Optional ERIC Coversheet — Only for Use with U.S. Department of Education Grantee Submissions

This coversheet should be completed by grantees and added to the PDF of your submission if the information required in this form **is not included on the PDF to be submitted**.

GRANTEE SUBMISSION REQUIRED FIELDS

Title of article, paper, or other content

Beyond Performance Analytics: Using Learning Analytics to Understand Learning Processes That Lead to Impro

All author name(s) and affiliations on PDF. If more than 6 names, ERIC will complete the list from the submitted PDF.

Last Name, First Name	Academic/Organizational Affiliation	ORCID ID
Vanacore, Kirk P.	Worcester Polytechnic Institute	
Lee, Ji-Eun	Worcester Polytechnic Institute	
Egorova, Alena	Worcester Polytechnic Institute	
Ottmar, Erin	Worcester Polytechnic Institute	

Publication/Completion Date—(if *In Press,* enter year accepted or completed) 2023

Check type of content being submitted and complete one of the following in the box below:

- ☐ If article: Name of journal, volume, and issue number if available
- ☐ If paper: Name of conference, date of conference, and place of conference
- ☑ If book chapter: Title of book, page range, publisher name and location
- ☐ If book: Publisher name and location
- ☐ If dissertation: Name of institution, type of degree, and department granting degree

Perspectives on Learning Analytics for Maximizing Student Outcomes, 168-187, IGI Global, PA, USA.

DOI or URL to published work (if available) 10.4018/978-1-6684-9527-8.ch009

Acknowledgement of Funding— Grantees should check with their grant officer for the preferred wording to acknowledge funding. If the grant officer does not have a preference, grantees can use this suggested wording (adjust wording if multiple grants are to be acknowledged). Fill in Department of Education funding office, grant number, and name of grant recipient institution or organization.

"This work was supported by U.S. Department of Education **[Office name]** Institute of Education Scie through **[Grant number]** R305A1804 to **Institution]** Worcester Poly. The opinions expressed are those of the authors and do not represent views of the **[Office name]** Institute of Education Scie or the U.S. Department of Education.

Chapter 9

Beyond Performance Analytics:
Using Learning Analytics to Understand Learning Processes That Lead to Improved Learning Outcomes

Kirk P. Vanacore
https://orcid.org/0000-0003-0673-5721
Worcester Polytechnic Institute, USA

Ji-Eun Lee
Worcester Polytechnic Institute, USA

Alena Egorova
Worcester Polytechnic Institute, USA

Erin Ottmar
Worcester Polytechnic Institute, USA

ABSTRACT

To meet the goal of understanding students' complex learning processes and maximizing their learning outcomes, the field of learning analytics delves into the myriad of data captured as students use computer assisted learning platforms. Although many platforms associated with learning analytics focus on students' performance, performance on learning related tasks is a limited measure of learning itself. In this chapter, the authors review research that leverages data collected in programs to understand specific learning processes and contribute to a robust vision of knowledge acquisition. In particular, they review work related to two important aspects of the learning process—students' problem-solving strategies and behavioral engagement—then provide an example of an effective math program that focuses on the learning process over correct or incorrect responses. Finally, they discuss ways in which the findings f rom this research can be incorporated into the development and improvement of computer assisted learning platforms, with the goal of maximizing students' learning outcomes.

DOI: 10.4018/978-1-6684-9527-8.ch009

1. INTRODUCTION

Learning is a complex process that requires exposure to content, thought, struggle, and memory (Bjork & Bjork, 2020; Koedinger et al., 2023; Lynch et al., 2018; Okano et al., 2000). To have learned something, a student must not only be able to perform a task in the moment, but demonstrate that they can retain the information or skill and apply it to new situations (Soderstrom & Bjork, 2015). Learning systems in which students encounter desirable difficulties (Bjork & Bjork, 2020) – such as varying presentation of content (Smith et al., 1978; Smith & Handy, 2014), interweaving knowledge components instead of presenting them sequentially (Rohrer et al., 2014), spacing content delivery (Cepeda et al., 2006) and retrieval practice (Karpicke & Zaromb, 2010) – increase the likelihood of learning. Yet, when desirable difficulties are designed into learning activities, students' performance often suffers, even as these design choices can positively affect learning as measured by distal outcomes (Roediger & Karpicke, 2006; Shea & Morgan, 1979; Smith & Rothkopf, 1984).

Despite the potential incongruence between performance and learning, Computer Assisted Learning Platforms (CALP) often rely heavily on students' performance within learning activities as the primary measure to evaluate students. The instructional design in these CALPs may vary greatly; some incorporate game-based learning (Siew et al., 2016), puzzles (Rutherford et al., 2010), and simulations (Martens et al., 2004; McCoy, 1996), while others focus on more traditional problem sets with tutorial instruction (Heffernan & Heffernan, 2014). While specific design and difficulty within these instructional methods may also differ, many of these CALPs use mastery learning (Barnes et al., 2016; Heffernan & Heffernan, 2014; Macaruso & Hook, 2007; Ritter et al., 2016). Mastery learning is based on the premise that students must master a knowledge component prior to progressing to new content (Bloom, 1968). Mastery of a skill is often determined using students' performance within the activities, either by modeling student knowledge or using simple criteria, such as solving three problems correctly in a row (Corbett & Anderson, 1994; Kelly et al., 2015; Reich, 2020; Yudelson et al., 2013). Early education technologies relied on modified versions of Rasch models which estimated the probability a student will get a problem correct based on a function of a problem's difficulty and a student's ability (Reich, 2020). Alternatively, knowledge tracing – which seeks to predict the probability that a student has mastered a knowledge component – has emerged as one of the main methods for assessing students' mastery of learning within a problem set (Corbett & Anderson, 1994; Yudelson et al., 2013). Furthermore, mastery learning systems often include interactive elements that provide assistance, which may further boost performance by reducing the mental effort necessary to produce correct responses, while potentially hindering learning (Koedinger & Aleven, 2007; Koedinger et al., 2008). Overall, these systems rely on students' performance data, most often represented by their binary outcomes on problems within an activity, to evaluate student learning and determine whether students should progress on to the next knowledge component.

Furthermore, a portion of Learning Analytics (LA) research also relies on proximal performance within a program. For example, A/B tests, which are commonly used in LA research both to evaluate features and to study learning mechanisms, often use proximal performance as an outcome measure. A/B tests are experiments that test the effect of conditions on a desired outcome through random assignment to a treatment (e.g., access to feature) and control (e.g., no feature access). They have commonly used the impact of evaluate features - such as hints, feedback, and reward systems - on student learning. Some of these tests rely solely on the students' performance directly after the problem to estimate the effects of features and interventions (Haim et al., 2022; Patikorn & Heffernan, 2020; Prihar et al., 2021; 2022), while others use a variety of performance and behavioral outcomes that were generated during or directly after the experimental problem set (Gurung, Baral, et al., 2023; Vanacore et al., 2023).

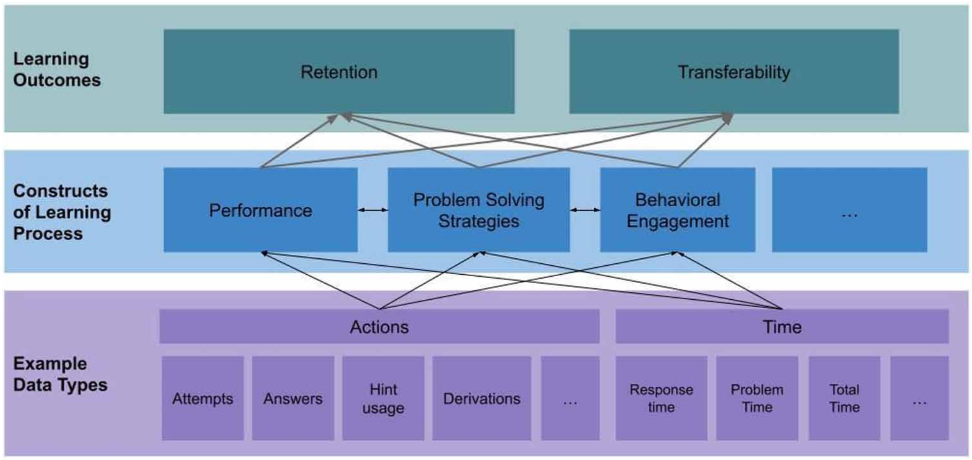
Many mastery learning programs are effective at improving students' learning outcomes, and analyses of performance have advanced our understanding of parts of the learning process (Anderson et al., 1995; Hurwitz & Vanacore, 2022; Kulik et al., 1990; Roschelle et al., 2016). Yet, there is evidence that alternative approaches to mastery learning, such as emphasizing students' learning processes as they solve problems, may have a greater impact on learning (Decker-Woodrow et al., 2023). For LA to fulfill the goal of understanding student learning and optimizing instructional systems (Long et al., 2011), the field has probed deeper into students' behaviors in CALP *as they solve problems* while incorporating this understanding into the development of learning systems. This process leverages different problem types which can range from traditional multiple-choice or fill-in-the-answer problems to more complex puzzles and games that provide different data about how students engage with the content. This undertaking requires analyzing and utilizing more than the proximal performance measures within the data, and focusing on the *learning processes* by which students arrive at the answers they submit.

Our chapter presents various LA research that goes beyond performance measures to understand students' learning processes as they solve problems in various CALP. The chapter is divided into two sections (Sections 2 and 3). In Section 2, we provide examples of research on both mastery learning programs and programs that use alternatives to proximal correctness when evaluating student learning, and explore how the field is moving toward a more nuanced understanding of learning. In Section 3, we provide an example of a program, which focuses on students' problem-solving processes, effort, and engagement, and discuss its impact on student learning.

To frame Section 2, we present a conceptual model (presented in Figure 1) in which data generated by CALP are used to understand how constructs that underlie students' learning processes lead to robust learning defined by retention and transferability. This model is not comprehensive, but it is a useful framework for understanding LA research. Studies in LA use a wide variety of log data collected in CALP. For the purpose of this chapter, we have grouped data into two general categories: *action data* and *time data. Action data* varies depending upon the program, but may include log-in/out data, attempts to solve problems, submitted answers, use of hints embedded within the programs, and steps students take to solve the problem (i.e., derivations). *Time data* can be aggregated at different levels to include students' response time for each action, the time taken to complete a problem, or their total time using the program. Section 2 focuses on how various data from CALP is being used in LA research to study two key constructs of learning: *problem-solving strategies* and *behavioral engagement*.

In Section 3, we use an online learning game, *From Here To There!* (FH2T) as an example of how systems that emphasize the learning processes can effectively promote students' mathematical learning. We outline the theoretical basis for FH2T and discuss evidence of its efficacy. Finally, we discuss new causal research on understanding key mechanisms and behaviors that drive the efficacy of FH2T as well as ways in which we can encourage effective learning behaviors within learning systems. As a whole, FH2T, and its associated research base, illustrate how CALP can improve our understanding of students' learning processes while positively impacting those processes by looking beyond performance.

Figure 1. Conceptual model of the components of learning analytics research which include the data captured in CALP systems, the components of the learning process, and learning outcomes
Lines represent the interconnected nature of the components across the model's levels



2. BEYOND PERFORMANCE: STUDYING LEARNERS' PROBLEM-SOLVING STRATEGIES AND ENGAGEMENT

2.1 Problem-Solving Strategies

In order to solve problems, students need to engage in cognitive processing to understand a problem, derive a solution, and conduct a sequence of steps to arrive at their answers. In many CALP, this process is *unobserved,* as programs simply ask the students to select or input their final answers into the system. In these cases, possible problem-solving paths may be inferred by speculating plausible processes for incorrect answers. While incorrect answers can help provide cues about student errors and inform targeted feedback and guidance, the variability in correct approaches or multiple strategies that students use while problem-solving is largely invisible in the data.

In traditional mastery learning CALP, researchers have attempted to derive students' processes from their given solution. For example, in a series of studies, Gurung et al. (2023a; 2023b) had teachers examine common wrong answers in a CALP, speculate how students would produce those answers, and then create automated feedback that directs toward the estimated paths. The feedback had mixed success in affecting students' performance, which may indicate that the teacher's speculations were at times incorrect. Notably, this system only works for incorrect answers and does not accommodate multiple paths to any solution. Furthermore, the students' true path to their solution remains unobserved.

Alternatively, when online learning programs focus on students' problem-solving processes or strategies rather than performance on tasks (i.e., producing correct answers), researchers can utilize *observable* students' actions in the form of log data collected in CALP to unpack traditionally *unobserved* cognitive processes (Rowe et al., 2021; Shute et al., 2016; Sun et al., 2022). For example, Shute et al. (2016) applied data mining methods to use student in-game action data as an indicator of problem-solving skills in an online mathematics learning game called "Use Your Brainz". The goal of this game was not just to produce a correct answer, but students needed to apply various problem-solving skills to achieve the goal (i.e., planting special plants on a lawn to prevent zombies from invading their houses). In this study, the authors divided students' use of problem-solving strategies into four sub-constructs: analyzing givens and constraints, planning a solution pathway, using tools and resources effectively, and monitoring and evaluation process. Then, they identified 32 students' in-game actions logged in the system (e.g., planting over three flowers before the zombies arrive), mapped them into each sub-construct of problem-solving skills (e.g., analyzing givens and constraints), and implemented this problem-solving model in the game using Bayesian networking. The results revealed that the in-game indicators of problem-solving skills significantly correlated with the results of two external tests of problem-solving skills, suggesting that the problem-solving skills assessment using in-game metrics is valid.

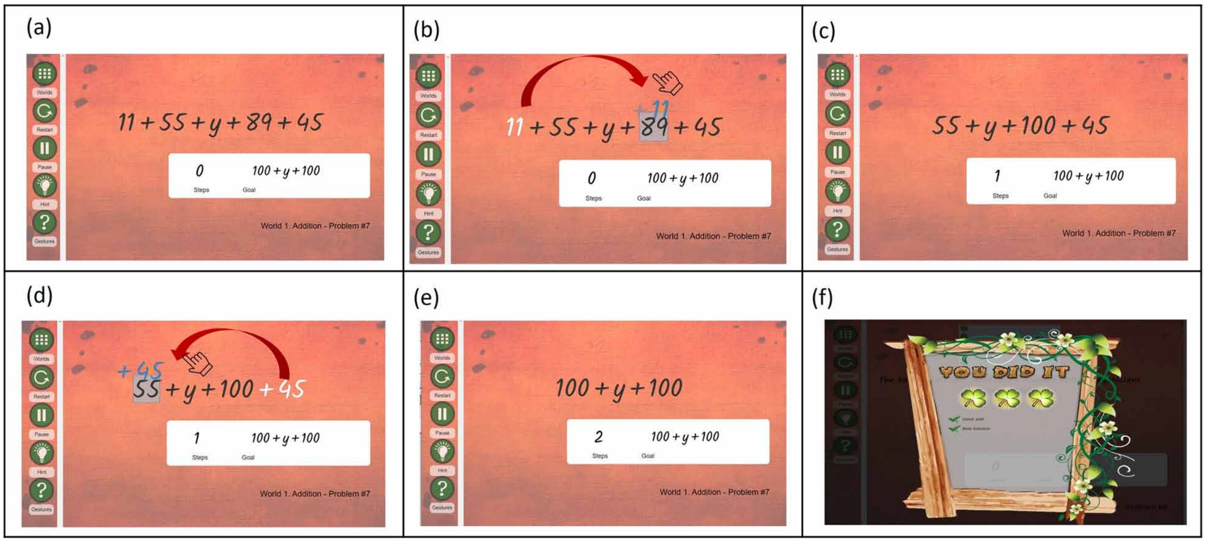
In another study, Rowe et al. (2021) used classification algorithms to create automated detectors of students' implicit problem-solving processes based on gameplay behaviors in an online mathematics game called "Zoombinis". Similar to Shute's (2016) study, they divided students' problem-solving strategies into four different phases (i.e., trial and error, systematic testing, systematic testing with a partial solution, implementing with a full solution) and used both raw log data (e.g., overall gameplay data) collected in the game system and hand-labeled data from observations of students gameplay (e.g., problem decomposition, pattern recognition) to build over 100 features (i.e., detectors) that might be indicative of these strategies. Then, they examined the relationships between these detectors and external post-assessment scores for validation and found that most of the detectors were significantly associated with the post-assessment scores.

Further, when in-game action data is combined with other types of data obtained from external sources (e.g., interaction among students), researchers can measure more complex learning processes, such as how a group of students uses their knowledge and skills to solve complex problems collaboratively. For example, Sun et al. (2022) investigated the relationship between collaborative problem-solving behaviors and solution outcomes in a physics learning game, "Physics Playground". They conducted qualitative coding of students' utterances during their gameplay and identified 19 indicators of collaborative problem-solving skills (e.g., constructing shared knowledge, negotiation). Then, they examined the relationship between these indicators and in-game performance (e.g., problem-solving efficiency) measured by action data. The results indicated that "proposing solution ideas contributed to desirable outcomes" was the most influential predictor of in-game performance out of 19 indicators of collaborative problem-solving, suggesting that collaborative problem-solving requires individual contributions as well as collective interactions.

The game-based algebraic CALP *From Here To There!* (FH2T; https://graspablemath.com/projects/fh2t), which was developed by Ottmar et al. (2015), utilizes a slightly different approach by directly logging *sequences* of students' problem-solving steps. In each problem of the game, students are asked to transform an algebraic expression or equations (e.g., "11+55+y+89+45" in Figure 2) into a mathematically equivalent but perceptually different goal state (e.g., "100+y+100" in Figure 2) using permissible touch-screen or mouse-based gesture-actions. The sequence of steps a student makes to reach the goal state (e.g., Figure 2a through f) is captured along with its timestamp in the system, allowing researchers and learning analysts to study students' various pathways to reach the solution.

In contrast to traditional mastery learning programs, FH2T does not evaluate students by producing answers to math problems but by how efficiently they reach the specified goal state in the game. For example, if a student reaches the goal state with the fewest steps possible to complete the problem (i.e., also called the "optimal step") using a two steps strategy (Student A in Figure 3: [start state: 11+55+y+89+45] → [step 1: 11+100+y+89] → [step 2: 100+y+100]), three clovers (i.e., rewards

Figure 2. A sample problem in FH2T in which shows the steps students take to manipulate the equation from the start to the goal state



in the game) are given, and an efficiency score (optimal step count/total step count) of 1 is assigned. However, the number of clovers they receive is deducted if the students exceed the minimum required number of steps to reach the goal state (e.g., Student B and Student C in Figure 3). As such, there is no correct/incorrect dichotomy in the game, but there are several different pathways to solve the problems, which allows researchers to explore variations of students' mathematical approaches and problem-solving processes to reach the solution.

Using the data collected in FH2T, we applied data visualization techniques and explored students' algebraic problem-solving processes in the game (Lee et al., 2022). Specifically, the study visualized individual students' step-by-step information about the problem-solving process (e.g., math strategies used, time spent between each action) using Indivisualizer (See Figure 3) and created Sankey diagrams (See Figure 4) to investigate how the productivity of the first step influenced the overall efficiency of problem-solving. The results showed a large variation in students' use of mathematical strategies to solve the problems, with some approaches being more efficient than others, and the productivity of the first step significantly predicted the overall efficiency of problem-solving. The findings suggested that these data visualizations depicting students' problem-solving processes can help unpack individual students' cognitive processing as well as variability in overall students' mathematical problem-solving strategies.

In this case, it is also possible to utilize *time data* to estimate processes, such as the amount of thought a student might be applying to the problem before they produce a solution. Using action data can help us grasp the invisible cognitive processes; however, there is always an amount of thinking ascribed to the action. In order to solve problems properly or efficiently, a student needs to examine a problem, pause, and strategically think about what to do next rather than rushing through the problem. Some studies have used "think aloud" methods to explore students' thinking and pause time (Charters, 2003), but these methods can be time-consuming to collect the data individually for each student. Time data automatically collected in CALP also can be used as a potential indicator of students' cognitive processing. As an example, we used time data to measure students' thinking and pausing before problem-solving in

FH2T (Chan et al., 2022). By computing the time students spent before making a first action on each problem, we found that students' pre-solving pause time was positively associated with strategy efficiency, which suggests that pause time may be a proxy indicator of students' strategic planning (i.e., thinking about pathways and solutions to the specific problem). Note that more details about FH2T are discussed in the later section.

Figure 3. An example of the Indivisualizer

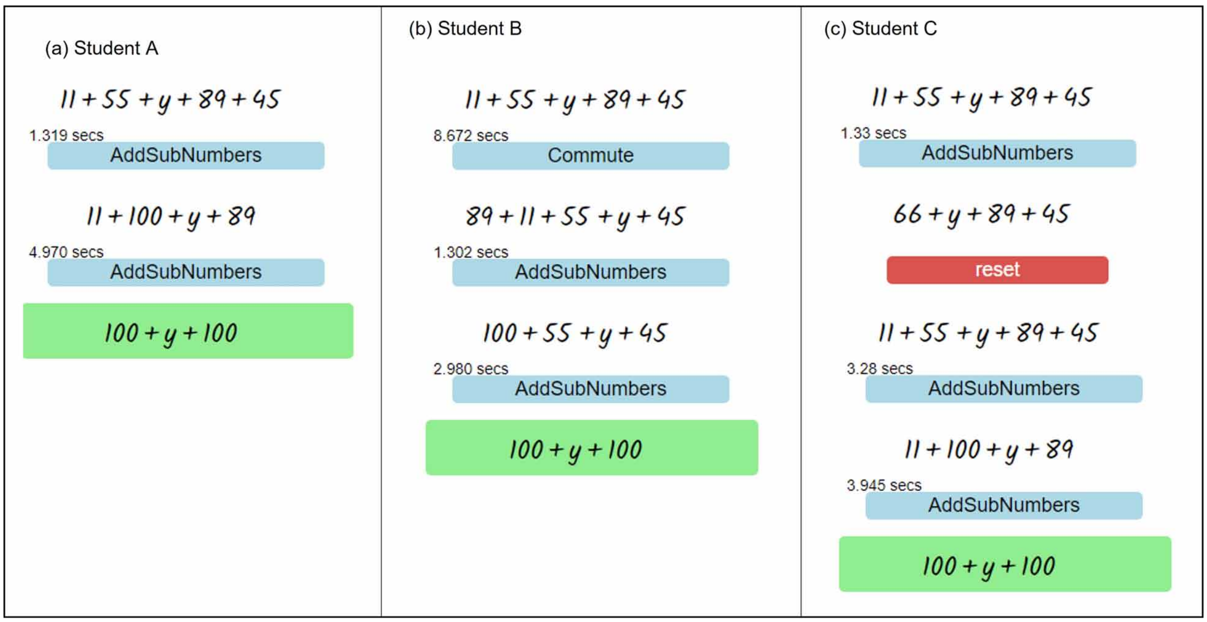
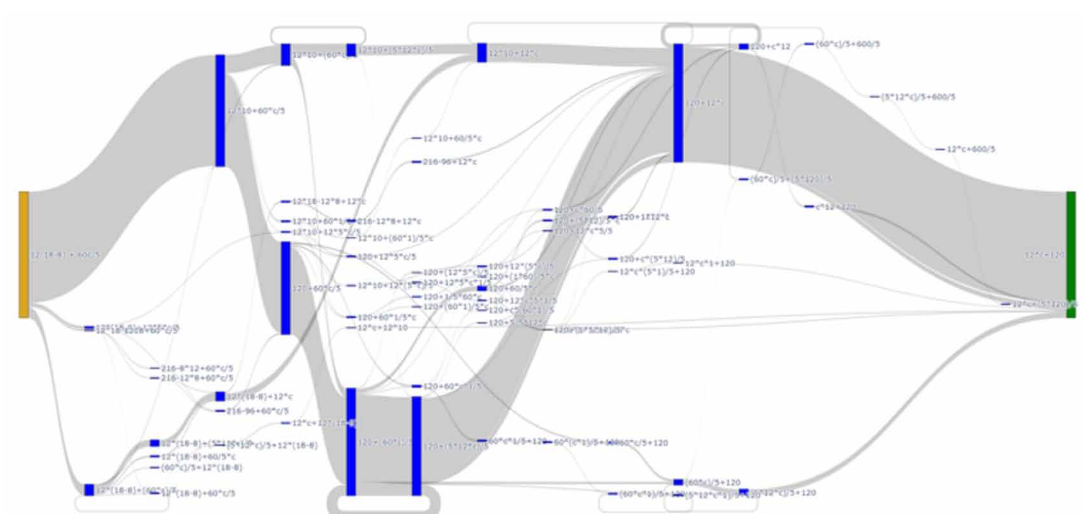


Figure 4. An example of a Sankey diagram showing all student pathways for a given problem



2.2 Students' Behavioral Engagement

Another important aspect of the learning process is students' *behavioral engagement*, which is generally defined as the student's involvement in one's own learning or academic tasks (Sinatra et al., 2015). In classrooms, researchers often measure students' behavioral engagement through class attendance and participation in activities, attentiveness, effort, persistence toward the tasks, or self-directed information-seeking actions when facing obstacles (Sinatra et al., 2015). Studies have found that students' behavioral engagement is significantly associated with positive proximal and distal learning outcomes, such as knowledge gain, better performance (Ladd & Dinella, 2009; Rohrer et al., 2014; Rutherford et al., 2014), and a lower probability of dropout (Archambault et al., 2009).

As with students' problem-solving strategies, behavioral engagement is often observed indirectly in CALP. In mastery learning CALP, LA researchers use process data to assess behavioral engagement, often by attempting to predict whether students will display suboptimal behavioral patterns. For example, researchers have exerted extensive effort to discern whether students are "gaming the system," which is attempting to succeed by exploiting properties of the system (e.g. requesting all the hints available) instead of actively engaging with the material being taught (Baker et al., 2008). For example, Baker et al. (2006) built a detector that identifies students' gaming behaviors using data collected through field observations and students' in-game action data. Alternatively, Paquette et al. (2018) identified 13 action patterns of gaming behaviors. Using these patterns, they identified gaming the system behaviors in several learning contexts and platforms with a satisfactory level of reliability. Notably, much of this work focuses on predicting rather than understanding learning behaviors, and when researchers have tried to understand why students game the system, they have looked into self reports as opposed to utilizing with-in program process data (Baker et al., 2008).

There are other examples of using process data to understand behavioral engagement (or disengagement) within mastery learning CALP. For example "Wheel-Spinning" – spending a substantial amount of time struggling to learn a subject without achieving mastery (Beck & Gong, 2013). As with gaming the system, most of this work has focused on predicting rather than understanding the behavior (Botelho et al., 2019A; Gong & Beck, 2015; Mu et al., 2020; Zhang et al., 2019). Another example is Botelho et al. (2019B) exploration of a behavior they called "stopout," for when students refuse to complete a problem. They used action data prior to the stopout behavior to understand its antecedents.

Although much of behavioral engagement research on mastery learning CALP focuses on negative behaviors of unproductive persistence or disengagement, some researchers have used students' response time data to estimate attentiveness to problems and hints. One example is Gurung et al. (2021) exploration of students' effort in ASSISTments - a math-focused CALP that includes mastery learning and traditional problem sets with immediate feedback. They estimated whether students exerted effort after requesting a hint by analyzing students' response times immediately after seeking help in combination with their subsequent action – which they refer to as "response time decomposition." Shih et al. (2008) also examined response times to estimate the time students spent thinking about bottom-out hints (i.e., worked examples that include problems' answers). Both of these analyses found that more time spent between accessing hints and the student's next action (requesting another hint or submitting an answer) was associated with better performance, thus they deduced that time to be an indicator of thoughtful effort.

Contrary to much of the research on mastery learning CALP, LA research on behavioral engagement with CALP that focus on learning processes is often aimed to understand desirable behavioral engagement. Similar to the works of Gurung and Shih, we examined the relationship between pause time and problem solving efficiency in FH2T (Chan et al., 2022). Using students' response times in conjunction with their derivations, we found that students who spent a higher proportion of pausing time before making a first action used more efficient problem solving strategies.

Another aspect of behavioral engagement explored in LA research is productive persistence in problem-solving. For example, Ventura and Shute (2013) used students' time on problems to assess their persistence within the physics game, Newton's Playground, which they validated using an external measure of persistence administered outside of the program. Another way of measuring persistence is to examine how frequently students replay problems. Replayability, which is a feature embedded within many CALP, provides students an opportunity to engage in several attempts to solve problems, often after they have achieved suboptimal performance on prior attempts (Boyce et al., 2011; Liu et al., 2017; Shang et al., 2006).

In FH2T, for example, multiple attempts at each problem are appropriate because there are many different pathways to each solution, and some paths are more optimal than others. This has allowed us to evaluate students' persistence. As mentioned earlier, students receive clovers upon completing a problem, which shows how efficient their solution was, and students are encouraged to replay that problem if their path is suboptimal. Our earlier work found that replaying a problem is associated with a higher likelihood of having optimal performance on the next problem (Liu et al., 2022) and also significantly predicts learning gains (Lee, et al., 2022).

Further, using FH2T data, we investigated whether specific types of feedback were associated with the number of optional replays students made to solve the problem (Liu et al., 2022; Vanacore et al., 2023). In FH2T, students can attempt to make incorrect arithmetic operations or can provide inefficient solutions to the problem. In the first case, the system provides students with error-feedback by shaking the number that the student was trying to put into the incorrect place. In the second case, reward-based feedback – the number of clovers given to students after problem completion – indicates the efficiency of their solution. Our analyses suggest that reward-based feedback can motivate students to attempt problems more than once.

Other studies have used replay behaviors to understand behavioral engagement with the program. For example, Liu et al. (2017) examined the relationship between students' replay patterns and their learning gains in a puzzle-like online mathematics game, ST Math. The results provide a nuanced picture of the association between replay behavior and learning. Students who replayed problems immediately after completing the level demonstrated the highest in-game performance, whereas replayed problems within their current level had lower learning gains. Similarly, Clark et al. (2011) used data on students replaying levels to understand the nuances of the effects of the physics game on student learning and affect. Notably, they did not find significant correlations between replay behaviors and any of their learning or affective outcomes.

In sum, the studies reviewed in Section 2 provide examples of how LA researchers can use data collected in CALP to understand not only whether the student gets an answer correct but also how the student arrives at a particular solution. When CALP is created to emphasize the process of problem-solving, as opposed to the performance of answering problems, data are produced in a way that helps unpack students' complex cognitive processes when solving problems. Analyses focused on how students arrive at a solution can provide valuable information and insights about implicit students' learning processes.

3. IMPACT OF FOCUSING ON LEARNING PROCESSES RATHER THAN MASTERY: WHAT WORKS AND WHY

3.1 Example of Impact: FH2T Efficacy Studies

Throughout the previous section, we have touched upon how LA can help us unpack complexities in students' learning processes using data from both mastery learning and alternative CALP. For this section, we use FH2T as an example of how a CALP focused on students' learning processes has been shown to outperform more traditional methods of instruction focused on mastery learning and performance in improving algebraic knowledge.

As mentioned above, FH2T is a game-based educational technology that aims to improve algebraic understanding (Ottmar et al., 2015). FH2T was developed based on theories of perceptual learning and embodied cognition. Perceptual learning theory posits that algebraic reasoning is inherently perceptual, which involves visual processing: seeing expressions or equations as structured objects, identifying symbols, and organizing them into groups (e.g., parsing $2 \times a + b \times 3$ as $(2 \times a) + (3 \times 4)$) (Goldstone et al., 2010). Embodied cognition theories suggest that students' physical experiences influence their thinking and reasoning in mathematics (Abrahamson et al., 2020). Integrating these theories into the game, FH2T made implicit mathematical metaphors and symbols into visual and tactile virtual objects. In this way, students can easily identify implicit algebraic structures through manipulation and transformations of the objects (e.g., touching and moving the symbols) in the game. While many math instructions tend to focus on memorizing abstract and seemingly arbitrary rules (e.g., multiplication and division before addition and subtraction), FH2T provides perceptual-motor experiences that help students acquire not only algebraic knowledge but also appropriate perceptual processing skills essential for fluency in algebraic problem-solving.

One of the integral parts of the game is that students can use any mathematically valid action, resulting in various strategies or pathways to solve each problem. Although there are the most efficient ways that students can solve the problem using the fewest steps possible, students can reach the goal state in an infinite number of mathematically valid ways. In this way, they have the freedom to think flexibly and creatively and realize that math problems can be solved in a large number of ways. Through these interactions, students experience dynamic algebraic transformations, rather than a series of static equations. In sum, FH2T emphasizes the *process of arriving at a mathematical solution* over the correctness of answers or completion of the problems, by providing students with a rich experience with puzzle-like problems that go beyond submitting or selecting correct or incorrect answers.

A number of Randomized Control Trials (RCT) consistently showed that students in the FH2T condition outperformed on algebra assessments compared to students who completed online problem sets, including multiple choice and fill-in-the-blank problem sets with automated hints and feedback (Chan et al., 2022; Decker-Woodrow et al., 2023; Hulse et al., 2019). Recently, Decker-Woodrow et al. (2023) conducted a large-scale efficacy RCT to examine whether FH2T improves 7th-grade students' ($N = 1{,}850$) algebraic understanding more than two other CALPs: a game-based learning program DragonBox (Kahoot!) and more traditional algebra problems sets presented with hints and immediate feedback (administered through ASSISTments). These conditions were all compared to the active control of traditional algebra problem sets with delayed feedback (also administered through ASSISTments). The results showed that students in the two game-based learning conditions (i.e., FH2T, DragonBox) showed larger learning gains in algebraic understanding compared to the control condition after 4.5 hours of intervention sessions, even after controlling for prior knowledge and demographic variables.

These results indicate that game-based programs which focus on the process of learning have benefits beyond providing more insight into students' learning processes; they also benefit students' learning. Notably, the availability of immediate hints and feedback did not produce significant differences in algebraic knowledge compared with the active control. This may be because hints can alleviate difficulties in the learning process, thus impacting students' performance (Patikorn & Heffernan, 2020; Prihar et al., 2021), while not impacting learning, due to a phenomenon known as the *Assistance Dilemma* (Koedinger & Aleven, 2007). Alternatively, programs that focus on the learning process through game-based tasks may help students engage in productive struggle, by allowing them to explore potential solutions to each problem and seek multiple paths to a solution. Thus, the emphasis on the learning process may create the condition which optimizes learning.

3.2 Providing a Deeper Understanding of Impact

Due to the rich problem-solving data produced as students use FH2T, the data from large efficacy studies provide opportunities to look beyond average treatment effects and understand the causal mechanisms driving these effects. Leveraging quasi-experimental methods, it is possible to use what we have learned about students' processes in correlational research to understand the causal relations between program features and desirable student behaviors as well as between those behaviors and their outcomes. This burgeoning area of work will lead to a better understanding of why some programs outperform others, why some students benefit more than others, and how we might improve programs to maximize learning for all students.

One example of how process data can be used to study the mechanisms of learning is through conducting research on replay behaviors. As explained above, students in FH2T are encouraged to replay problems when they have suboptimal performance, and this behavior is associated with a higher likelihood of performance within the game (Chen et al., 2020; Liu et al., 2022). To go beyond evaluating the association between behaviors and performance, we applied a quasi-experimental method – fully latent principal stratification (Sales & Pane, 2019) – to FH2T efficacy study data in order to estimate the effects of FH2T for students with a high propensity to replay problems (Vanacore et al., 2023). We found that the effect of FH2T was about twice as large for students with a high propensity to replay. This suggests that the ability to replay problems in FH2T is a key mechanism in the effectiveness of the program. Notably, the process orientation of FH2T allows for the replay feature. Unlike traditional mastery learning programs in which students can submit one correct answer, FH2T allows students to take multiple correct paths to the answer. Those who take this opportunity benefit more from the program.

The obvious next step is to understand how to effectively encourage students to replay more problems. To this end, we evaluated the impact of the performance-feedback based systems in FH2T on students' likelihood of replaying each problem after they have suboptimal performance using another quasi-experimental method, regression discontinuity design (Vanacore et al., 2023). We found that when students received the lowest level performance-feedback (a score of one out of three) based on their performance, they were more likely to replay a problem compared to receiving a higher performance-feedback (a score of two out of three). We speculate that the students view receiving the lower performance-feedback as a game-based -failure, which motivates them to retry the problems. This suggests that adjusting performance-based feedback systems to communicate game-based failures while also providing opportunities to reattempt the problems can encourage students towards productive persistence.

In sum, we have identified a feature of the program (e.g, ability to replay problems) which, when coupled with a behavior (e.g., replaying problems), increases the impact of the CALP on students' learning. Then, we have studied how to influence this behavior within the program. In the future, we can adjust the program in order to encourage replay behavior and test whether this improves the program's efficacy. Thus, this serves as an example of how understanding the learning process can lead to an interactive cycle of improving CALP to help students learn.

4. CONCLUSION AND FUTURE DIRECTIONS

In order to understand the complexity of students' learning processes, LA has turned to rich data sources produced by CALP to study students' learning processes, including their problem solving strategies and behavioral engagement. Furthermore, to maximize student outcomes, CALP have provided robust learning environments that go beyond submitting or selecting answers to statically presented problems, towards creating dynamic and interactive learning environments. This chapter provided examples of how LA can be used to reveal learning processes and learning-related behaviors. The knowledge gained through this research is then leveraged to create programs that optimize systems for student learning.

Yet to be truly meaningful, this research area must grow in two ways. In their article describing the importance of LA, Wise et al. (2021) suggest that impactful LA should focus on (1) "closing the loop" by connecting the learners to actionable interventions through data collection and analyses and (2) creating an iterative cycle which integrates research findings into improving learning systems. They also emphasize that understanding the process of learning using data from programs that emphasize that process is only the beginning for LA. The research on the learning process needs to be connected to interventions for learners and used to improve the programs themselves.

"Closing the loop" involves going beyond learners using a CALP, data collection within the CALP, and analysis/metricization of the data from the CALP. This loop is incomplete without using the analyses and metrics to drive interventions (Clow, 2012). The research presented above does not utilize the analysis of the learning process to differentiate instruction within the programs or help teachers do so in their classrooms. Using our understanding of students' learning process to drive dynamic systems that tailor instruction to students' needs is the necessary and logical next step of this work. Furthermore, helping teachers access understandable data and metrics about their students' strategies and behaviors will help them address their students' misconceptions and provide informed feedback as they teach.

The learning analytics cycle also requires a feedback loop of program improvement (Wise et al., 2021). This is still missing from much of the research on students' learning processes. The field has identified some strategies and behaviors associated with better learning outcomes, and we have started to pinpoint causal factors that connect these strategies with greater impact on learning. Yet, more work should be done to understand how to guide students toward better learning strategies and higher engagement. In Section 3.3, we present one example of a behavior – replaying problems after suboptimal performance – which is both associated with positive outcomes and influenced by a feature of the program. More work must be done to close the loop between identifying positive learning strategies and behaviors and influencing students such that they develop better strategies and behaviors.

Notably, we focused on problem-solving strategies and behavioral engagement as key constructs of the learning process, as we point out in the introduction, they are not the only relevant constructs. For example, students' affect and mindsets are influential factors in how they engage with learning tasks and can be added to the conceptual model proposed in Figure 1. Both of these have been studied at length in LA (e.g., Andres et al., 2019; Baker et al., n.d.; Dillon et al., 2016; Stone et al., 2019; Vanacore et al., 2023; Wang et al., 2015). For example, Baker et al. (2012) detected students' affect using derivations of students' action and time data. As with problem-solving strategies and behavioral engagement LA, more work must be done to discern students' affect and mindsets, as well as understand how we might positively influence these aspects of learning.

In conclusion, LA is a field of iterative design and evaluation with the aim of improving learning environments. Researchers and educators know that learning is complex and nuanced. To effectively meet the aim of LA, we must continue to focus on studying learning as a nuanced and complex process. This focus will help researchers and educators understand students as they learn, and ensure that systems are optimally designed to assist the process of learning.

REFERENCES

Adjei, S. A., Baker, R. S., & Bahel, V. (2021). Seven-year longitudinal implications of wheel spinning and productive persistence. *22nd International Conference, AIED 2021*, (pp. 16–28), Springer. 10.1007/978-3-030-78292-4_2

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences*, *4*(2), 167–207. doi:10.120715327809jls0402_2

Andres, J., Ma, A. L., Ocumpaugh, J., Baker, R. S., Slater, S., Paquette, L., Jiang, Y., Karumbaiah, S., Bosch, N., Munshi, A., Moore, A., & Biswas, G. (2019). Affect sequences and learning in Betty's brain. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, (pp. 383–390). Springer. 10.1145/3303772.3303807

Archambault, I., Janosz, M., Fallu, J.-S., & Pagani, L. S. (2009). Student engagement and its relationship with early high school dropout. *Journal of Adolescence*, *32*(3), 651–670. doi:10.1016/j.adolescence.2008.06.007 PMID:18708246

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.

Baker, R. S., Nasiar, N., Ocumpaugh, J. L., Hutt, S., Andres, J. M. A. L., Slater, S., Schofield, M., Moore, A., Paquette, L., Munshi, A., & Biswas, G. (2021.). Affect-targeted interviews for understanding student frustration. In *Artificial Intelligence in Education: 22nd International Conference, AIED 2021 Proceedings, Part I* (pp. 52-63). Cham: Springer International Publishing.

Baker, R. S. J. D., Corbett, A. T., Koedinger, K. R., & Roll, I. (2006). Generalizing detection of gaming the system across a tutoring curriculum. In M. Ikeda, K. D. Ashley, & T.-W. Chan (Eds.), *Intelligent Tutoring Systems* (pp. 402–411). Springer Berlin Heidelberg. doi:10.1007/11774303_40

Baker, R. S. J. D., Gowda, S. M., Wixon, M., Kalka, J., Wagner, A. Z., Salvi, A., Aleven, V., Kusbit, G. W., Ocumpaugh, J., & Rossi, L. (2012). Towards sensor-free affect detection in cognitive tutor algebra. In International Educational Data Mining Society. https://eric.ed.gov/?id=ED537205

Barnes, T., Chi, M., & Feng, M. (2016). MATHia X: The next generation cognitive tutor. *Proceedings of the 9th International Conference on Educational Data Mining*.

Beck, J. E., & Gong, Y. (2013). Wheel-Spinning: Students who fail to master a skill. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), *Artificial Intelligence in Education* (pp. 431–440). Springer. doi:10.1007/978-3-642-39112-5_44

Bjork, R. A., & Bjork, E. L. (2020). Desirable difficulties in theory and practice. *Journal of Applied Research in Memory and Cognition*, *9*(4), 475–479. doi:10.1016/j.jarmac.2020.09.003

Bloom, B. S. (1968). Learning for mastery. Instruction and curriculum. Regional Education Laboratory for the Carolinas and Virginia, Topical Papers and Reprints. *Evaluation Comment, 1*(2). https://eric.ed.gov/?id=ED053419

Botelho, A. F., Varatharaj, A., Patikorn, T., Doherty, D., Adjei, S. A., & Beck, J. E. (2019). Developing early detectors of student attrition and wheel spinning using deep learning. *IEEE Transactions on Learning Technologies*, *12*(2), 158–170. doi:10.1109/TLT.2019.2912162

Botelho, A. F., & Varatharaj, V. Van Inwegen, Eric. G., & Heffernan, N. T. (2019). Refusing to try: Characterizing early stopout on student assignments. In *The 9th International Learning Analytics & Knowledge Conference (LAK19)*. Springer. 10.1145/3303772.3303806

Boyce, A., Doran, K., Campbell, A., Pickford, S., Culler, D., & Barnes, T. (2011). BeadLoom Game: Adding competitive, user generated, and social features to increase motivation. *Proceedings of the 6th International Conference on Foundations of Digital Games*, (pp. 139–146). Springer. 10.1145/2159365.2159384

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3), 354–380. doi:10.1037/0033-2909.132.3.354 PMID:16719566

Chan, J. Y.-C., Lee, J.-E., Mason, C. A., Sawrey, K., & Ottmar, E. (2022). From Here to There! A dynamic algebraic notation system improves understanding of equivalence in middle-school students. Journal of Educational Psychology, 114(1), 56-71. https://doi.org/10.1037/edu0000596

Chan, J. Y.-C., Ottmar, E. R., & Lee, J.-E. (2022). Slow down to speed up: Longer pause time before solving problems relates to higher strategy efficiency. Learning and Individual Differences, 93, 102109. https://doi.org/10.1016/j.lindif.2021.102109

Charters, E. (2003). The use of think-aloud methods in qualitative research: An introduction to think-aloud methods. *Brock Education Journal*, *12*(2), 2. doi:10.26522/brocked.v12i2.38

Chen, F., Cui, Y., & Chu, M.-W. (2020). Utilizing game analytics to inform and validate digital game-based assessment with evidence-centered game design: A Case study. *International Journal of Artificial Intelligence in Education*, *30*(3), 481–503. doi:10.100740593-020-00202-6

Clark, D. B., Nelson, B. C., Chang, H.-Y., Martinez-Garza, M., Slack, K., & D'Angelo, C. M. (2011). Exploring Newtonian mechanics in a conceptually-integrated digital game: Comparison of learning and affective outcomes for students in Taiwan and the United States. *Computers & Education*, *57*(3), 2178–2195. doi:10.1016/j.compedu.2011.05.007

Clow, D. (2012). The learning analytics cycle: Closing the loop effectively. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, (pp. 134–138). Springer. 10.1145/2330601.2330636

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. doi:10.1007/BF01099821

Dillon, J., Ambrose, G. A., Wanigasekara, N., Chetlur, M., Dey, P., Sengupta, B., & D'Mello, S. K. (2016). Student affect during learning with a MOOC. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, (pp. 528–529). IEEE. 10.1145/2883851.2883960

Fang, Y., Nye, B., Pavlik, P., Xu, Y. J., Graesser, A., & Hu, X. (2017). *Online learning persistence and academic achievement*. 10th International Conference on Educational Data Mining. EDM 2017, Wuhan, China.

Goldstone, R. L., Landy, D. H., & Son, J. Y. (2010). The education of perception. *Topics in Cognitive Science*, *2*(2), 265–284. doi:10.1111/j.1756-8765.2009.01055.x PMID:25163789

Gong, Y., & Beck, J. E. (2015). Towards detecting wheel-spinning: Future failure in mastery learning. *Proceedings of the Second (2015) ACM Conference on Learning @ Scale*, (pp. 67–74). Springer. 10.1145/2724660.2724673

Gurung, A., Baral, S., Vanacore, K. P., Mcreynolds, A. A., Kreisberg, H., Botelho, A. F., Shaw, S. T., & Hefferna, N. T. (2023). Identification, exploration, and remediation: Can teachers predict common wrong answers? *LAK23: 13th International Learning Analytics and Knowledge Conference* (pp. 399–410). https://doi.org/10.1145/3576050.3576109

Gurung, A., Lee, M., Baral, S., Sales, A., Vanacore, K., McReynolds, A., Kreisberg, H., Heffernan, C., Haim, A., & Heffernan, N. (2023). How common are common wrong answers? Crowdsourcing remediation at scale. Learning@Scale 2023. https://doi.org/10.1145/3573051.3593390

Haim, A., Prihar, E., & Heffernan, N. T. (2022). *Toward improving effectiveness of Crowdsourced, on-demand assistance from educators in online learning platforms.* Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners' and Doctoral Consortium: 23rd International Conference, AIED 2022, Durham, UK. 10.1007/978-3-031-11647-6_5

Heffernan, N. T., & Heffernan, C. L. (2014). The ASSISTments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *International Journal of Artificial Intelligence in Education*, *24*(4), 470–497. doi:10.100740593-014-0024-x

Hulse, T., Daigle, M., Manzo, D., Braith, L, Harrison, A, & Ottmar, E. (2019). From here to there! Elementary: A game-based approach to developing number sense and early algebraic understanding. Educational Technology Research and Development, 67, 423-441.

Hurwitz, L. B, & Vanacore, K. P. (2022). Educational technology in support of elementary students With reading or language-based disabilities: A cluster randomized control trial. Journal of Learning Disabilities, 00222194221141093. https://doi.org/10.1177/00222194221141093

Karpicke, J. D., & Zaromb, F. M. (2010). Retrieval mode distinguishes the testing effect from the generation effect. *Journal of Memory and Language*, *62*(3), 227–239. doi:10.1016/j.jml.2009.11.010

Kelly, K., Wang, Y., Tamisha, T., & Neil, H. (2015). Defining mastery: Knowledge tracing versus consecutive correct responses. *In the Proceedings of the 8th International Conference on Educational Data Mining*. Educational Data Mining.

Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, *19*(3), 239–264. doi:10.100710648-007-9049-0

Koedinger, K. R., Carvalho, P. F., Liu, R., & McLaughlin, E. A. (2023). An astonishing regularity in student learning rate. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(13), e2221311120. doi:10.1073/pnas.2221311120 PMID:36940328

Koedinger, K. R., Pavlik, P., McLaren, B. M., & Aleven, V. (2008). Is it better to give than to receive? The assistance dilemma as a fundamental unsolved problem in the cognitive science of learning and instruction. *Proceedings of the 30th Annual Conference of the Cognitive Science Society*. Cognitive Science Society, Austin, TX.

Kulik, C.-L. C., Kulik, J. A., & Bangert-Drowns, R. L. (1990). Effectiveness of mastery learning programs: A Meta-Analysis. *Review of Educational Research*, *60*(2), 265–299. doi:10.3102/00346543060002265

Ladd, G. W., & Dinella, L. M. (2009). Continuity and change in early school engagement: Predictive of children's achievement trajectories from first to eighth grade? *Journal of Educational Psychology*, *101*(1), 190–206. doi:10.1037/a0013153 PMID:23060683

Lee, J.-E, Chan, J. Y.-C, Botelho, A, & Ottmar, E. (2022). Does slow and steady win the race?: Clustering patterns of students' behaviors in an interactive online mathematics game. Educational Technology Research and Development, 70(5), 1575–1599. https://doi.org/10.1007/s11423-022-10138-4

Lee, J.-E., Stalin, A., Ngo, V., Drzewiecki, K., Trac, C., & Ottmar, E. (2022). Show the flow: visualizing students' problem-solving processes in a dynamic algebraic notation tool.

Liu, A., Vanacore, K., & Ottmar, E. (2022). Reward-, but not error-based, feedback systems create micro-failures that support persistence-related learning behaviors [Manuscript Under Review].

Liu, Z., Cody, C., & Barnes, T. (2017). The antecedents of and associations with elective replay in an educational game: Is replay worth It? *Proceedings of the 10th International Conference on Educational Data Mining*. Educational Data Mining.

Long, P., Siemens, G., Conole, G., & Gašević, D. (2011, February 27). Message from the LAK 2011 General & Program Chairs. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*. LAK 2011: 1st International Conference on Learning Analytics and Knowledge, Banff Alberta Canada. https://dl.acm.org/doi/proceedings/10.1145/2090116

Lynch, S. D., Hunt, J. H., & Lewis, K. E. (2018). Productive struggle for all: Differentiated instruction. *Mathematics Teaching in the Middle School*, *23*(4), 194–201. doi:10.5951/mathteacmiddscho.23.4.0194

Macaruso, P., & Hook, P. E. (2007). Computer assisted instruction: Successful only with proper implementation. *Perspectives on Language and Literacy*, *33*(4).

Martens, R. L, Gulikers, J., & Bastiaens, T. (2004). The impact of intrinsic motivation on e-learning in authentic computer tasks. Journal of Computer Assisted Learning, 20(5), 368–376. https://doi.org/10.1111/j.1365-2729.2004.00096.x

McCoy, L. P. (1996). Computer-based mathematics learning. Journal of Research on Computing in Education, 28(4), 438–460. https://doi.org/10.1080/08886504.1996.10782177

Mu, T., Jetten, A., & Brunskill, E. (2020). Towards suggesting actionable interventions for wheel-spinning students. *Proceedings of The 13th International Conference on Educational Data Mining*. EDM.

Okano, H., Hirano, T., & Balaban, E. (2000). Learning and memory. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(23), 12403–12404. doi:10.1073/pnas.210381897 PMID:11035781

Paquette, L., & Baker, R. S. (2019). Comparing machine learning to knowledge engineering for student behavior modeling: A case study in gaming the system. *Interactive Learning Environments*, *27*(5–6), 585–597. doi:10.1080/10494820.2019.1610450

Paquette, L., Baker, R. S., & Moskal, M. (2018). A System-general model for the detection of gaming the system behavior in CTAT and LearnSphere. In C. Penstein Rosé, R. Martínez-Maldonado, H. U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren, & B. Du Boulay (Eds.), *In Artificial Intelligence in Education: 19th International Conference* (Vol. 10948, pp. 257–260). Springer International Publishing. 10.1007/978-3-319-93846-2_47

Paquette, L., Baker, R. S., & Ocumpaugh, J. (2014). *Reengineering the feature distillation process: A case study in the detection of gaming the system*. Springer.

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

Park, S. (2023). Discovering unproductive learning patterns of wheel-spinning students in intelligent tutors using cluster analysis. *TechTrends*, *67*(3), 489–497. doi:10.100711528-023-00847-9

Patikorn, T., & Heffernan, N. T. (2020). Effectiveness of crowd-sourcing on-demand assistance from teachers in online learning platforms. *L@S 2020 - Proceedings of the 7th ACM Conference on Learning @ Scale*, (pp. 115–124). ACM. 10.1145/3386527.3405912

Prihar, E., Patikorn, T., Botelho, A., Sales, A., & Heffernan, N. (2021). Toward personalizing students' education with crowdsourced tutoring. *L@S 2021 - Proceedings of the 8th ACM Conference on Learning @ Scale*, (pp. 37–45). ACM. 10.1145/3430895.3460130

Prihar, E., Syed, M., & Ostrow, K. (2022). *Exploring common trends in online educational experiments*. Springer.

Reich, J. (2020). Algorithm-Guided Learning At Scale. In Failure to Disrupt: Why Technology Alone Can't Transform Education (pp. 47–76). Harvard University Press. doi:10.2307/j.ctv322v4cp.6

Richey, J. E., Zhang, J., Das, R., Andres-Bray, J. M., Scruggs, R., Mogessie, M., Baker, R. S., & McLaren, B. M. (2021). Gaming and confusion explain learning advantages for a math digital learning game. In I. Roll, D. McNamara, S. Sosnovsky, R. Luckin, & V. Dimitrova (Eds.), *In Artificial Intelligence in Education: 22nd International Conference* (pp. 342–355). Springer International Publishing. 10.1007/978-3-030-78292-4_28

Ritter, S., Yudelson, M., Fancsali, S. E., & Berman, S. R. (2016). How mastery learning works at scale. *Proceedings of the Third (2016) ACM Conference on Learning @ Scale*, (pp. 71–79). ACM. 10.1145/2876034.2876039

Roediger, H. L. III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255. doi:10.1111/j.1467-9280.2006.01693.x PMID:16507066

Rohrer, D., Dedrick, R. F., & Burgess, K. (2014). The benefit of interleaved mathematics practice is not limited to superficially similar kinds of problems. *Psychonomic Bulletin & Review*, *21*(5), 1323–1330. doi:10.375813423-014-0588-3 PMID:24578089

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

Rowe, E., Almeda, M. V., Asbell-Clarke, J., Scruggs, R., Baker, R., Bardar, E., & Gasca, S. (2021). Assessing implicit computational thinking in Zoombinis puzzle gameplay. *Computers in Human Behavior*, *120*, 106707. doi:10.1016/j.chb.2021.106707

Rutherford, T., Farkas, G., Duncan, G., Burchinal, M., Kibrick, M., Graham, J., Richland, L., Tran, N., Schneider, S., Duran, L., & Martinez, M. E. (2014). A randomized trial of an elementary school mathematics software intervention: Spatial-Temporal Math. *Journal of Research on Educational Effectiveness*, *7*(4), 358–383. doi:10.1080/19345747.2013.856978

Rutherford, T., Kibrick, M., Burchinal, M., Richland, L., Conley, A., Osborne, K., Schneider, S., Duran, L., Coulson, A., Antenore, F., Daniels, A., & Martinez, M. E. (2010). *Spatial temporal mathematics at scale: An innovative and fully developed paradigm to boost math achievement among all learners*. In Online Submission. https://eric.ed.gov/?id=ED510612

Sales, A. C., & Pane, J. F. (2019). The role of mastery learning in an intelligent tutoring system: Principal stratification on a latent variable. *The Annals of Applied Statistics*, *13*(1), 420–443. doi:10.1214/18-AOAS1196

Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology. Human Learning and Memory*, *5*(2), 179–187. doi:10.1037/0278-7393.5.2.179

Shute, V. J., Wang, L., Greiff, S., Zhao, W., & Moore, G. (2016). Measuring problem solving skills via stealth assessment in an engaging video game. *Computers in Human Behavior*, *63*, 106–117. doi:10.1016/j.chb.2016.05.047

Siew, N. M, Geofrey, J, & Lee, B. N. (2016). Students' Algebraic Thinking and Attitudes towards Algebra: The Effects of Game-Based Learning using Dragonbox 12 + App. The Research Journal of Mathematics and Technology, 5(1). https://doi.org/2163-0380

Sinatra, G. M., Heddy, B. C., & Lombardi, D. (2015). The Challenges of defining and measuring student engagement in science. *Educational Psychologist*, *50*(1), 1–13. doi:10.1080/00461520.2014.1002924

Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, *6*(4), 342–353. doi:10.3758/BF03197465

Smith, S. M., & Handy, J. D. (2014). Effects of varied and constant environmental contexts on acquisition and retention. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *40*(6), 1582–1593. doi:10.1037/xlm0000019 PMID:24797444

Smith, S. M., & Rothkopf, E. Z. (1984). Contextual enrichment and distribution of practice in the classroom. *Cognition and Instruction*, *1*(3), 341–358. doi:10.12071532690xci0103_4

Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science*, *10*(2), 176–199. doi:10.1177/1745691615569000 PMID:25910388

Stone, C., Quirk, A., Gardener, M., Hutt, S., Duckworth, A. L., & D'Mello, S. K. (2019). Language as thought: Using natural language processing to model noncognitive traits that predict college success. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, (pp. 320–329). ACM. 10.1145/3303772.3303801

Sun, C., Shute, V. J., Stewart, A. E. B., Beck-White, Q., Reinhardt, C. R., Zhou, G., Duran, N., & D'Mello, S. K. (2022). The relationship between collaborative problem solving behaviors and solution outcomes in a game-based learning environment. *Computers in Human Behavior*, *128*, 107120. doi:10.1016/j.chb.2021.107120

Vanacore, K., Ottmar, E., Sales, A., & Liu, A. (2023). Evaluating In-Program Decisions by Leveraging Cut Points and Regression Discontinuity Analysis for Causal Inference. In (Sales, A., Chair) *Causal Modeling of Log Data in Edu Tech Symposium*. National Council on Measurement in Education's Annual Meeting, Chicago, IL. USA.

Vanacore, K., Sales, A., Liu, A., & Ottmar, E. (2023). Benefit of gamification for persistent learners: Propensity to replay problems moderates algebra-game effectiveness. *Tenth ACM Conference on Learning @ Scale (L@S' 23)*. Learning @ Scale, Copenhagen, Denmark. https://doi.org/10.1145/3573051.3593395

Vanacore, K. P., Gurung, A., McReynolds, A. A., Liu, A., Shaw, S. T., & Heffernan, N. T. (2023). Impact of non-cognitive interventions on student learning behaviors and outcomes: An analysis of seven large-scale experimental inventions. *LAK23: 13th International Learning Analytics and Knowledge Conference*, 10. https://doi.org/10.1145/3576050.3576073

Wang, Y., Heffernan, N. T., & Heffernan, C. (2015). Towards better affect detectors: Effect of missing skills, class features and common wrong answers. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, (pp. 31–35). ACM. 10.1145/2723576.2723618

Wise, A. F., Knight, S., & Ochoa, X. (2021). What makes learning analytics research matter. *Journal of Learning Analytics*, *8*(3), 1–9. doi:10.18608/jla.2021.7647

Yudelson, M. V., Koedinger, K. R., & Gordon, G. J. (2013). Individualized Bayesian knowledge tracing models. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), *Artificial Intelligence in Education* (pp. 171–180). Springer. doi:10.1007/978-3-642-39112-5_18

Zhang, C., Huang, Y., Wang, J., Lu, D., Fang, W., Fancsali, S., Holstein, K., & Aleven, V. (2019). Early detection of wheel spinning: Comparison across tutors, models, features, and operationalizations. *Proceedings of The 12th International Conference on Educational Data Mining*. EDM.

KEY TERMS AND DEFINITIONS

Action Data: Data on the actions students take within a program, including, but not limited to their submission of answers to questions or problems, use of hints and scaffolding, and derivations in problem-solving. These data will vary based upon the actions available in the program.

Computer Assisted Learning Platforms (CALP): Programs that provide automated learning content and/or problems with the goal of students gaining knowledge or skills. These can include programs with varying levels of adaptivity from intelligent tutors with automated targeted feedback to digitally presented problems.

Desirable Difficulties: Elements of learning programs, systems, or courses that create conditions for productive struggle which improve learning.

Learning: Permanent changes in abilities or knowledge, including long term retention of information and transferability of skills outside of the direct context in which they were learned.

Performance: Execution of a task during the learning activity, including whether a problem was answered correctly or a task completed sufficiently.

Randomized Control Trial: A research design in which units (often students) are randomized into conditions, allowing for an evaluation of the effect of those conditions in the unit.

Time Data: Data on the time it takes a student to take an action or series of actions.

Quasi-Experimental Studies: Research designs that estimate the effects of a condition, though the units are not randomized into conditions. This must be done by accounting for confounding that occurs, which influences what units experience what conditions. Common quasi-experimental methods include propensity score matching and regression discontinuity design.