what level of effectiveness is needed to effect change in other dimensions of the learning experience. In this model, we operationalize emotion through a measure of frustration and confidence, with higher values representing an emotional state more favorable to learning. Whenever the value of emotion falls below the 80%, the intervention is activated resulting in improvement in the emotion state. In Fig. 10, we see that although most graphs are very similar

to the baseline, the emotion intervention simulated in Run 2 shows a pulse that works, with effectiveness that follows a binomial distribution of 59%, to sustain emotion at a higher level. Because emotion is an indirect predictor of success or mastery, the effect on these values is minimal.

Simulation 3: Intervention to improve emotion for a slower mastery rate of change

The third simulation reflects the experience when an intervention is used to sustain positive emotion at 80% of highest score for a situation where the rate of change of mastery is slower. This could be due to many factors, for example a student who is not engaged within the ITS or a student who is struggling with the content. Fig. 11 provides the graphical displays of these relationships. Similar to Simulation 2, the intervention is activated to sustain the emotion measure at the higher level, but dissimilar to Simulation 2, we simulated a lower rate of mastery increase for this student. In this scenario, we see that students attempt a larger proportion of problems at a lower problem difficulty rate, and that problem difficulty increases at a lower rate. Similarly, mastery levels remain lower longer, increasing at a lower rate. We also see that the lower level of mastery has a reducing effect on expectation of success, which in turn reduces the rate of increase for grit and emotion. One of the benefits of using the simulation model is the ability to see the combined direct and indirect effects of these variables on the outcome variables of interest, something that is difficult to see in a closed system of equations without feedback loops.

A link to an interactive online version of the model can be found as follows:

<https://preview.tinyurl.com/y4jazqz5>

Discussion

This research resulted from interdisciplinary work between motivation in learning sciences, artificial intelligence in education, system dynamics, and structural equation modeling. The outcomes from it are rich and complex, as provided by an analysis of the line of inquiry and research questions as follows. This research investigated the following two areas of inquiry and 5 specific research questions.

1. We sought a deeper understanding of the interrelationships among learning processes by exploring the validity of the Cyclical SRL model, testing the sensitivity of the system to change, and using the SD model to bolster good decision making regarding improvements and interventions within the system.

RQ1: Does log data from MathSpring along with additional self-report measures of grit and expectation of success support an interconnected mathematical model of self-regulated learning with connections to the Zimmerman and Moylan (2009) model? ‘

The Zimmerman and Moylan (2009) Cyclical SRL Model is supported using log data from MathSpring using the sample from Argentina. The model was tested and found to fit the data well, and most regression coefficients were statistically significant. Further, the Massachusetts

sample provided further validation for a subset of the model, namely the interaction between the performance and evaluation phases.

RQ2: What learnings do we gain from significant regression paths in the hypothesized Cyclical SRL model?

The Cyclical SRL model regression coefficients provide strong evidence of relationships aligned with SRL theory between measures in the feedback loop. Here is what we found. In the feedback loop between the evaluation phase and the performance phase, we found that greater success on problems predicts greater mastery, and higher mastery predicts higher problem difficulty. Some of the relationships seemed counterintuitive but made sense upon further investigation. For example, higher mastery was associated with lower success because as mastery increases, problem difficulty also increases, resulting in the student facing more difficult problems, potentially beyond their zone of proximal development, resulting in lower success. Increased use of hints was associated with lower success, again a result of controlling for problem difficulty. When problems are more difficult, students seek more help, but as problem difficulty increases, the hints are not sufficient to support success.

Between the evaluation phase and the forethought phase, we found a series of relationships that provides a new insight on a mechanism for change in grit. We found that higher mastery in the evaluation phase predicted higher expectation of success in the forethought phase, and higher expectation of success predicted an increase in grit. To interpret this, high expectation of success means the student believes they will do well, understand, and be successful on the future task of working in MathSpring. As a student experiences greater success and grows in their knowledge of the topic, these experiences validate their efforts, supporting favorable and hopeful attitudes towards their prospect of success. Over time, the sustained experience of hopeful, valid and encouraging results is the mechanism for changing grit by validating hard work ethic, strengthening student resilience to setbacks, and encouraging their diligence.

RQ3: Using the coefficients to inform the Simulation Model, how can results from simulations of interventions using the Cyclical SRL model be used to support good decision making on live interventions?

The SD Simulation model was used to run three simulations, but it is clear that many more scenarios could be investigated. We ran a baseline model, an investigation to improve emotion using average parameters, and an investigation to improve emotion for a learner whose mastery growth is modeled to reflect a decelerated pace of learning. All three simulations provided interesting and useful results to inform how interventions can be designed to improve the ITS. In practice, interventions are chosen for a variety of reasons related to student age, alignment with the content, and cost, but this model provides additional information on how to gauge whether an intervention is effective enough to make a difference on other outcomes in the model. The intervention function within system dynamics could be added to other sections of the Cyclical SRL model such as hints. This simulation mechanism provides valuable input to balance competing priorities, understand tradeoffs, and choose interventions most effectively.

2. We sought to use the SD model to investigate the relationship between the Cyclical SRL model and emotions as measured by frustration and confidence.

RQ4: In what way do the emotions of frustration and confidence predict measures in the SRL model phases of Forethought, Performance and Self-regulation?

We hypothesized that emotion, as measured by frustration and confidence, would be significant predictors of help seeking, as measured by hints and time, within the performance phase. We did not find significant paths from emotion to the performance phase, but the effect of emotion on hints within the performance phase was mediated by grit. This finding was surprising and is an area for future research.

RQ5: Do changes in mastery performance predict consequent changes in emotion within MathSpring?

Changes in mastery performance do not have a direct effect on emotion, but rather a mediated effect. In this case, mastery has a mediated effect through expectation of success on emotion. This finding is also remarkable and is an area for future research.

The holistic nature of the Simulation Model allows us to ask even more questions. As illustrated in Fig. 10, empirical results confirmed that grit in the forethought phase predicts success in the performance phase through hints. The level of mastery in the evaluation phase predicts success and then success predicts mastery in a feedback loop. We also saw that mastery in the evaluation phase predicts expectation of success in the forethought phase. Mastery affects emotion through its effect on expectation of success, and emotion is a significant predictor of grit. These results represent the first time empirical research has validated the feedback loop of the SRL model using a dataset from a single set of students measured during a common set of experiences within a SRL environment, namely an ITS. Prior research by Callan and Cleary (2019) did not find evidence that processes during the forethought phase were predictive of metacognitive processes in the self-reflection phase, whereas we found evidence that processes from forethought were mediated through performance in how they affected self-reflection. The resultant model streamlines inquiry into student learning through enabling mathematical simulations depicting a variety of student background types and intervention styles and supporting deeper explorations of dimensions of learning.

The structural depiction of the Cyclical Phase SRL model in this research represents a high degree of adherence to the original theory of student learning and modeling graphical design. Initially, we included more stocks, flows, and paths representing hypothesized relationships. We expected time spent on a problem to be a predictive covariate to success, but once we controlled for hints, time was not significant, and hence we trimmed time from the path model. This means that as a student solves problems within the tutoring system, and as they encounter difficulties, they request more hints to support their learning. The trimming of time as a covariate is encouraging because it tells us that students do not just sit back and try to figure the problem out while the clock ticks away; they activate help-seeking behaviors such as requesting hints.

We were initially surprised by the inverse relationship between success on the problem and mastery, and as mentioned above, this led to a deeper dive into examining this phenomenon. The use of path modeling with many linear equations on a rich dataset can result in unexpected coefficient signs, such as the negative coefficient of mastery as a predictor of success on the problem. In a static world where students are attempting an unchanging problem stream, higher mastery would always predict higher success. But an ITS is not that type of model. In

an ITS, as students succeed in problems, they improve mastery, and the ITS increases the level of difficulty, thus providing them with more and more difficult problems (Arroyo et al. 2014). As the student attempts these problems, they find they are less successful at them because the level of difficulty is approaching a level beyond what they can master at that point, that is beyond the zone of proximal development (Vygotsky 1980). This phenomenon is what is reflected in the negative coefficient of mastery in the success equation, supported by results from the Argentina data sample and the Massachusetts data sample. Similarly, we were surprised by the negative coefficient of hints, meaning we hypothesized that hints and success would be directionally similar, but they were not. The effect of changing problem difficulty might also affect the coefficient of hints in that as students complete harder problems, they request greater hints, but once they move outside their zone of proximal development, the hints may no longer be effective in improving success.

The Cyclical SRL Simulation Model has not only provided validation of the Zimmerman and Moylan (2009) cyclical phase model of SLR and validation of the role of emotion in the feedback loop, it also provides a platform for simulating student experiences within an ITS to explore the effect of interventions on student performance, including how different types of students may react. The three simulations presented here, the baseline case, the intervention to improve emotion, and the intervention to improve emotion for a slower learning student, are just a tip of the iceberg of the types of questions about student learning that could be explored. Further simulations could investigate the effect of using different starting values of problem difficulty and mastery, of the effect of interventions, and of the experiences of students with diverse learning patterns. Using the estimated coefficients provides strong validity evidence for the Cyclical SRL model and an understanding of what happens on average, but individual coefficients and change rates could be revised to more accurately simulate patterns for students with disabilities, gifted students, students more easily frustrated and so on.

Limitations and recommendations for future research

The results reflected in the Cyclical SRL Simulation Model provide a unique contribution to what we know about how students experience the ITS environment, but there are limitations and recommendations for future research. The data examined represent two samples, one taken in a middle school in Massachusetts in the United States and one collected from middle school students in Argentina. Although the key relationships were supportive of our final model, these samples were not selected at random and may not be generalizable to all student experiences in an ITS. Data from students in Argentina may be different than United States students due to differences in their experiences based upon language, culture, or level of challenge taken during their time in the ITS, and math topics selected for the learning experience. Future research should include larger samples collected over a longer time period.

This research was limited in development by limitations in available measures. Students were asked self-report questions within the ITS and also given self-report measures outside the ITS, creating a potential for fatigue. We sought a balance between collecting meaningful measures but also in a way that protected our student participants. We were not able to collect data on goal setting, and the Massachusetts sample did not include the self-report measures of grit and expectation of success. Measuring a variety of emotions and other affective states at different time points could support theories regarding the interplay of emotions with performance over time. Further, the measures of emotion represented an aggregation from questions regarding level of frustration or confidence, and thus may have underrepresented the construct

as we conceptualized it. Further, we hypothesized that a student's goal orientation as operationalized by measures of mastery approach, work-avoidance, and performance-avoidance within the forethought phase would be significant predictors of the amount of time a student spent on a problem, the number of hints a student requested, and their emotional state. Current learning theory continues to suggest that these metacognitive states have a strong influence on student behavior in a self-regulated learning environment, but unfortunately, due to limitations in data availability, we were unable to include them in our model.

Future research needs to examine change patterns in student cognitive and metacognitive attributes in the ITS. In this research, we tested time as a covariate of success on the problem, suggesting that longer periods of time working on the problem would predict different success rates. Time plays another role in representing patterns of change - of growth - in these constructs over time. Using longitudinal or mixture modeling techniques would provide techniques for examining how quickly students achieve mastery and how long it takes to change emotion, for example. Growth mixture modeling would provide models depicting typologies of student growth patterns using all the data from the ITS experience (Kooken et al. 2019). For example, some students could grow quickly to mastery, others more slowly, others in an erratic pattern, others may need more hints or metacognitive supports. These patterns could then be used to more precisely simulate the experience of profiles of students with similar growth characteristics and ultimately to provide more precise calibration of the ITS to meet student learning needs.

The use of SEM in estimating model parameters for a non-recursive model usable to calibrate the SD simulation model introduces some areas for future research. In this project, we used established techniques within SEM to identify the best fitting model, but as with any project using SEM with real data, we cannot know if the model has been properly identified. SEM provides estimates of residual (error) variances, which is an indication of error in model fit (Kline 2011). Measures of cognitive, affective, and metacognitive constructs are susceptible to substantial measurement error. Future research might consider how to incorporate residual variances into the SD model to provide a more realistic simulation (Houghton et al. 2014).

Finally, the questions we are asking are directed toward understanding phenomena at the student - problem level, and so multilevel models reflecting student - problem level could provide greater insights to coefficients and relationships. As we expand the level of complexity of the model to reflect longitudinal, multilevel analysis, this will require larger and more complex samples.

Implications for theory and practice

The results of this research have implications to both design of ITS and learning theory. This research produced an SD simulation model of a student's experience within an ITS calibrated using coefficients estimated from real data analyzed with the well-established method of structural equation modeling. These results represent the first-time real data has been used to validate the full Zimmerman and Moylan (2009) Cyclical SRL Model. Empirical evidence supports the theory illustrating how grit and expectation of success in the forethought phase predict success in the performance phase. Success is then predictive of mastery in the self-evaluation phase, which we showed is a significant predictor of expectation of success back in the forethought phase. Further, emotion is connected to mastery through expectation of success and also to grit. These results provide

validation of what many have theorized. Higher grit and expectation of success can lead to greater success, success leads to greater mastery, and then mastery leads back to greater expectation of success and emotion, ultimately leading to a mechanism for change in grit, which we measured as increasing by approximately 10% of one standard deviation. Completion of this study supports the use of the SD Simulation model to simulate experiences in the ITS to improve functionality and intervention quality, complete sensitivity analysis using data from different samples, and can lead to greater understanding and improvement of how students learn both in the ITS and in more traditional environments.

Funding Preparation of this article was supported by funding provided by the National Science Foundation (R1551594; PI Ivon Arroyo).

Compliance with ethical standards

Conflict of interest There are no potential conflicts of interest to disclose.
This research involves human subjects.
All participants provided informed consent.

Appendix

Fig. 12 Structural Equation Model with estimated path coefficients from preferred model

References

- Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. (2003). Help seeking and help design in interactive learning environments. *Review of Educational Research*, 73(3), 277–320. <https://doi.org/10.3102/00346543073003277>.
- Arroyo, I., Burleson, W., Tai, M., Muldner, K., & Woolf, B. P. (2013). Gender differences in the use and benefit of advanced learning technologies for mathematics. *Journal of Educational Psychology*, 105(4), 957–969. <https://doi.org/10.1037/a0032748>.
- Arroyo, I., Cooper, D. G., Burleson, W., Woolf, B. P., Muldner, K., & Christopherson, R. (2009). Emotion sensors go to school. *AIED*, 400, 17–24.
- Arroyo, I., Cooper, D. G., Burleson, W., & Woolf, B. P. (2010). Bayesian networks and linear regression models of students' goals, moods, and emotions. In C. Romero, S. Ventura, M. Pechenizkiy, & R. S. J. Baker (Eds.), *Handbook of Educational Data Mining* (pp. 323–338).
- Arroyo, I., Karumbaiah, S., Woolf, B., Muldner, K., Lizarralde, R., Allesio, D., Wixon, N., & Burleson, W. (under review). Addressing students' emotions in personalized digital learning environments. *Submitted to the International Journal of Artificial Intelligence in Education*.
- Arroyo, I., Micciollo, M., Casano, J., Ottmar, E., Hulse, T., & Rodrigo, M. M. (2017). Wearable learning: multiplayer embodied games for math. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, 205–216.
- Arroyo, I., Woolf, B. P., Burleson, W., Muldner, K., Rai, D., & Tai, M. (2014). A multimedia adaptive tutoring system for mathematics that addresses cognition, metacognition and affect. *International Journal of Artificial Intelligence in Education*, 24(4), 387–426.
- Arroyo, I., Woolf, B.P., Cooper, D.G., Burleson, W., Muldner, K. (2011). The Impact of animated pedagogical agents on girls' and boys' emotions, attitudes, behaviors, and learning. *ICALT 2011, IEEE's International Conference on Advanced Learning Technologies*, Athens, GA.
- Baker, R. S., Corbett, A. T., & Koedinger, K. R. (2004). Detecting student misuse of intelligent tutoring systems. In J. C. Lester, R. M. Vicari, & F. Paragauçu (Eds.), *Intelligent Tutoring Systems* (pp. 531–540). Springer. https://doi.org/10.1007/978-3-540-30139-4_50.
- Baker, R., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., & Koedinger, K. (2008). Why students engage in “gaming the system” behavior in interactive learning environments. *Journal of Interactive Learning Research*, 19(2), 185–224.
- Butler, K. L. (2016). Motivation for mathematics: The development and initial validation of an abbreviated instrument. 198.
- Callan, G. L., & Cleary, T. J. (2019). Examining cyclical phase relations and predictive influences of self-regulated learning processes on mathematics task performance. *Metacognition and Learning*, 14(1), 43–63. <https://doi.org/10.1007/s11409-019-09191-x>.
- Credé, M., Tynan, M. C., & Harms, P. D. (2017). Much ado about grit: a meta-analytic synthesis of the grit literature. *Journal of Personality and Social Psychology*, 113(3), 492.
- Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4(4), 253–278. <https://doi.org/10.1007/BF01099821>.
- Corbett, A. T., & Bhatnagar, A. (1997). Student modeling in the act programming tutor: adjusting a procedural learning model with declarative knowledge. In A. Jameson, C. Paris, & C. Tasso (Eds.), *User modeling* (pp. 243–254). Springer. https://doi.org/10.1007/978-3-7091-2670-7_25.
- Corbett, A. T., Koedinger, K. R., & Anderson, J. R. (1997). Chapter 37-intelligent tutoring systems. In M. G. Helander, T. K. Landauer, & P. V. Prabhu (Eds.), *Handbook of human-computer interaction* (second edition) (pp. 849–874). North-Holland. <https://doi.org/10.1016/B978-0-44481862-1.50103-5>.
- Credé, M. (2018). What shall we do about grit? A critical review of what we know and what we don't know. *Educational Researcher*, 47(9), 606–611.
- du Boulay, B. (2018). Intelligent tutoring systems that adapt to learner motivation. *Tutoring and Intelligent Tutoring Systems*, 103–128.
- Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92(6), 1087–1101. <https://doi.org/10.1037/0022-3514.92.6.1087>.
- Duckworth, Angela Lee. (2020). Grit: The power of passion and perseverance. Retrieved March 19, 2020, from https://www.ted.com/talks/angela_lee_duckworth_grit_the_power_of_passion_and_perseverance.
- Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the short grit scale (Grit-S). *Journal of Personality Assessment*, 91(2), 166–174. <https://doi.org/10.1080/00223890802634290>.
- Eccles, J. S., & Wigfield, A. (1999). In the mind of the actor: The structure of adolescents' achievement task values and expectancy related belief. *Cognitive and Moral Development and Academic Achievement in Adolescence*, 131–196.

- Forrester, J. W. (1968). *Principles of systems*. Cambridge, MA: Wright-Allen Press.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park: Sage Publications Inc.
- Houghton, J., Siegel, M., Goldsmith, D., Mouton, A., Madnick, S., & Wirsch, A. (2014). A survey of methods for data inclusion in system dynamics models: Methods, tools, and applications. In *Proceedings of the 32nd International Conference of the System Dynamics Society*. Netherlands: Delft.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55.
- Jaitley, U. (2019, December 5). Why data normalization is necessary for machine learning models. <https://medium.com/@urvashilluniya/why-data-normalization-is-necessary-for-machine-learning-models-681b65a05029>.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling*. Guilford Publications.
- Kookon, J., McCoach, D. B., & Chafouleas, S. M. (2019). The Impact and Interpretation of Modeling Residual Noninvariance in Growth Mixture Models. *The Journal of Experimental Education* 87(2), 214–237.
- Midgley, C., Kaplan, A., Middleton, M., Maehr, M. L., Urdan, T., Anderman, L. H., Anderman, E., & Roesser, R. (2013). Patterns of adaptive learning scales [Data set]. *American Psychological Association*. <https://doi.org/10.1037/t19870-000>.
- Muenks, K., Wigfield, A., Yang, J. S., & O'Neal, C. R. (2017). How true is grit? Assessing its relations to high school and college students' personality characteristics, self-regulation, engagement, and achievement. *Journal of Educational Psychology*, 109(5), 599.
- Murray, T., & Arroyo, I. (2002). Toward measuring and maintaining the zone of proximal development in adaptive instructional systems. In S. A. Cerri, G. Gouardères, & F. Paraguaçu (Eds.), *Intelligent tutoring systems*, (pp. 749–758). Springer. https://doi.org/10.1007/3-540-47987-2_75.
- Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user's guide* (Eighth ed.). Los Angeles, CA: Muthén & Muthén.
- National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards for mathematics*. Washington, DC: Authors.
- Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). Scaling procedures: Issues and applications. In *Thousand Oaks*: Sage.
- Oladokun, M. G., & Aigbavboa, C. O. (2018). Simulation-based analysis of energy and carbon emissions in the housing sector: A system dynamics approach. In *Springer*.
- Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for Research. *Frontiers in Psychology*, 8, 1–28.
- Park, D., Yu, A., Baelen, R. N., Tsukayama, E., & Duckworth, A. L. (2018). Fostering grit: Perceived school goal-structure predicts growth in grit and grades. *Contemporary Educational Psychology*, 55, 120–128. <https://doi.org/10.1016/j.cedpsych.2018.09.007>.
- Pekrun, R., Elliot, A. J., & Maier, M. A. (2009). Achievement goals and achievement emotions: Testing a model of their joint relations with academic performance. *Journal of educational Psychology*, 101(1), 115–135.
- Pekrun, R., Frenzel, A. C., Goetz, T., & Perry, R. P. (2007). The control-value theory of achievement emotions: An integrative approach to emotions in education. In *Emotion in education* (pp. 13–36). San Diego: Academic Press.
- Pekrun, R., Goetz, T., Daniels, L. M., Stupnisky, R. H., & Perry, R. P. (2010). Boredom in achievement settings: Exploring control–value antecedents and performance outcomes of a neglected emotion. *Journal of Educational Psychology*, 102(3), 531–549.
- Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, 37(2), 91–105.
- Price, L. R., Gonzalez, D. P., & Whittaker, T. A. (2019). Performance of nonrecursive latent variable models under misspecification. *Structural Equation Modeling: A Multidisciplinary Journal*, 26(1), 12–23. <https://doi.org/10.1080/10705511.2018.1463164>.
- Raaijmakers, S. F., Baars, M., Paas, F., van Merriënboer, J. J. G., & van Gog, T. (2019). Effects of self-assessment feedback on self-assessment and task-selection accuracy. *Metacognition and Learning*, 14(1), 21–42. <https://doi.org/10.1007/s11409-019-09189-5>.
- Rahmandad, H., Oliva, R., & Osgood, N. D. (2015). *Analytical methods for dynamic modelers*. Cambridge: MIT Press.
- Richardson, G. P. (1991). *Feedback thought in social science and systems theory*. Philadelphia: University of Pennsylvania.
- Richmond, B. (2013). An introduction to systems thinking: STELLA software (Nachdr.). In *ISEE Systems*.
- Rovers, S. F. E., Clarebout, G., Savelberg, H. H. C. M., de Bruin, A. B. H., & van Merriënboer, J. J. G. (2019). Granularity matters: Comparing different ways of measuring self-regulated learning. *Metacognition and Learning*, 14(1), 1–19. <https://doi.org/10.1007/s11409-019-09188-6>.

- Saeed, K. (2008). Trend forecasting for stability in supply chains. *Journal of Business Research*, 61(11), 1113–1124. <https://doi.org/10.1016/j.jbusres.2007.11.005>.
- Sanz, M. T., Arnau, D., González-Calero, J. A., & Arevalillo-Herráez, M. (2017). Using system dynamics to model student performance in an intelligent tutoring system. *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, 385–386. <https://doi.org/10.1145/3079628.3079635>.
- Schmitz, B., Klug, J., & Schmidt, M. (2011). Assessing self-regulated learning using diary measures with University students. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 251–266). New York, NY: Routledge.
- Scott, B. A., Awasty, N., Johnson, R. E., Matta, F. K., & Hollenbeck, J. R. (2019). Origins and destinations, distances and directions: Accounting for the journey in the emotion regulation process. *Academy of Management Review*. <https://doi.org/10.5465/amr.2017.0448>.
- Sottolare, R. A., Graesser, A., Hu, X., & Holden, H. (2013). Design recommendations for intelligent tutoring systems. *US Army Research Laboratory*.
- Steinmayr, R., Weidinger, A. F., & Wigfield, A. (2018). Does students' grit predict their school achievement above and beyond their personality, motivation, and engagement? *Contemporary Educational Psychology*, 53, 106–122.
- Sterman, J. (2000). *Business dynamics: Systems thinking and modeling for a complex world*. Boston, MA: Irwin/McGraw-Hill.
- Sumpter, A. L. (2017). *Grit and Self-control: Independent contributors to achievement goal orientation and implicit theories of intelligence* [Ohio Dominican University]. https://etd.ohiolink.edu/pg_10?0::NO:10:P10_ACCESSION_NUM:oduhonors149425540515589.
- Tang, X., Wang, M.-T., Guo, J., & Salmela-Aro, K. (2019). Building Grit: The Longitudinal Pathways between Mindset, Commitment, Grit, and Academic Outcomes. *Journal of Youth and Adolescence*, 48(5), 850–863. <https://doi.org/10.1007/s10964-019-00998-0>.
- Von Culin, K. R. V., Tsukayama, E., & Duckworth, A. L. (2014). Unpacking grit: Motivational correlates of perseverance and passion for long-term goals. *The Journal of Positive Psychology*, 9(4), 306–312. <https://doi.org/10.1080/17439760.2014.898320>.
- Vygotsky, L. S. (1980). Mind in society: The development of higher psychological processes. In *Harvard University Press*.
- Wigfield, A., & Eccles, J. S. (2000). Expectancy–Value Theory of Achievement Motivation. *Contemporary Educational Psychology*, 25(1), 68–81. <https://doi.org/10.1006/ceps.1999.1015>.
- Wigfield, A., & Eccles, J. S. (2002). *Development of achievement motivation*. San Diego: Academic Press.
- Wolters, C. A., & Hussain, M. (2015). Investigating grit and its relations with college students' self-regulated learning and academic achievement. *Metacognition and Learning*, 10(3), 293–311. <https://doi.org/10.1007/s11409-014-9128-9>.
- Wood, H., & Wood, D. (1999). Help seeking, learning and contingent tutoring. *Computers & Education*, 33(2), 153–169. [https://doi.org/10.1016/S0360-1315\(99\)00030-5](https://doi.org/10.1016/S0360-1315(99)00030-5).
- Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. (2009). Affect-aware tutors: recognising and responding to student affect. *International Journal of Learning Technology*, 4(3–4), 129–164.
- Woolf, B. P., Arroyo, I., Muldner, K., Burleson, W., Cooper, D. G., Dolan, R., & Christopherson, R. M. (2010). The effect of motivational learning companions on low achieving students and students with disabilities. In V. Aleven, J. Kay, & J. Mostow (Eds.), *Intelligent tutoring systems* (pp. 327–337). Springer. https://doi.org/10.1007/978-3-642-13388-6_37.
- Zaini, R. M., Elmes, M. B., Pavlov, O. V., & Saeed, K. (2016). Organizational dissent dynamics: a conceptual framework. *Management Communication Quarterly*, 31(2), 258–277.
- Zaini, R. M., Pavlov, O. V., Saeed, K., Radzicki, M. J., Hoffman, A. H., & Tichenor, K. R. (2017). Let's talk change in a university: a simple model for addressing a complex agenda. *Systems Research and Behavioral Science*, 34(3), 250–266. <https://doi.org/10.1002/sres.2408>.
- Zaini, R. M., Elmes, M. B., Pavlov, O. V., & Saeed, K. (2019). Organizational dissent dynamics in universities: simulations with a system dynamics model. *Management Communication Quarterly*. <https://doi.org/10.1177/0893318919846625>.
- Zeidner, M. (2007). Chapter 10 - test anxiety in educational contexts: concepts, findings, and future directions. In I. P. A. Schutz & R. Pekrun (Eds.), *Emotion in education* (pp. 165–184). Academic Press. <https://doi.org/10.1016/B978-012372545-5/50011-3>.
- Zimmerman, B. J. (2013). From cognitive modeling to self-regulation: A social cognitive career path. *Educational Psychologist*, 48(3), 135–147.
- Zimmerman, B. J., & Martinez-Pons, M. (1988). Construct validation of a strategy model of student self-regulated learning. *Journal of Educational Psychology*, 80(3), 284–290. <https://doi.org/10.1037/0022-0663.80.3.284>.

Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. *Handbook of Metacognition in Education*, 299–315.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.