The Findings and Developed Tools from the Exploration of Procedural Pathways to Bolster Algebraic Knowledge

By

Siddhartha Prasad Pradhan

A thesis submitted in partial satisfaction of the
requirements for the degree of
Master of Science

in

Data Science

in the

Graduate Division

of the

Worcester Polytechnic Institute, Worcester

April 2024

The committee in charge:

Professor Erin Ottmar, Advisor

Professor Roee Shraga, Reader

Abstract

Algebra is essential for delving into advanced mathematical topics and STEM courses (Chen, 2013), requiring students to apply various problem-solving strategies to solve algebraic problems (Common Core, 2010). Yet, many students struggle with learning basic algebraic concepts (National Mathematics Advisory Panel (NMAP), 2008). Over the years, both researchers and developers have created a diverse set of educational technology tools and systems to support algebraic learning, especially in facilitating the acquisition of problem-solving strategies and procedural pathways. However, there are very few studies that examine the variable strategies, decisions, and procedural pathways during mathematical problem-solving that may provide further insight into a student's algebraic knowledge and thinking. Such research has the potential to bolster algebraic knowledge and create a more adaptive and personalized learning environment.

This multi-study project explores the effects of various problem-solving strategies on students' future mathematics performance within the gamified algebraic learning platform From Here to There! (FH2T). Together, these four studies focus on classifying, visualizing, and predicting the procedural pathways students adopted, transitioning from a start state to a goal state in solving algebraic problems. By dissecting the nature of these pathways—optimal, sub-optimal, incomplete, and dead-end—we sought to develop tools and algorithms that could infer strategic thinking that correlated with post-test outcomes. A striking observation across studies was that students who frequently engaged in what we term 'regular dead-ending behavior', were more likely to demonstrate higher post-test performance, and conceptual and procedural knowledge. This finding underscores the potential of exploratory learner behavior within a low-stakes gamified framework in bolstering algebraic comprehension. The implications of these findings are twofold: they accentuate the significance of tailoring gamified platforms to student behaviors and highlight the potential benefits of fostering an environment that promotes exploration without retribution. Moreover, these insights hint at the notion that fostering exploratory behavior could be instrumental in cultivating mathematical flexibility. Additionally, the developed tools and findings from the studies, paired with other commonly used student performance metrics and visualizations are used to create a collaborative dashboard–with teachers, for teachers.

Acknowledgements

I would like to express my sincere gratitude to my advisor, Professor Erin Ottmar, for leading, supporting, and entrusting me with this research. I believe my work would not have been complete without her continuous support and feedback for my research and thesis work. I thank her for taking time out of her busy schedule to revise and improve my thesis.

I am very grateful to Professor Roee Shraga for his valuable time advising me and reading my thesis, which helped me to improve the quality and readability of this thesis.

I also thank all the members in the WPI MAPLE lab and my peers in the larger LST community for sharing their valuable experiences and insights. Many thanks to Kirk Vanacore, who mentored me and introduced me to academic research, and Ashish Gurung, my guru, for his invaluable guidance in the initial phases of my thesis.

Additionally, the research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through an Efficacy and Replication Grant (R305A180401) and an NSF CAREER Grant (2142984) to Worcester Polytechnic Institute. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education.

Finally, I want to dedicate this work to my beloved family, who offer me selfless support and unconditional love. Get well soon Mama.

– Sid (April 2024)

Table of Contents

List of Figures

List of Tables

Introduction

Motivation

Over the past year, my role as a Graduate Research Assistant in the MAPLE Learning Sciences Lab has allowed me to delve deeply into data-driven methodologies, analysis, and software engineering for education in an applied interdisciplinary setting. Within this capacity and through working on several NSF and IES funded grants, we have utilized various data science and computer science techniques, including modeling student behaviors, predicting student learning outcomes, and creating interactive visualizations to make sense of large-scale action-level log data as students solve problems in various math technologies. By embracing an interdisciplinary outlook, we have harnessed the power of data mining and data modeling within learning analytics research, paving the way for the adoption of more sophisticated analytical tools in the field. Much of this research has involved leading projects and writing of conference papers and manuscripts. Since beginning my Masters, we have first-authored two accepted conference papers (Pradhan et al., 2024a; 2024b), have collaborated on several other projects, and am currently working on submitting two first-authored manuscripts to journals. We have also engaged in several other educationally driven research projects, working alongside teachers, students, data analysts and faculty with varying expertise to broaden my understanding of learning.

For my thesis, we will focus on a subset of these projects that have guided a majority of my work. In particular, this thesis will describe the process of building a tool to classify students' procedural pathways when solving algebraic problems; developing visualizations for students' attempts and the respective classifications; and conducting data analyses to explore whether the choice of procedural pathways predicts differentiated learning outcomes. By using a data-driven approach we plan on narrowing the gap in educational data mining regarding the study of students' variation in procedural pathways.

Thesis Introduction

Algebra is an essential topic for success in advanced mathematical topics and other STEM courses (Chen, 2013). To be proficient in algebra, students need to have a strong understanding of algebraic notations, conceptual and procedural knowledge, as well as the ability to use that knowledge flexibly (Rittle-Johnson et al., 2015; Schoenfeld, 2007). Further, students need to be able to apply various problem-solving strategies to efficiently solve algebraic problems (Common Core, 2010). Yet, many students struggle with learning basic algebraic concepts and applying that knowledge or appropriate problem-solving strategies to unfamiliar problems (National Mathematics Advisory Panel [NMAP], 2008).

To effectively apply algebraic knowledge to unseen or unfamiliar problems, students need to develop not only the ability to solve problems accurately but also *procedural flexibility* to make strategic choices adaptively (Baroody, 2003; Threlfall, 2009; Verschaffel et al., 2009). Over the years, both researchers and developers have created a diverse set of tools and systems to bolster algebraic learning, especially in facilitating the acquisition of problem-solving strategies. Notably, rule-based approaches have stood out in the development of Intelligent Tutoring Systems (ITS), with the primary goal of strengthening the acquisition of procedural knowledge that's pivotal to algebraic comprehension. These approaches find their foundation largely in the cognitive theories presented in ACT-R (Anderson, 2014). A variety of cognitive tutors have emerged over time, all aiming to aid learners in achieving mastery across various subjects (Anderson et al., 1995; Corbett et al., 1997; Ritter & Fancsali, 2016). While these cognitive tutors are designed to provide learners with adaptive feedback and personalized guidance (Aleven et al., 2006), the procedural pathways available to students are often constrained by what the problem creator considers essential for mastering the core concepts (see Figure 1a).

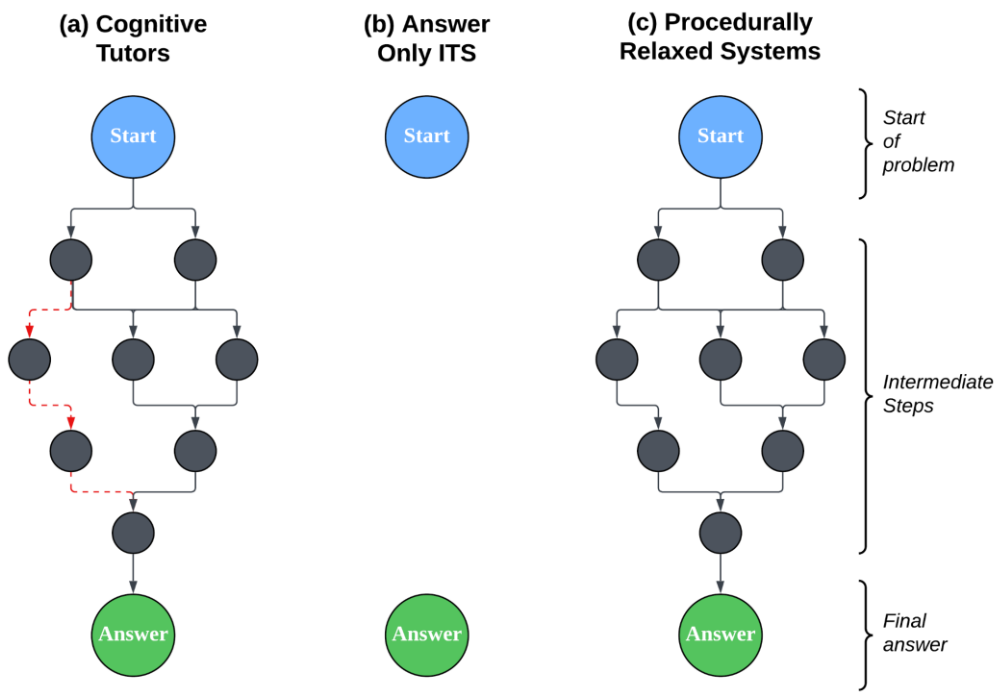


Figure 1. Procedural pathways in digital learning systems

In contrast, some ITS have adopted an alternative approach, developing systems that merely require students to enter their answers to the problems (see Figure 1b), without demonstrating the procedural comprehension necessary to solve them (Heffernan & Heffernan, 2014). Despite the ambiguity surrounding the procedural pathways chosen by the learner, the use of such systems has led to better learning outcomes (Murphy et al., 2020) and the availability of feedback has been observed to be beneficial in enhancing learning (Koedinger et al., 2010; Mendicino et al., 2009).

Though both approaches effectively facilitate the acquisition of mathematical knowledge (Murphy et al., 2020; Pane et al., 2014; Steenbergen-Hu & Cooper, 2013), little is known regarding the various procedural pathways learners might potentially employ in formulating a solution. With the rapid development in technology and ensuing innovations in Intelligent Tutoring Systems (ITS), researchers and developers have been investigating the efficacy of implementing novel methodologies to aid learners in acquiring algebraic knowledge. A subset of these educational technologies, such as Graspable Math (Weitnauer et al., 2016) and From Here to There! (FH2T; freely available at graspablemath.com/projects/fh2t; Ottmar et al., 2015), have embraced dynamic procedural pathways for teaching algebra. Specifically, FH2T adopts a distinctive gamified and dynamic procedural approach: learners are presented with an algebraic expression as the starting state and a transformed version of that expression as the goal state. Students can traverse any procedural pathway they prefer, with all mathematically valid transformations being permissible (see Figure 1c). This architecture inherently grants learners the autonomy to explore various procedural avenues in their exploration of the transformations necessary to attain the goal state. Such a modality can shed light on the diversity of procedural pathways chosen by learners.

While most traditional math learning tools guide students towards producing correct and efficient answers, gamified learning environments such as FH2T, employ a different approach to help students learn core math concepts and develop procedural flexibility. Gamified environments have been shown to help students remain engaged as they explore multiple possible solution paths and practice core mathematical concepts, rather than solely producing a correct answer (Clark et al., 2016; Jere-Folotiya et al., 2014; Wouters et al., 2013). Several studies (Chan et al., 2022; Decker-Woodrow et al., 2023; Hussein et al., 2019; Wouters et al., 2017; Young et al., 2012) have found that gamified math learning applications are more effective than non-gamified counterparts in improving students' algebraic learning. Additionally, various prior studies have reported on their exploration of the efficacy of FH2T's gamification model in improving different aspects of the in-game behavior that can predict better learning outcomes (Chan et al., 2022b; Vanacore et al., 2023). However, to the best of our knowledge, very little exploration regarding the variation in the procedural pathways adopted by students in their attempt to reach the goal state has been studied.

Contribution

Our contributions are summarized as follows:

- We developed a novel tool called 'MathFlowLens' that classifies the *entire* procedural pathway of students as optimal, sub-optimal, incomplete, and dead-end. Subsequently, the tool uses the classifications and represents students' procedural pathways as a sequential network visualization (Study 1).
- We explore the relation between the classifications and the constructs of algebraic knowledge: conceptual knowledge, procedural knowledge, and procedural flexibility. Surprisingly, the results indicate that students who have a high average number of dead-end attempts per problem have higher conceptual and procedural knowledge (Study 2).
- We further dissect the nature of dead-end attempts and reveal that students who exhibit what we call 'Regular Dead-ending' behavior have significantly higher algebraic knowledge scores (Study 3).
- We develop a data and research-driven interactive dashboard (mathflowlens.com) by following a co-design methodology and collaborating with teachers. In addition to using the developed MathFlowLens tool from Study 1, we incorporate various other visualizations and student performance metrics in the dashboard to provide teachers

with a detailed overview of their students' performance. This dashboard is generalizable and usable by any learning platform that logs detailed student transactional data (Study 4).

Study 1. MathFlowLens: A Classification And Visualization Tool for Analyzing Students' Procedural Pathways.

** This study is part of a paper currently under review in the Journal of Educational Technology Research and Development. Edits have been made to improve the readability and flow of the thesis.*

Citation: Pradhan, S., Ottmar, E., Gurung, A. & Lee, J. (*in review*). MathFlowLens: A Classification And Visualization Tool for Analyzing Students' Procedural Pathways. *Education Tech Research and Development*.

Introduction

Previous studies regarding student learning in gamified learning environments have mainly examined correlations between in-game behaviors and positive learning outcomes. Students exhibiting behaviors such as consistent in-game progress (Hulse et al., 2019; Martin et al., 2015, Shute et al., 2015), longer pause time before problem solving (Chan et al., 2022), higher propensity to reattempt problems (Chen et al., 2020; Vanacore et al., 2023) and exploratory behavior (Pradhan et al., 2024a) tend to have better learning outcomes. As students' in-game behavioral patterns play a significant role in student learning (Chan et al., 2023), developers of gamified learning platforms may need to pay close attention to the game design to ensure they do not penalize more exploratory or creative but less efficient behaviors. Consistently employing such problem-solving strategies may give students a more nuanced understanding of mathematical structures and transformations (Pradhan et al., 2024). However, very few studies have examined how students' variable problem-solving strategies influence the underlying mechanism of learning and other behaviors when interacting with the game (Hulse et al., 2019).

While the studies conducted by Lee et al. (2022b; 2022c) partially explored the variability in students' solution strategy by examining the efficiency and productivity of the first step; to the best of our knowledge, there have been no studies that classify the *entire* problem-solving pathways from start to finish and explore the variation of strategies in online learning tools or platforms. This is in part because a majority of educational technologies record students' entered answers rather than log moment-by-moment mathematical transformations and expressions that show how students arrive at their solutions. This often results in a lack of process-based data to explore variability in students' mathematical ideas and problem-solving strategies (Lee et al. 2022c). However, documenting and visualizing students' preferred solution pathways and the variation in strategy may provide valuable insight into student's mathematical thinking and knowledge, particularly compared to binary problem correctness information.

Furthermore, in order to support students' algebraic learning and flexibility, it is vital for teachers to examine the variable strategies, decisions, and behaviors that students exhibit during problem-solving. However, in traditional learning environments, teachers often have difficulties keeping track of and monitoring students' problem-solving strategies and learning progress (Asquith et al., 2007). Understanding the variation in students' problem-solving strategies and preferred pathways in gamified learning applications may help teachers identify students' common misconceptions and gaps in knowledge. By knowing different student problem-solving pathways and strategies, teachers may actively compare different solution strategies with subsequent student discussions to highlight the differences and similarities of the different pathways.

In sum, by acknowledging the variability in students' strategies (e.g. creative, exploratory, and efficient), both teachers and students may stop relying exclusively on a single procedural pathway over others that are equally effective or more efficient, leading to improved procedural flexibility (Star et al., 2016). Further, developers and teachers alike could consider the individual differences in students and their behaviors to create more adaptive and supportive learning environments.

The current study

In this study, we address this gap in research and practice by leveraging process-level log data from a digital algebraic learning game. In particular, we introduce a novel tool called "*MathFlowLens*" that utilizes graph search algorithms such as A-star or Dijkstra's shortest path algorithm (Mehlhorn & Sanders, 2007) to reveal problem-solving pathways that are optimal, sub-optimal, incomplete, and occasionally paths that result in dead-ends. Additionally, MathFlowLens utilizes these classifications to create interactive *network visualizations* that represent students' paths of algebraic transformations, problem-solving pathways, and solution strategies in the game. Specifically, the aims of this study are to:

- *Development Phase 1:* Create a tool that can classify various problem-solving pathways and strategies that students employ while problem-solving.
- *Development Phase 2:* Create interactive network visualizations of students' problem-solving pathways and variations in their employed strategies.

Background

Utilizing Log-file data from game-based learning platforms to explore student problem-solving.

Digital game-based learning is a modern and technologically driven instructional approach for mathematics learning, in which students may explore content in a more low-stakes and relaxed setting compared to other math learning technologies, where students solve traditional textbook problems. Such game-based learning platforms can provide students with multiple learning pathways, conceptual reinforcement, and cognitive enhancement with the use of sound, image, and interactivity (Dede, 2009; Gee, 2003). The interactivity and flexibility of game-based learning environments may prompt students to display exploratory behaviors (Pradhan et al., 2024a) and encourage re-attempting problems (Chen et al., 2020; Vanacore et al., 2023), both of which lead to better learning outcomes. Further, these aspects of games help engage students while practicing mathematical concepts (Clark et al., 2016; Jere-Folotiya et al., 2014; Wouters et al., 2013), especially those who struggle in a traditional classroom setting (Moses & Cobb, 2002; Kiili et al. 2015).

A systematic literature review (Hussein et al., 2019) comparing game-based learning with other methods of instruction in K-12 mathematics education revealed that a majority of studies reported positive outcomes for knowledge acquisition. Similarly, for perceptual and cognitive skills, and affective, motivational, and behavioral changes, most studies reported positive outcomes with a game-based intervention. Furthermore, several studies revealed the effectiveness of the gamified system in decreasing mathematical errors and improving student understanding (Chan et al., 2022, 2023; Decker-Woodrow et al., 2023; Hulse et al., 2019; Ottmar et al., 2015).

In addition to improving learning and engagement and decreasing mathematical errors, click-stream or log-file data collected in gamified learning applications can provide researchers and developers with a plethora of vital information about students' problem-solving processes (Gobert et al., 2013). The fine-grained and detailed log data contains aspects of students' behaviors and provides researchers the opportunity to explore the relation between interaction patterns and learning outcomes (Crossley et al., 2019). There is growing research that suggests log-file data is strongly predictive of short- and long-term engagement, interest, and learning in mathematics (Ocumpaugh et al., 2016). Most studies in the field of game-based learning have primarily used log-file data to understand students' in-game behaviors, or the validation of the gamified elements (Alonso-Fernandez et al., 2019; Cano et al., 2018). While these studies provide useful information regarding the relationship between behavioral patterns and student learning, Hulse et al. (2019) claim that researchers and developers must pay close attention to the underlying learning mechanisms and behaviors that are triggered when students interact with the game. Yet, to the best of our knowledge, researchers have rarely used log-file data to explore variability in student's problem-solving strategies and pathways. While Lee et al. (2022b; 2022c) partially achieved this by hand-coding the productivity of the first step to explore factors leading to higher productive first steps, they did not classify the productivity or efficiency of the *entire* solution strategy. Raw log-file data, especially those that record moment-by-moment transformations and expressions made by students while solving a problem, may contain valuable information regarding the employed solution strategies and students' mathematical thinking. Further, the variation in problem-solving strategies and pathways employed by students when exploring and reattempting problems in a gamified learning system may provide valuable insight into students' mathematical knowledge.

Using Data Visualizations to Identify Student Problem-Solving Pathways

Identifying student strategies or problem-solving pathways in online learning environments is often challenging for researchers or teachers due to the complexity of pathways and the uncertainty about which patterns to explore in the data (Wang et al., 2017). Applying *data visualizations* to student problem-solving pathways can help researchers and teachers quickly identify what students do, and where they are stuck, and better understand the students' cognitive process during problem-solving activities.

Several empirical studies have applied data mining and advanced data visualization techniques to identify student strategies or problem-solving pathways (Hurtig et al., 2022;

Sinha & Aleven, 2015; Wang et al., 2017; Xia et al., 2021). One of the most widely used data visualization techniques to present students' learning pathways is the *Sankey diagram*. For example, Wang et al. (2017) developed a visualization tool called "PathViewer" to identify the sequence of students' problem-solving paths in programming works. They utilized the Sankey diagrams to depict the students' problem-solving pathways and natural language processing to identify the most prevalent sequences and loops. Using PathViewer, they successfully identified sequential patterns of students' coding and common causes of failure. Similarly, Xia et al. (2021) created a tool named "QLens" to visualize elementary and middle students' trajectory data across various steps and stages in an online math learning platform. They also used Sankey diagrams to depict students' problem-solving logic, engagement, and difficulties encountered. The following case studies and interviews showed the usefulness and effectiveness of QLens in providing information about students' problem-solving processes and engagement. As such, while Sankey diagrams are useful for visualizing various problem-solving pathways, they illustrate the data flowing in only one direction (e.g., start state to goal state or correct answer). They do not depict *multidirectional pathways* within the learning process, for example, how students withdraw their strategy, return to the previous state, correct errors, and take a more efficient problem-solving pathway.

One alternative way to visualize student problem solving pathways is through *sequential network visualization.* Sequential network visualizations illustrate not only the direction and frequency of the pathway through arrows but also how nodes (e.g., concepts, events) are interconnected with each other. As an example, Hurtig et al., (2022) investigated students' pathways to correct answers in college chemistry classes. They represented students' attempts, both incorrect and correct, as nodes in a network and visualized how they reached correct answers. The usability test of this tool with faculty members showed a satisfactory level, and they found the tool useful for monitoring and better understanding students' comprehension. In our study, we represented problem-solving steps as nodes in networks and created sequential network visualizations to illustrate multidirectional student problem-solving pathways to reach goal states in an online mathematics learning game.

Context of the Study

Game Description

FH2T, a gamified learning platform that was developed by Ottmar and colleagues (2015), is a dynamic gamified learning application developed based on several learning theories such as perceptual learning, embodied cognition, and gamification. While traditional algebraic teaching in school focuses on the memorization of abstract and arbitrary rules, FH2T helps improve students' knowledge of arithmetic for algebra learning by helping students identify the structure of algebraic expressions and think more flexibly about mathematical operations and properties (Chan et al., 2022). Several empirical studies have shown that FH2T improves students' procedural learning, conceptual understanding, understanding of equivalence, and flexibility in algebra, as well as decreasing mathematical errors (Chan et al., 2022; Decker-Woodrow et al., 2023; Hulse et al., 2019; Ottmar et al., 2015).

FH2T incorporates interactive virtual objects to represent algebraic equations, allowing students to apply mathematical principles through dynamic gestures (e.g. tapping or dragging) in a virtual environment. In doing so, students can realize that mathematical transformations or steps are dynamic, rather than procedural steps. In particular, students learn algebra through puzzle-based problems, in which each problem has a starting expression (i.e. 'start state') that needs to be transformed into a predefined and explicit ending expression (i.e. 'goal state'). While these two states are mathematically equivalent, they differ perceptually (see Figure 2a). In Figure 2, one may solve the problem by transforming the start state (i.e. "44+56+a+72+28") into the goal state (i.e. "100+a+100") using mathematically valid actions.

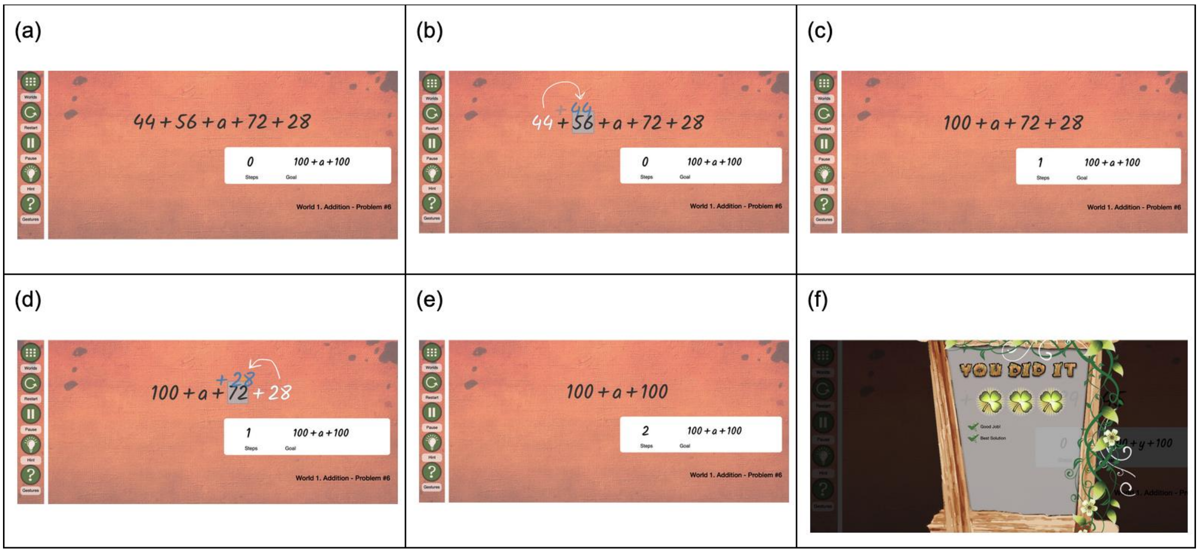


Figure 2. Example solution steps for problem 6 in FH2T

The game also rewards students based on efficiency (i.e. the number of steps a student took to complete a particular problem). It rewards three clovers to students who solve the problem with efficient solutions (see Figure 2f). The number of clovers decreases if a student takes more solution steps than the efficient solution. In addition, the game allows students to *reset* at any point while attempting a problem. The game also allows students to *reattempt* a completed problem more efficiently.

The game comprises 252 problems organized into 14 'worlds'. Each world covers various mathematical concepts (e.g. addition, multiplication, fraction). The game orders problems based on difficulty, and only allows students to advance to the next world when they complete 14 consecutive problems (Lee et al., 2022a).

Dataset Used for MathFlowLens Development

In order to classify students' problem-solving strategy and create data visualizations, we used data collected from a large Randomized Control Trial (RCT) comparing student learning from FH2T to two technological interventions and an active control condition as described by Ottmar (2023). In the RCT, the authors recruited a total of 4,092 7th-grade students from 11 middle school students within a large suburban district in the United States from September 2020 to

April 2021 amidst the COVID-19 pandemic. The results of this RCT are available in Decker-Woodrow et al. (2023).

In the RCT, researchers randomly assigned 1,649 students to the FH2T condition, out of which 52.6% were male and 47.4% were female. In terms of racial identity, a majority of the sample identified as White (49.8%), 24.8% as Asian, and 16.4% as Hispanic/Latino. The remaining 9% identified with either multiple races or various other racial categories. The racial distribution of the sample was found to be representative of the school district. On average, the students attempted 112.57 problems (*SD* = 55.06) and completed 111.06 problems (*SD* = 55.05) out of 252 problems in the game.

While the larger dataset from the study contains various levels of aggregation (e.g. student and problem level), we use the log-file data containing real-time or in-the-moment data that show how students arrive at their solutions. The FH2T platform records detailed logs of mathematical expressions or steps logged after each transformation a student made, including timestamps and the different types of errors made by students. This feature-rich raw log file data in FH2T enabled us to identify students' various problem-solving pathways at the fine-grained level and build network visualizations. The dataset is publicly available for researchers on OSF after a data-sharing agreement (link: https://osf.io/r3nf2/).

Development of MathFlowLens

In an effort to address a gap in game-based algebraic learning research regarding variability in students' solution strategies, we developed a tool, *MathFlowLens*, which is capable of identifying and classifying various types of attempts. In the following subsections, we describe the creation, technical components, and inner workings of this tool.

Development Phase 1: Creation of a classification tool identifying various problem-solving pathways and strategies.

Solution Steps as a Directed Graph

Each mathematical expression or state in a solution is represented as a node in a network. Similarly, each transformation from one expression to another is represented as a directed edge from one node to the other. In doing so, we represent a series of actions (i.e. steps) taken by a student to solve a problem as a directed graph. For example, for the sequence of steps in Figure 2, the start state (i.e. "44+56+a+72+28", Fig. 2a), step 1 (i.e. "100+a+72+28", Fig. 2c), and goal state (i.e. "100+a+100", Fig. 2e) would each constitute a node in the graph. Similarly, the transformations: start state to step 1 (i.e. ["44+56+a+72+28" → "100+a+72+28"], see Fig. 2b) and step 1 to the goal state (i.e. ["100+a+72+28" → "100+a+100"], see Fig. 2d), would represent directed edges to and from the respective nodes in the graph.

Alternatively, similar to most problems in FH2T, one can solve the example presented above (see Figure 2) with multiple equally efficient solutions. In the case above, one may choose to add "72+28" from the start state as a first step to get "44+56+a+100" as step 1 rather than "100+a+72+28". Next, one can add "44+56" as the last step to reach the goal state (i.e. "100+a+100"). The graph structure can represent any number and combination of procedural pathways, regardless of the solution strategy employed.

After populating a graph with several solution steps (i.e. problem attempts), we used efficient graph traversal algorithms such as A* or Dijkstra's shortest path algorithm (DSPA) to identify all efficient solutions or paths from the start state to the goal state. In our implementation, we chose DSPA for its ability to calculate weighted distances, providing us with the flexibility to reward or penalize certain transformations. Such graph traversal algorithms find the shortest distance between two nodes in a weighted and directed graph (Mehlhorn & Sanders, 2007).

Creating Graphs for All Problems in FH2T

Using the expressions and transformations recorded in the log data, we created and populated individual graphs for each problem. The graphs were created in Python 3.11 using NetworkX (Hagberg, 2008) and pandas (McKinney, 2010) packages. Additionally, as the creation of a single graph relies solely on raw data from a single problem, we used multi-threading to make this process scalable and efficient. For each problem, we included all expressions (or steps) in the log data, including the start state and goal state, as nodes in the graph and the respective transformations to and from nodes as the directed edges. Furthermore, we recorded the number of students who made a particular transformation, which allowed us to determine the preferred or common solution strategies and transformations. Similarly, we recorded the total number of errors students made in a particular state to help identify expressions that led to more errors. This process resulted in 252 graphs, one for each problem in the FH2T application. Note, that for this initial exploratory study, we assigned equal weights to each transformation to find the optimal paths using DSPA, and consequently, the minimum number of steps required to reach the goal state.

Identified Classifications for Solution Steps/Attempts

Based on the graphs created, we identified three types of problem attempts: optimal, sub-optimal, and incomplete paths. *Optimal* paths are procedural pathways that use the least number of steps or transformations to reach the goal state. Similarly, *sub-optimal* paths are inefficient pathways that use more steps than the optimal paths to reach the goal state. Conversely, incomplete paths are attempts that did not reach the goal state. However, we observed two distinct types of incomplete attempts: 1) incomplete paths that were unique, and no student had reused them to reach the goal state, and 2) incomplete paths that students had reused to reach the goal state. We labeled the former as *dead-end* paths and the latter simply as *incomplete* paths. In other words, an incomplete pathway occurs when students stop using the current procedural pathway and decide to reset, even though that pathway leads to the solution. Distinctly, a dead-end path represents a pathway in which students cease progress before completing the problem, unlike incomplete pathways, no student has ever traversed them successfully to reach the goal state.

For example, in the problem given in Figure 2, if the raw log contains the four problem attempts from Figure 3, we can create a single graph representing all problem attempts. In doing so, one can identify the different attempt classifications. The attempt with optimal classification completes the problem with the least possible number of transformations (i.e., two transformations), whereas the sub-optimal attempt is complete but with a greater number of transformations than the optimal classification (i.e., three transformations). Next, the dead-end attempt is a unique incomplete attempt (i.e. the path has never been fully traversed to reach the goal state). Finally, the incomplete attempt makes the same first transformation as sub-optimal and optimal solutions, however, the attempt is reset before reaching the goal state. Note that it is this resemblance to a completed attempt that leads this attempt to be classified as incomplete rather than a dead-end pathway.

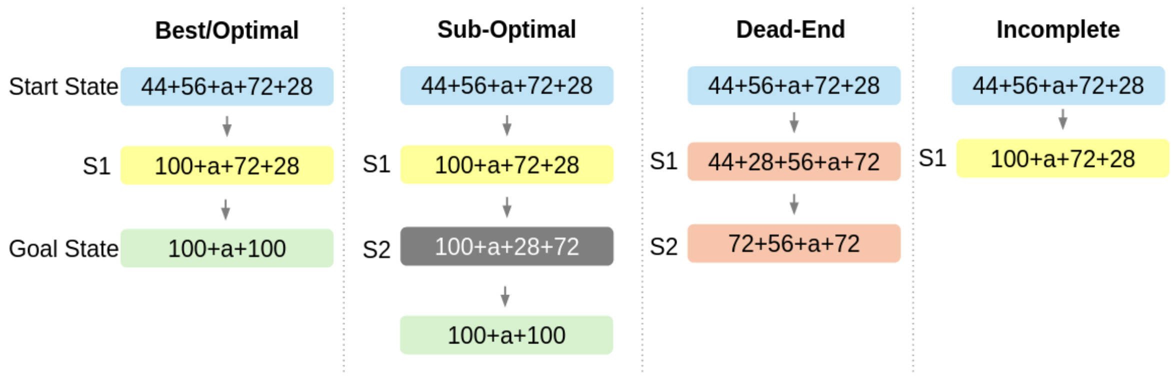


Figure 3. Identified classifications for student pathways

Alternative Classifications Using Binning Factor

As described above, the classifications inherently depend on the attempts that are included or excluded from the graph. For example, take an incomplete attempt by student A: [‘Start State’ → ‘Step 1’]. If we populated a graph with the entirety of the sample’s attempts, there may exist a student B, presumably with higher prior knowledge, who took the same path as student A but reached the goal state, with the attempt: [‘Start State’ → ‘Step 1’ → ‘Step 2’→ ‘Goal State’]. Based on the definition provided above, the algorithm would classify this path as incomplete. However, researchers may be interested in defining attempts based on various grouping or binning factors, for example: pre-knowledge scores, class ID, teacher ID, student ID, etc. By altering the students included in a particular network, we can constrain or relax the definitions of the classifications. This flexibility of creating multiple networks for a single problem based on a binning factor allows us to explore variability in the defined classifications and localize the definitions to the bins.

Development Phase 2: Applying classifications to create interactive visualizations of student problem-solving pathways and strategies.

While the graph creation process described above helped us identify four distinct pathways, the resulting graph data can then be used to create interactive sequential network visualizations, allowing us to examine students' various solution pathways and strategies. In our implementation, we used JavaScript and the d3.js library to render the visualizations. Each node in the visualization is color-coded based on the identified classifications from above. To distinguish the start and goal states from other states, we color-coded them blue and green, respectively. Similarly, we color-coded the optimal states and transformations as yellow, sub-optimal pathways as gray, and dead-end pathways as red (See Figure 4). Note that, students have reused that pathway in an optimal or sub-optimal manner to reach the goal state based on the definition of incomplete attempts. As a result, such incomplete attempts can not be visualized directly.

Furthermore, in the visualization, the thickness of an edge connecting two nodes and the size of the directional arrow is proportional to the number of attempts that made that transformation. This allows us to quickly identify the students’ preferred or common transformations and pathways. An example of this visualization is given in Figure 4, which represents all attempts in the log-file data for problem 6 in FH2T (start state: 44+56+a+72+28, goal state: 100+a+100).

In this case, there are two distinct optimal solutions (i.e., paths colored in yellow), which both require two steps to reach the goal state. The visualization has a small number of pathways indicating that there was little variation in students' problem-solving strategies. Additionally, based on the thickness of the edges and the size of the arrows, most students employed optimal problem-solving strategies.

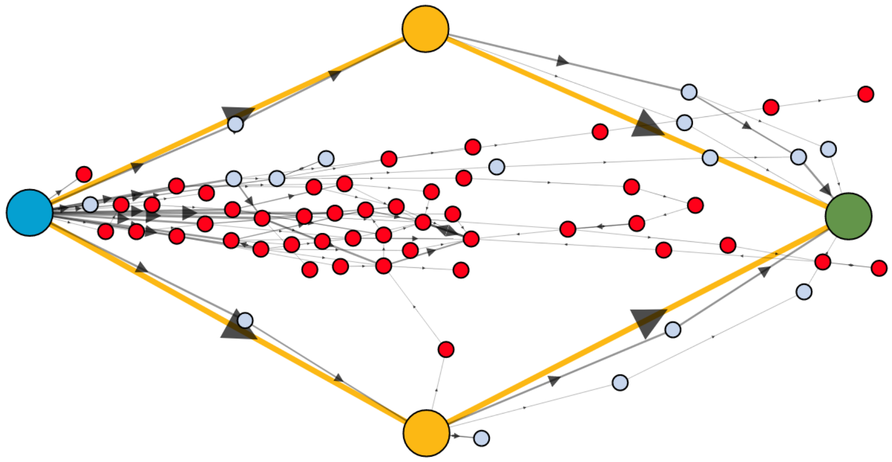


Figure 4. Network visualization for all student attempts for problem 6 in FH2T

Interactive elements of the visualization

An important feature of this visualization is *interactivity*, which allows users to drill down and focus on specific pathways or nodes. For example, hovering on the nodes using a mouse reveals additional information, such as the mathematical expression and the node's classification (see Figure 5a). Similarly, users can drag and drop each node, allowing them to focus on the relevant and interesting pathways. Additionally, researchers and teachers using this visualization may be interested in knowing what pathways lead to a particular node, or, conversely, what pathways lead to the goal state from a particular node. By holding the 'control' key while clicking a node, the visualization displays all paths that lead to the node from the start state and all paths that lead to the goal state from that node. Finally, as the visualization can get cluttered when there are many distinct pathways, users may choose to only show the optimal pathways and hide the sub-optimal and dead-end pathways (See Figure 5b).

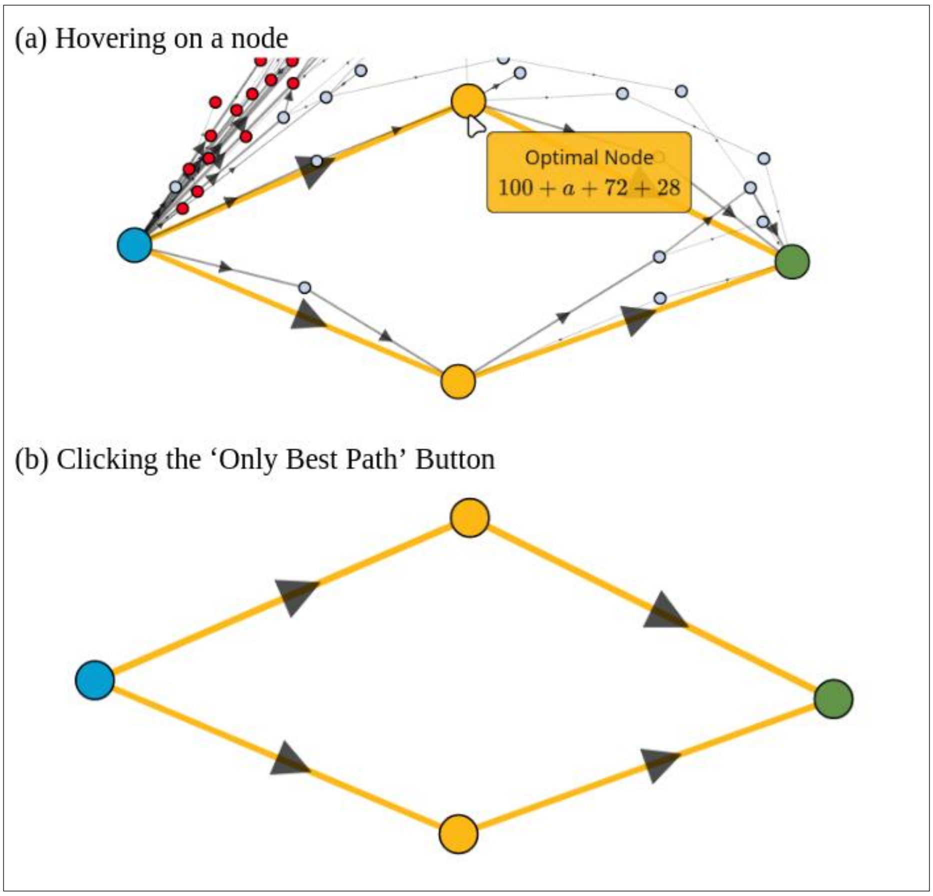
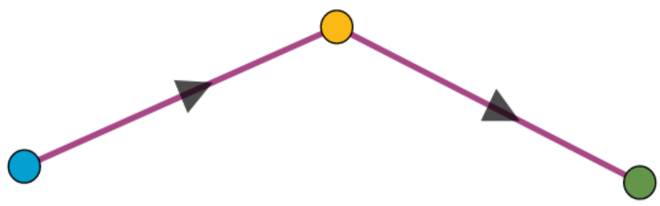
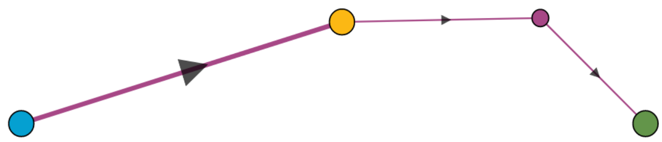
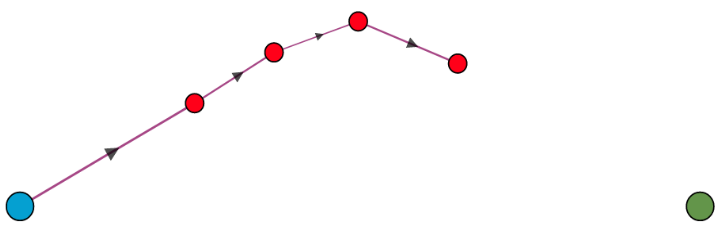
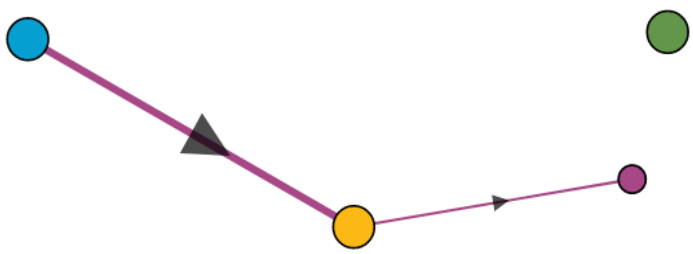


Figure 5. Demonstration of interactive features of MathFlowLens visualization

Another important feature of the interactive graph is its ability to visualize individual student attempts. By selecting a student ID and a problem attempt, the graph highlights the student's chosen problem-solving pathway in purple. An example of this feature is given in Table 1, in which we have presented 4 student attempts each belonging to the four identified attempt classifications: optimal, sub-optimal, dead-end, and incomplete. A live demonstration of this visualization can be found at mathflowlens.com in the Graph Diagram component.

Table 1. Visualization of sample student attempts and respective classifications

Student's Solution Strategy and Steps	Attempt Pathway Visualization
Optimal Attempt $44 + 56 + a + 72 + 28$ $100 + a + 72 + 28$ $100 + a + 100$	
Suboptimal Attempt $44 + 56 + a + 72 + 28$ $100 + a + 72 + 28$ $100 + a + 28 + 72$ $100 + a + 100$	
Dead-end Attempt $44 + 56 + a + 72 + 28$ $44 + 84 + a + 72$ $128 + a + 72$ $200 + a$ $a + 200$	
Incomplete Attempt $44 + 56 + a + 72 + 28$ $100 + a + 72 + 28$ $100 + a + 28 + 72$	

Discussion

Algebra is an essential topic to delve into advanced mathematical topics and other STEM courses (Chen, 2013), yet most students struggle with basic algebraic concepts (NMAP, 2008).

Despite widespread acknowledgment regarding the importance of procedural flexibility, and the facets of algebraic knowledge in general, limited works have prioritized classifying strategy pathways or explored what type of student's problem-solving strategies or pathways lead to improved algebraic learning. Furthermore, previous studies regarding student learning in gamified learning environments have mainly focused on correlations between in-game behaviors and positive learning outcomes (Hulse et al., 2019; Martin et al., 2015, Shute et al., 2015). This study aimed to narrow the gaps in practice by introducing a novel tool, 'MathFlowLens,' to reveal problem-solving pathways, and also to address gaps in research by exploring students' problem-solving strategies and the variation in solution pathways.

Classifying and Visualizing Student Problem-Solving Pathways

In this study, we presented the development process of MathFlowLens, which used raw log data from FH2T to classify each student attempt using graph theory and shortest path finding algorithms, and subsequently represented students' solution attempts as a *sequential network visualization* (Hurtig et al., 2022). While our earlier studies (Lee et al., 2022b; 2022c) explored the partial pathways of student problem-solving, this study investigated the entire problem-solving pathways in the gamified learning application. As noted by Gobert et al. (2013), clickstream data collected in gamified learning applications indeed provided us with comprehensive information about students' mathematical problem-solving processes to build a visualization tool. The developed tool, MathFlowLens, successfully identified four distinct classifications for students' problem-solving strategies: optimal, sub-optimal, dead-end, and incomplete. Our results corroborate the findings of other research (Hurtig et al., 2022; Sinha & Aleven, 2015; Wang et al., 2017; Xia et al., 2021), which showed visualizing student problem-solving pathways may help researchers and teachers quickly identify students' understanding (i.e., best/optimal pathways), where they are stuck (i.e., incomplete pathways), as well the as the strategies they employ.

Next, in the development Phase 2 of MathFlowLens, we proposed an interactive graph visualization that showcases the variability in students' problem-solving strategies and provides an overall summary of a sample's preferred solution pathways. Additionally, the classifications from Phase 1 were used to color code the nodes and transformations in the Network diagram. The descriptive statistics of the pathway classifications indicated that there was variability in students' problem-solving pathways; the students exhibited a high average frequency of not only optimal pathways but also incomplete pathways. This suggests that this gamified learning environment can help students explore multiple possible problem-solving trajectories, not just reach a single solution (Clark et al., 2016; Jere-Folotiya et al., 2014; Wouters et al., 2013). In addition, unlike other visualization tools such as "PathViewer" (Wang et al., 2017) and "QLens" (Xia et al., 2021) mainly used Sankey Diagrams to represent student problem-solving paths, MathFlowLens uses network diagrams to visualize student pathways. While Sankey diagrams typically depict the transformations and flow in one direction, we were able to illustrate *multidirectional pathways* within problem-solving, for example, how the students returned to the previous states, corrected errors, and took other problem-solving pathways, as demonstrated in Hurtig et al.'s study (2022). These approaches, documenting and visualizing students' problem-solving pathways and the variation in strategy, may offer valuable insight into students' mathematical thinking and knowledge, as well as students' common misconceptions and gaps in their knowledge.

Practical implications for researchers and developers

Researchers can use MathFlowLens to classify students' problem attempts to further explore the complex nature of algebraic learning and student's problem-solving strategies. Researchers can also utilize the proposed visualization to gain additional insight into students' mathematical thinking and problem-solving strategies to inform new methodologies and prompt new research questions.

This study also has implications for developers of gamified applications and tutoring platforms. They may choose to implement the presented tool directly into their existing infrastructure, allowing for real-time attempt classifications and visualizations, along with the automatic generation of progress reports to further help and inform teachers.

Practical implications for teachers

This work also has direct implications for teachers in several ways. First, the developed tool is capable of automatically classifying students' problem-solving strategies and pathways which are often imperceptible to teachers (Asquith et al., 2007). Having a better understanding of what strategies a student uses to solve a problem and the variation in the employed solution pathways may provide teachers with additional insight into the student's current algebraic knowledge. Next, teachers can utilize the presented graph visualization to quickly identify a cohort's preferred problem-solving pathways. The interactive features of this visualization allow teachers to explore common misconceptions and errors and discover multiple equally efficient pathways. As such, the proposed visualization can be used to inform and improve teachers' instruction. Finally, teachers can use the identified pathways to actively compare problem-solving strategies in the classroom. This side-by-side presentation of solution strategies coupled with subsequent student discussion to highlight the similarities and differences among procedural pathways has been shown to improve students' procedural flexibility (Star et al., 2015).

Limitations and future directions

Several important limitations need to be considered. The presented tool that classifies students' problem attempts requires highly granular data containing step-by-step student transformations. However, a majority of educational technologies record students' entered answers rather than moment-by-moment mathematical transformations and expressions that show how students arrive at their solutions (Lee et al., 2022c).

While the proposed visualization presents a simplified view of a cohort's problem-solving strategies and pathways, its effectiveness and usefulness may depend on the user's ability to interpret and interact with it. We partially address this by making the visualization interactive, such as: allowing users to only show optimal pathways in the visualization. However, more work, such as usability testing for teachers or researchers, needs to be done to enhance the interpretability and usability of the visualization.

Overall, we view this work as an initial step towards classifying and understanding students' problem-solving strategies and the variation in their employed pathways. Possible extensions of this study include analyzing the sequence of problem-solving strategies rather than their frequencies, creating an automatic productivity coder for students' first steps using the proposed tool, and creating effective dashboards for teachers using the visualization and insights from the tool.

Conclusion

In this paper, we presented a classification and visualization tool, called MathFlowLens, that uses raw log data from FH2T to analyze students' procedural pathways. The tool classifies the *entire* student attempt as either optimal, sub-optimal, dead-end, or incomplete. Subsequently, MathFlowLens represents students' procedural pathways as a sequential network visualization. The findings of the current study demonstrate that MathFlowLens can provide valuable information regarding the employed solution strategies and students' mathematical thinking from raw log data. Additionally, the proposed visualization can be used by researchers and teachers to quickly identify the variation in a cohort's problem-solving pathways.

Study 2. Application of MathFlowLens: Examining the relations between identified pathways and the facets of algebraic knowledge.

* *This study is part of a paper currently under review in the Journal of Educational Technology Research and Development. Citation:* Pradhan, S., Ottmar, E., Gurung, A. & Lee, J. (submitted). MathFlowLens: A Classification And Visualization Tool for Analyzing Students' Procedural Pathways *Education Tech Research Development* (submitted).

Specifically, this study was incorporated as a component in the submitted ETRD manuscript as a potential application example, primarily, to showcase the predictive ability of the classifications.

Introduction

In the context of educational data mining and learning analytics, machine learning has been used to make various long-term predictions about student learning and future performance. For example, Chui et al. (2020) used data collected from a virtual learning environment and demographic information to predict marginal or at-risk college students. Similarly, Hodges & Mohan (2019) harnessed neural networks to predict gifted students with high accuracy. Such classifiers can help answer questions such as "is this an appropriate differentiation strategy" or "is this child showing need for additional support" (Hodges & Mohan, 2019). This predictive ability allows educators and teachers to intervene when necessary to facilitate student success.

With the development of the MathFlowLens tool complete, we leveraged the classifications generated from the tool to test its applicability in data-driven learning analytics research. In this short exploratory study, we apply the classifications in Study 1 by examining correlations between the identified classifications and students' algebraic knowledge. In other words, we use the classification data to predict the different aspects of students' algebraic knowledge. As mentioned above, algebraic knowledge consists of conceptual knowledge, procedural knowledge, and procedural flexibility.

Research Question

How do the different types of identified problem-solving pathways correlate with the various facets of algebraic knowledge?

Background

Facets of Algebraic Learning: Procedural and Conceptual Knowledge and Flexibility.

The mastery of algebra is essential for learning further advanced topics in mathematics (NMAP, 2008). However, many middle school and high school students struggle with basic algebraic concepts such as determining valid transformations and decomposing numbers

(NMAP, 2008), as well as converting simple algebra story problems to mathematically equivalent equations (Koedinger & Nathan, 2004). Students who struggle with these concepts may have a difficult time learning more advanced topics, most of which are usually represented in algebraic form.

For Algebraic proficiency, students require both *conceptual* and *procedural knowledge* and the ability to use this knowledge efficiently and *flexibly* (Schneider et al., 2011). Rittle-Johnson et al. (2001) defines *conceptual knowledge* as students' verbal and nonverbal knowledge of algebraic concepts and principles, including familiarity with algebraic symbols and syntactic conventions. Next, they define *procedural knowledge* as the understanding of the rules and procedures for solving an algebraic problem (e.g. the order of steps or transformations to solve a problem). While conceptual knowledge builds the foundation for procedural knowledge, students can use procedural practice to develop procedural knowledge as well (Rittle-Johnson et al., 2001, 2015; Schoenfeld, 2007). The combination of these two types of knowledge contributes to procedural flexibility–the ability to select the most efficient and effective solution for a particular problem (Star et al., 2016).

In a systematic review of procedural flexibility, Hong et al. (2023) highlights the emphasis in both research and practice on procedural flexibility as a learning outcome in mathematics. Across the globe, including Australia, China, Singapore, and the United States, educational ministries and committees have highlighted flexibility as an educational goal in mathematics learning (Australian Education Systems Officials Committee, 2006; Hästö et al., 2019; Hong et al., 2023; Ministry of Education of Singapore, 2006; Ministry of Education of the People's Republic of China, 2022; National Council of Teachers of Mathematics, 2014). In addition, there has been plenty of evidence demonstrating the importance of procedural flexibility. Blöte et al. (2001) and Rittle-Johnson and Star (2007, 2009) showed that procedural flexibility facilitates solving unfamiliar problems. Similarly, Rittle-Johnson et al. (2012) found that it increases conceptual and procedural knowledge in mathematics, the effect of which may spill over to other STEM domains such as physics and chemistry (Hästö et al., 2019). Various studies (Robinson et al., 2006; Venkat et al., 2019) suggest that strategic efficiency and flexibility are significant indicators of a student's understanding of the inherent mathematical structures.

Despite the widespread acknowledgment of the importance of mathematical flexibility, there are few pedagogical recommendations for improving students' flexibility and most of them tend to focus mainly on the use of a few strategies, rather than comparing multiple possible solutions (Verschaffel et al., 2009). Furthermore, to apply procedural flexibility in different situations, students need to develop a well-connected mental representation of the core concepts in algebra, such as equivalence, order of operation, and use of parentheses (Knuth et al. 2006; Ottmar et al. 2012; Welder 2012), allowing them to identify when, how, and which strategies are effective and applicable (Baroody, 2003; Threlfall, 2009; Verschaffel et al., 2009). In an effort to assist students in comparing different solution strategies and pathways, we created a tool that can identify various problem-solving pathways and strategies that students employ while problem-solving.

Method

In the RCT, students took a pre-assessment of their algebraic knowledge (hereafter, pretest scores) prior to the start of the intervention, and a post-assessment of their algebraic knowledge (hereafter, posttest scores) after the intervention. Out of the 1,649 students assigned to the FH2T condition, 1,139 completed the pre-test and only 778 completed both the pre- and post-test. For this analysis, we omitted students who did not complete both the pre-and post-test, resulting in an analytical sample of 778 students. For the pretest, students solved each item sequentially without feedback. Ten multiple-choice items were adapted from a previously validated measure of algebraic understanding (Star et al., 2015; Cronbach's $\alpha = 0.89$), consisting of three sub-constructs: conceptual knowledge (4 items), procedural knowledge (3 items), and procedural flexibility (3 items; see Appendix A). The post-test items replicated the pretest items but used different equations in the questions and choices. The average pretest score was 4.71 ($SD = 2.68$), and the posttest score was 4.50 ($SD = 2.93$) out of 10. Additional details about the dataset can be found in Ottmar et al. (2023).

The outcome variables (i.e., conceptual, procedural, and flexibility scores) were at the student level, hence, we aggregated the one-hot-encoded attempt-level classifications (i.e., independent variables) to the student level. First, for each problem, we summed the occurrences of each attempt type (i.e. total number of optimal, sub-optimal, dead-end, and incomplete attempts for a problem), resulting in one row for each problem for each student. Next, for each student, we took the average across problems attempted. Thus the columns in the data represent the average number of attempts belonging to the respective attempt type. The descriptive statistics of the outcome variables and the classifications have been given in Table 2. The most frequent pathways that the students took were optimal, followed by incomplete, sub-optimal, and dead-end pathways.

To identify what types of pathways (i.e. the identified attempt classifications) led to better conceptual knowledge, procedural knowledge, and procedural flexibility, we estimated three linear models. Model 1 predicted the conceptual knowledge scores (i.e. the number of conceptual items a student got correct in the posttest; see Appendix A) based on the identified attempt classifications while accounting for pretest scores (mean-centered). Similarly, Model 2 predicted procedural knowledge scores, and Model 3 predicted procedural flexibility scores after controlling for pretest scores.

Table 2. Descriptive Statistics for the student-level variables included in the models

Variable	Min	Median	Mean	Max	SD
Independent variables					
Optimal	0.40	0.80	0.84	4.36	0.25
Sub-optimal	0.00	0.33	0.35	2.09	0.13
Dead-end	0.00	0.11	0.13	2.14	0.12
Incomplete	0.00	0.62	0.70	6.79	0.44
Dependent variables					
Conceptual Score	0.00	2.00	1.89	4.00	1.37
Procedural Score	0.00	1.00	1.42	3.00	1.04
Flexibility Score	0.00	1.00	1.19	3.00	1.01

Results

As shown in Table 3, Model 1 explained 39.3% of the variation in conceptual scores ($F(5, 772) = 99.96$, $p < .0001$). Similarly, Model 2 accounted for 30.6% of the variation in procedural scores ($F(5, 772) = 68.22$, $p < .0001$), and Model 3 explained 30.8% of the variation in flexibility scores ($F(5, 772) = 68.64$, $p < .0001$). Surprisingly, students taking dead-end pathways were positively related to conceptual score ($B = 0.84$, $p = .031$) and procedural score ($B = 1.01$, $p = .001$), after accounting for pretest scores. These results were surprisingly counter-intuitive, as we originally hypothesized that dead-ends indicate poor procedural and conceptual knowledge and would consequently lead to lower procedural and conceptual scores. Conversely, students taking optimal paths was negatively related to flexibility score ($B= -0.58$, $p = .021$), after controlling for pretest scores. Neither incomplete nor sub-optimal attempts were significant predictors of conceptual, procedural, or flexibility scores. The results also suggest that pretest scores, the control variable, was a significant predictor of conceptual ($B = 0.31$, $p < .001$), procedural ($B = 0.20$, $p < .001$), and flexibility scores ($B = 0.20$, $p < .001$).

Table 3. Regression results of predicting conceptual, procedural, and flexibility scores from the identified attempt classifications while accounting for prior algebraic knowledge.

(*N* = 778)	**Conceptual Score**			**Procedural Score**			**Flexibility Score**		
Predictors	*B*	*CI*	*p*	*B*	*CI*	*p*	*B*	*CI*	*p*
(Intercept)	2.02	1.73, 2.31	**<0.001**	1.43	1.19, 1.66	**<0.001**	1.28	1.05, 1.51	**<0.001**
Optimal	-0.28	-0.62, 0.05	0.099	-0.17	-0.44, 0.10	0.216	-0.31	-0.58, -0.05	**0.021**
Dead-end	0.84	0.08, 1.60	**0.031**	1.01	0.39, 1.63	**0.001**	0.43	-0.27, 1.03	0.161
Incomplete	-0.17	-0.39, 0.05	0.133	-0.11	-0.29, 0.07	0.234	-0.09	-0.27, 0.09	0.315
Sub-optimal	0.36	-0.33, 1.05	0.311	0.23	-0.33, 0.79	0.414	0.52	-0.02, 1.06	0.060
Pre Total Math Score	0.31	0.28, 0.34	**<0.001**	0.20	0.18, 0.23	**<0.001**	0.20	0.18, 0.23	**<0.001**
R^2	0.393			0.306			0.308		
R^2 Adjusted	0.389			0.302			0.303		

Discussion

As a potential research application of MathFlowLens, we examined the relations between the average frequency of students' solution strategies across problems with the three constructs of math performance: conceptual knowledge, procedural knowledge, and flexibility scores. The results indicated that students with frequent dead-end attempts had higher conceptual and procedural knowledge scores, whereas frequent optimal attempts led to lower flexibility scores.

Dead-ending behavior may represent a form of exploratory play, in which students try different solution strategies in a low-stakes gamified environment. This exploratory behavior may help students identify unproductive transformations to avoid, potentially leading to a better understanding of algebraic knowledge (Pradhan et al., 2024a). This finding is aligned with that productive failure through exploratory behavior may facilitate learning gains by emphasizing opportunities to explore the constraints of a problem space (i.e., boundary testing), and testing multiple solution pathways (Owen et al., 2016).

On the other hand, as indicated by the results, focusing solely on efficient solutions may lead to a decrease in procedural flexibility. A plausible explanation for this negative correlation is that forcing students to only take a particular and efficient pathway may limit their mathematical flexibility and/or the likelihood of learning from mistakes (Francome & Hewitt, 2020).

Furthermore, based on the findings from the linear models, developers of learning platforms should pay close attention to platform design to ensure exploratory behaviors, such as frequent dead-end attempts, are not penalized. Such exploratory or creative attempts may be less efficient but may provide students with a more nuanced understanding of mathematical structures and operations. By acknowledging the variability in students' strategies (e.g. creative, exploratory, and efficient), developers can consider the individual differences in students and their behaviors to create more adaptive and supportive learning environments.

Limitations and Future Directions

The analysis using the linear models focused on data derived from one specific context, FH2T. It is unknown whether these results are generalizable across different platforms. To strengthen the findings from this study, it would be beneficial for future work to replicate and test the generalizability of these results. Additionally, future studies could run alternative linear models with the inclusion of students' prior individual component scores (i.e. prior conceptual, prior procedural, and prior flexibility scores) rather than prior algebraic performance. Doing so would allow researchers to examine the correlation between the frequency of a type of procedural pathway and the associated increase in score for the individual components of algebraic knowledge.

Conclusion

In this initial exploratory study, we surprisingly found that students who take dead-end pathways more frequently have higher conceptual and procedural scores. This result highlights the need to foster exploratory behavior and creativity to bolster algebraic knowledge. Finally, students frequently taking optimal pathways have lower procedural flexibility, which underscores the importance of teaching multiple solution pathways and strategies.

Study 3. Gamification and Dead-ending: Unpacking Performance Impacts in Algebraic Learning.

** This study was presented and is published in the proceedings as a paper at the International Learning Analytics and Knowledge Conference (LAK 2024, Kyoto, Japan). A few edits have been made to improve the readability and flow of the thesis.*

Citation: Pradhan, S., Gurung, A., Ottmar, E. (2024, March). Gamification and Dead-ending: Unpacking Performance Impacts in Algebraic Learning. In *LAK24: 14th International Learning Analytics and Knowledge Conference*.

Introduction

Despite the widespread acknowledgment of the correlation between strong algebraic knowledge and enhanced performance in future advanced topics, a disconcerting number of middle school and high school students struggle with fundamental algebraic concepts. Difficulties making such as valid transformations and decomposition (NMAP, 2008), and challenges in converting simple story problems into mathematically equivalent equations (Koedinger & Nathan, 2004), are indicative of the potential struggles these students might face as they encounter more advanced topics typically expressed in algebraic form.

Solving algebraic problems requires students to utilize a broad spectrum of problem-solving techniques. These techniques enhance students' ability to synthesize solutions, shape their mathematical intuition, and reinforce their methodological approaches to problem-solving. As students' progress to more advanced mathematical domains, mastering these foundational strategies becomes paramount. Indeed, proficiency in algebraic concepts is intimately linked with the acquisition of a wide array of problem-solving techniques (NMAP, 2008). Especially in K-12 mathematics education, efficiency and flexibility in problem-solving strategies are prioritized (Common Core, 2010), and efficient students often employ fewer steps or transformations (Xu et al., 2017). This is supported by various studies that suggest that strategic efficiency is a significant indicator of a student's understanding of the inherent mathematical structures (Robinson et al., 2006; Venkat et al., 2019).

However, the surprising results from Study 2 suggest that relying solely on efficient problem-solving strategies and pathways may have a negative effect on students' procedural flexibility. Ultimately, this may limit a student's ability to generate alternative procedural pathways for unseen and new problems that may be more intuitive and transferable to more advanced mathematical topics.

Additionally, the underlying mechanisms that prompt students to discontinue their current procedural approach, leading to dead-end and incomplete pathways, remain unclear. However, various factors, both positive and negative, can sway a learner's decision to discontinue. Positive triggers might include realizing that a path will only yield a sub-optimal outcome or foreseeing a challenging state ahead. Conversely, negative factors could include frustration from an inability to solve a problem or reaching a genuine impasse where the student is unable to identify the next state.

It is important to note that the results from Study 2 gave insight into the relation between the classifications and the individual constructs of algebra. It is unknown whether the choice of procedural pathways lead to differentiated algebraic learning outcomes as a whole. To further investigate the surprising results in Study 2 and unpack the algebraic performance impacts of the identified classification, this paper aims to explore the implications of encountering dead-ends within the network of strategic pathways. Accordingly, we explore the following research questions:

Research Questions

RQ1: Does the choice of procedural pathways in algebraic problem-solving lead to differentiated learning outcomes?

RQ2: In what ways do dead-end attempts within a gamified environment impact algebraic learning?

Data

The data and sample in this study were the same as Study 1 and 2, i.e. the FH2T raw log data from the RCT. Specifically, we used the classification data generated from the MathFlowLens tool for the analysis.

Results

RQ 1: Exploring what types of attempts lead to better learning outcomes.

To address RQ1 and identify problem-solving strategies that led to better learning outcomes, we estimate two linear models. Model 1 predicts the post-test scores of students based on the identified pathways or classifications, while the second model accounts for prior algebraic knowledge (mean-centered) in addition to the classifications. Table 4 contains the results of running the linear models. The results of model one suggest that at the student level, neither classification of optimal (β = -0.38, p = 0.407) nor sub-optimal (β = 1.46, p = 0.118) was a significant predictor of post-test scores. In contrast, the classification of incomplete (β = -1.69, p <0.001) and classification of dead-end (β = 5.17, p <0.001) were significant predictors of post-test scores. In model two, we observed that higher prior knowledge was correlated with higher post-test performance (β = 0.72, p <0.001). The optimal path (β = -0.76, p = 0.028) was also a significant predictor of post-test scores. Additionally, while the effect decreased, dead-end (β = 2.28, p = 0.004) was still a significant predictor of higher post-test performance.

Overall, the results of these models suggest that after accounting for prior algebraic knowledge, the average student exhibiting dead-ending behavior is more likely to succeed. On the other hand, students who exhibit efficient problem-solving behavior tend to perform worse on the post-test. These findings strengthen and align with our findings from Study 2. The positive correlation between dead-ending behavior and student post-test performance indicates the likelihood that the underlying mechanism that results in dead-ending behavior is likely positive in nature. Such learners are able to identify that the path will only yield a sub-optimal path or foresee a challenging state ahead. As such, we posit that a dead-end attempt may, in fact, be an indicator of 'exploratory play', an in-game behavior that potentially leads to a more nuanced understanding of the transformations to avoid or the ability to identify problematic states when solving algebraic problems to reach the total state. Consequently, resulting in a better algebraic post-test performance.

Table 4. Student-level linear regression results predicting post-test score.

	Model 1 post test math score			Model 2 post test math score		
Predictors	*Estimates*	*CI (α=0.05)*	*p*	*Estimates*	*CI (α=0.05)*	*p*
(Intercept)	4.79	[4.02, 5.56]	**<0.001**	4.70	[4.12, 5.29]	**<0.001**
Class Best	-0.38	[-1.27, 0.52]	0.407	-0.76	[-1.44, -0.08]	**0.028**
Class Incomplete	-1.69	[-2.27, -1.11]	**<0.001**	-0.35	[-0.80, 0.10]	0.129
Class Sub-optimal	1.46	[-0.37, 3.29]	0.118	1.08	[-0.31, 2.46]	0.129
Class Dead-end	5.17	[3.17, 7.16]	**<0.001**	2.28	[0.75, 3.81]	**0.004**
Pre Total Math Score				0.72	[0.66, 0.77]	**<0.001**
Observations	774			774		
R^2/R^2 *adjusted*	0.057/0.052			0.459/0.456		

RQ 2: Exploring the effect of regular dead-ending on algebraic learning outcomes.

Building on the results of the exploratory analysis in Study 2 and RQ1, we further investigated the relationship between dead-ending (or exploratory behavior) and higher post-test scores. We examined potential variance in the dead-end states across students by constructing individual networks per student per problem. The data for this analysis was generated using the binning classification feature of the MathFlowLens tool. Student-level networks were generated to identify dead-end paths of students that were potentially masked by their peers' attempts. For example, if a student had an exploratory attempt ('start state' → 'a'), and another student used the same path to reach the goal state sub-optimally ('start state' → 'a' → 'b' → 'goal state'), the student's exploratory attempt would be masked and classified as incomplete. By identifying dead-end paths on student-level networks, we localize the definition of dead-end paths to individual students' attempts. It is important to note that this modification does not change the classification for optimal or sub-optimal attempts, as the optimal paths found from the entire sample are used for this classification.

Table 5. Summary Statistics of Dead-end Count and Percentage

Statistic	Mean	St. Dev.	5%	Median	95%
Dead-end Count	27.1	16.1	4	24.5	56
Dead-end Percentage	23.4	7.5	12.05	23.2	35.5

Next, we examined the frequency of dead-ending behavior per student by examining the total number of problems in which the student had at least one dead-end attempt. Similarly, we calculate the percentage of problems with at least one dead-end attempt. These results can be found in Table 5. Since the dead-end count and percentages were not normally distributed, and certain students were regularly utilizing the dead-end pathways in comparison to their peers, we classified the students into 'regular dead-enders' and 'occasional dead-enders' by utilizing a cutoff point at the 5th percentile of the dead-ending behavior distribution. We ran a mixed-effects model at the attempt level, predicting post-test scores while accounting for prior knowledge (mean-centered), using the student-level network classification and an indicator for the students' regular usage of dead-end paths. As the data is at the attempt level, we introduce random intercepts for the problem ID, attempt number, and pre-test scores.

Table 6. Exploring the correlation between different types of procedural pathways taken by individual students and their post-test performance.

	post test math score		
Predictors	*Estimates*	*CI (α=0.05)*	*p*
(Intercept)	4.90	[4.50, 5.29]	**<0.001**
Attempt Best	0.34	[0.31, 0.38]	**<0.001**
Attempt Deadend	0.05	[0.01, 0.08]	**0.006**
Attempt Sub-optimal	0.07	[0.04, 0.11]	**<0.001**
Pre Total Math Score	0.65	[0.53, 0.76]	**<0.001**
Regular Deadending	0.24	[0.18, 0.31]	**<0.001**
Random Effects			
σ^2	4.07		
$\tau_{00\,problem\ id}$	0.54		
$\tau_{00\,attempt\ number}$	0.08		
$\tau_{00\,pre\ total\ math\ score}$	0.38		
ICC	0.20		
$N_{attempt\ number}$	94		
$N_{pre\ total\ math\ score}$	11		
$N_{problem\ id}$	252		
Observations	179575		
Marginal R^2/Conditional R^2	0.391/0.512		

Table 6 suggests that for a student with an average score on the pretest, the use of optimal or best paths correlates significantly with higher scores on the post-test ($\beta = 0.33$, $p < 0.001$), especially when compared to the reference category of incomplete paths. This trend is also seen with sub-optimal paths ($\beta = 0.07$, $p < 0.001$) and dead-end paths ($\beta = 0.05$, $p = 0.006$), both showing a positive correlation with the students' post-test scores. Similar to the results of RQ1, the pre-test score remains a significant predictor of the post-test scores ($\beta = 0.65$, $p < 0.001$). Interestingly, students who regularly adopt dead-ending strategies in their problem-solving tend to perform better than those who use such strategies less frequently ($\beta = 0.24$, $p < 0.001$).

Discussion

The findings of this study have two major implications. Firstly, the positive effect of gamified systems on algebraic learning outcomes depends on the behaviors exhibited by the student. Past studies such as Chan et al. (2022b), Lee et al. (2022a), and Vanacore et al. (2023), have shown that different in-game behaviors are predictive of algebraic learning outcomes. In particular, the studies by Chan et al. (2022b) and Lee et al. (2022a) showed that students who paused before answering tend to perform better in the post-test. Similarly, Vanacore et al. (2023) showed that students with a higher propensity for persistence benefit more from the gamified system. In the current study, using log data, we provided additional evidence suggesting that the effect of gamified platforms on learning outcomes depends on the behaviors and intentions of the user.

The second major implication is that in-game behaviors exhibited by students may be the main driving force behind improved algebraic knowledge in gamified systems. Desirable behaviors, such as the exploratory behavior identified in this study, should not be penalized. If the results presented in this study are consistent for similar gamified systems, there are profound impacts on the design of gamified platforms to foster exploration. Additionally, our results suggest that in order to develop math flexibility, students may need to explore various procedural pathways. In the long run, this may allow students to develop the important skill of choosing efficient problem-solving strategies. Overall, the results from this study support, and further explain the strange results from Study 2.

Limitations and Future Directions

In considering the outcomes of this study, several important caveats should be acknowledged. To begin with, our analysis was narrowly focused on data derived from the FH2T platform. This specificity introduces potential limitations on the generalizability of the results. There remains an open question about the replicability of the observed student behaviors and interactions across a wider range of platforms that employ similar dynamic procedural pathways. To strengthen the findings of this study, it would be instructive for subsequent investigations to explore the generalizability of our findings further. Additionally, the insights extrapolated here might be more germane to gamified environments rather than to traditional tutoring platforms such as the Cognitive Tutor (Anderson et al., 1995) and ASSISTments (Heffernan & Heffernan, 2014) mentioned earlier. These platforms, with varying affordances regarding procedural requirements, might influence student behavior differently, possibly reducing the propensity for the kind of exploratory action observed in our analysis.

While this study aimed to identify and understand the implications of various procedural pathways in solving algebraic problems within a gamified setting, the broader implications of these classifications must be acknowledged. Future research should investigate the effects of hints on the paths and explore variations in their effective utilization. Prior research has underscored the value of using response times as a metric to infer productive hint usage (Gurung et al., 2021) and the formulation of optimal solutions (Chan et al., 2022b). Additionally, several studies have highlighted the benefits of providing error-specific feedback to frequently occurring incorrect answers (Gurung et al., 2023a; 2023b). The models established in this research can greatly enhance our understanding of the mechanisms underlying the procedural pathways that lead to these common errors. Similarly, insights into these pathways can improve the quality of automated grading and feedback generation for student responses in open-ended algebraic problems (Baral et al., 2023a; 2023b) by helping mitigate potential biases (Gurung et al., 2022) by facilitating an objective understanding of the potential mechanisms influencing the students' responses.

It would also be of academic interest for subsequent studies to investigate the interplay between these classifications and various demographic or evaluative indicators, such as levels of math anxiety. Such a focus can illuminate nuanced patterns of interaction across heterogeneous student groups. By doing so, we can better inform and adapt educational strategies, aiming to enhance both the inclusivity and efficacy of gamified instructional methodologies.

Conclusion

In this study, we find that students who exhibit regular dead-ending behavior have a higher post-test score (i.e. higher learning outcome), than students who are irregular dead-enders. In other words, students exhibiting regular dead-ending behavior (i.e. exploratory), gain more from the gamified system. This suggests that students who display regular exploratory (i.e. dead-ending) behavior may be learning the various algebraic rules and notations in a low-stakes gamified environment, eventually leading to better algebraic understanding. Such students take pathways usually deemed incorrect or inefficient but overall, it may lead to a productive learning process.

Study 4. MathFlowLens Dashboard: Creating a collaborative dashboard with teachers, for teachers.

** This study is part of a larger NSF project (CAREER 2142984: PI Ottmar) which focuses on using Codesign methodology and professional development sessions with teachers to iteratively design and development MathFlowLens. The manuscript for that project is currently under preparation to be submitted later this summer. My contributions to that project were focused on the data processing, data analysis, prototype and UX design based on teachers input, and software engineering.*

Citation: Thompson, T., St. John, J., Pradhan, S., & Ottmar, E. (in preparation). *MathFlowLens Dashboard: Co-Designing Teacher Orchestration Tools to Engage in Discourse Around Students' Mathematical Strategies.*

Introduction

Previous studies revealed the significance of classifying, predicting, and valuing multiple solution pathways and showed how the identified classifications differentially predicted learning outcomes. As highlighted in the background, teachers need to be aware of their students’ performance, their exhibited behaviors (Walkoe et al., 2013), and employ pathways and understand that variation in pathways. As highlighted by the results in Studies 2 and 3, even incorrect or inefficient pathways are a productive process for learning. Knowing students’ procedural pathways, common misconceptions, and performance allows teachers to intervene when necessary, and create a personalized and supportive learning environment. If teachers were provided with information about students' variations in the problem-solving process, teachers modify their instruction and make them more aware of their student’s mathematical processes and knowledge in an online setting.

Teacher orchestration tools and dashboards can aid teachers in identifying and interpreting relevant information and providing impactful and actionable feedback. For example, the use of dashboards has the potential to improve teachers' understanding of students’ knowledge, perceptual and mathematical strategies, and misconceptions (Walkoe et al., 2013). However, this information about students' mathematical problem solving and processes is rarely given to teachers. Providing teachers with more information about students' strategies and behaviors when solving problems may enhance teachers’ pedagogical insights allowing them to modify their mathematical instruction accordingly (Holstein et al., 2019). While educators aspire to gain insights into students' thought processes and identify any misconceptions (Holstein et al., 2019), teachers tend to employ technology platforms primarily for assessing student performance rather than delving into a nuanced understanding of their comprehension. This is in part due to the lack of time for teachers to examine individual students’ problem attempts (Feldon, 2007; Holstein et al., 2017), and the lack of technological tools and dashboards to convey relevant metrics and visualizations. Further, simply presenting teachers with dashboards and data is not enough to sufficiently help teachers identify and interpret relevant information about students’ problem-solving processes. Teacher tools must prioritize design that helps teachers quickly find relevant information and accurately.

Tools such as teacher dashboards summarize student learning and performance data to help teachers monitor and be more aware of their students' progress, ultimately facilitating pedagogical decision making (Verbert et al., 2014). Yet, there are disparities between researchers' intent for the developed dashboards and their effectiveness and usability for teachers (Hopfenbeck, 2020; Schwarz et al., 2021). Thus, developing tools using a codesign methodology–by including teachers in the design and conceptualization of such tools–is paramount. By working with teachers to better understand their needs and the various challenges they face in the classroom, researchers may devise tools that are truly supportive and meaningful for teachers.

Overview of the Codesign Sessions: Understanding What Teachers Want.

Last summer, two Learning Science graduate students and I, along with Erin Ottmar (PI), led a 2-day professional development and co-design workshop with five middle and high school mathematics teachers. Through engaging with these teachers for several sessions, this project sought to iteratively design and develop a research-based technology dashboard that provided insights into students' mathematical problem solving, with extensive input from experienced educators. This project had 3 distinct phases: 1) familiarizing teachers with existing dashboards and online educational platforms, which allowed teachers to select or request key metrics and preferred visualizations; 2) iterative and collaborative proof-of-concept designing of tools with teachers for the various components; and 3) developing and programming the final dashboard.

In the first phase, prior to the first codesign session, we requested the participating teachers to get familiarized with an online learning platform: Graspable Math (GM; Weitnauer et al., 2016). Like FH2T–the gamified learning platform whose data was used for the conceptualization of the MathFlowLens tool in Study 1, GM logs moment-by-moment transactional data that records each step a student makes. This familiarization process was intended to give teachers a student's experience while solving problems in the digital platform and allow them to begin narrowing the insights they would like to gain from a teacher dashboard. Next, we held a conceptualization and brainstorming session in which we presented existing visualization techniques–including the developed sequential network visualization from Study 1–and commonly used student performance metrics. This session allowed us researchers to better understand what the teachers sought for: the likes and dislikes of the presented analytical and visual components; factors contributing to cognitive overload that may lead to unwillingness to use the tool; and the existing tools, both technological and pedagogical, that teachers used and would want included in a dashboard.

In the second phase, we consolidated the needs of and comments by the teachers in the first session to prepare prototypes of visualizations and key student performance metrics. Then we held a final session to present the prototypes and allow teachers to give additional feedback. In this session, we also allowed teachers to manipulate the layout, placement and inclusion of the prototypes, giving them complete control over the final appearance of the dashboard.

In the final phase, based on the feedback from teachers, and following the iterative co-design methodology, we used the prototypes and suggested layout designs to create a web-based dashboard that conveys key metrics, including the classifications of strategies, and overall student performance to teachers (link: mathflowlens.com). The following sections describe the final phase and the individual components in detail.

Development of the MathFlowLens Dashboard: Developing Tools that Support Teachers.

One of the main goals of this project was to create a dashboard that is teacher-centric, interactive, and provides actionable feedback regarding students' performance and solution strategies. This study also paves the way for future studies to conduct additional user-design research with teachers to better understand how these classifications and visualizations can be useful for their mathematics teaching and student learning.

Another principal goal for this project was to make the tool generalizable, thereby allowing external learning platforms to use and display the strategic information and visualizations on the developed dashboard. As such, we initially collaborated with and utilized data from GM to develop the proof-of-concept dashboard and successfully preprocess and display visualizations and metrics using their externally housed data. Additionally, we programmed this dashboard to be easily displayed as an iFrame to allow for a more robust integration with the external learning platform’s infrastructure.

The following sections describe the design process behind the initial prototype and final implementation of the individual components included in the current dashboard.

Network Component: Utilizing Graph Theory to Visualize and Classify Student Strategies.

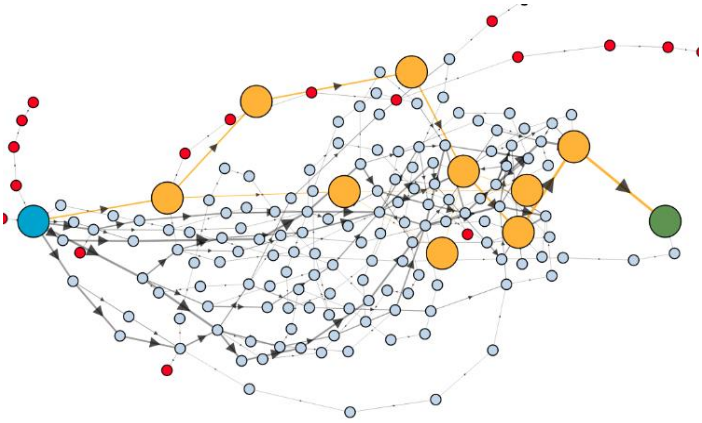
The first step in this development was to generate visualization and categorizations of students' strategies. To do so, we used the methodologies outlined in Pradhan et. al (2024a) and the “MathFlowLens” tool developed by Pradhan et. al (under review) to analyze, classify and visualize students’ problem-solving procedural pathways. By applying graph theory and shortest path finding algorithms on transactional log data, MathFlowLens classifies the procedural pathways as optimal, sub-optimal, dead-end, or incomplete. An optimal pathway occurs when students solve a problem in the least number of steps; conversely, a sub-optimal pathway occurs when they use more steps than needed, i.e. use an inefficient problem-solving strategy. While both dead-end and incomplete attempts are unfinished attempts, which indicates the student reset or restarted the problem, dead-end attempts are unique as there are no procedural pathways leading to the goal state. On the other hand, incomplete pathways can eventually be used to reach the goal state. Next, after classifying all student attempts in the raw data, MathFlowLens uses sequential network visualizations to represent the procedural pathways (see Figure 6a) using color coding based on the identified classifications. A summary of the nodes is given in Table 7.

Table 7. Summary of nodes on the network diagram

Node Color	Indicates	Description
Blue	Start State	The initial expression before students begin working.
Gold	Optimal Step	Efficient steps that lead to an optimal attempt.
Grey	Sub-optimal Step	Inefficient steps that lead to a sub-optimal attempt.
Red	Dead End Step	Nodes that lead to a unique, but incomplete path (never reaches the goal state).
Green	Goal State	The final expression that the student transforms the Start State into.

For the network diagram component of the dashboard, we used the same implementation as the MathFlowLens tool. Specifically, we used Python 3.11 and the pandas package for the preprocessing and classifications, and JavaScript and the d3.js library to render the visualizations. The initial prototype presented to the teachers was static (i.e. no interactivity; see Figure 6a), as the figure represented an overall view of the variation of pathways. However, after the initial presentation of this visualization to the teachers, both teachers and researchers agreed that adding interactivity to drill down on specific attempts and pathways would be beneficial (see Figure 6b).

(a) Initial Prototype



(b) Final Implementation (in fullscreen mode)

Figure 6. Initial and Final Design of the Sequential Network Visualization

Consequently, in the final implementation of the network diagram component, we added various interactive features. Users can use their mouse to hover over the nodes for additional information such as the node's classification and expression (see Figure 7a). Similarly, users can click on the 'Only Best Path' button to hide all nodes and pathways besides those with an optimal classification (see Figure 7b). Next, to allow teachers to isolate a specific student's progress and attempt, we implemented a feature that highlights the selected attempt in the sequential network diagram in purple and lists the individual steps taken by the selected student in a table (see Figure 7c). Additional interactivity features such as making the visualization fullscreen, and temporarily disabling interaction were also implemented.

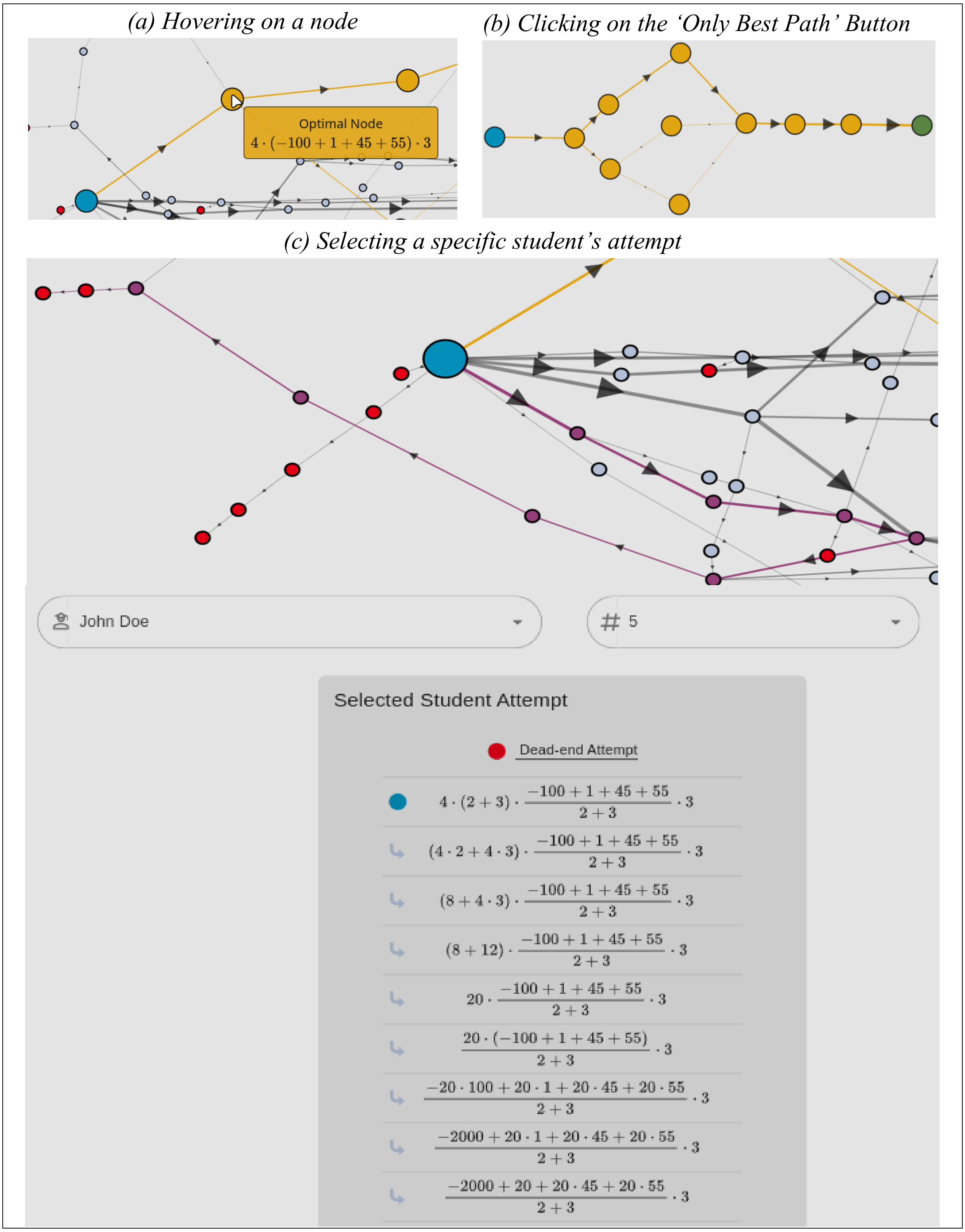


Figure 7. Interactive elements of the final network component

Treemap Component: Visualizing the frequency of the first step

The initial needs analysis revealed that teachers often prioritize students' first steps, primarily because it indicates the overall productivity of the attempt. As a result, we suggested using treemap diagrams (Johnson & Shneiderman, 1998) in the dashboard to display the various first steps taken by students. Treemap diagrams display nested and hierarchical structures as a set of rectangles, and the size of the rectangles is proportional to their value. In this case, each rectangle represents a first step taken by a student, and the size of the rectangle is determined by the number of attempts in which that first step was taken.

This treemap visualization was originally conceived to be separate from the MathFlowLens' classifications and sequential network diagram. In other words, us researchers viewed this component to be isolated from the other components. Hence, the colors of the nodes were based on its frequency. However, after presenting this version of the diagram to the teachers in the initial codesign session, the teachers suggested using the same colors (based on the classifications derived from the MathFlowLens tool: gold, gray, and red) for consistency and to promote additional discourse in the classroom.

After the first session, based on the feedback of teachers we updated the treemap component to use consistent color coding. In the second codesign session, the teachers suggested adding partial interactivity to improve usability. Besides adding interactivity, the initial prototype and final design of the treemap component was the same. We used Python 3.11 and the pandas package to preprocess the data and JavaScript and Plotly.js library to render the visualizations. Each rectangle displays the step's expression and the percentage of attempts that used that step. Additionally, colors derived from the MathFlowLens tool were used to indicate the step's classification (see Figure 8). This figure suggests that the two most common first steps (i.e. "4*5*(-100+1+45+55)/(2*3)*3" and "4*(2+3)*(99+45+55)/(2*3)*3") both lead to a sub-optimal attempt.

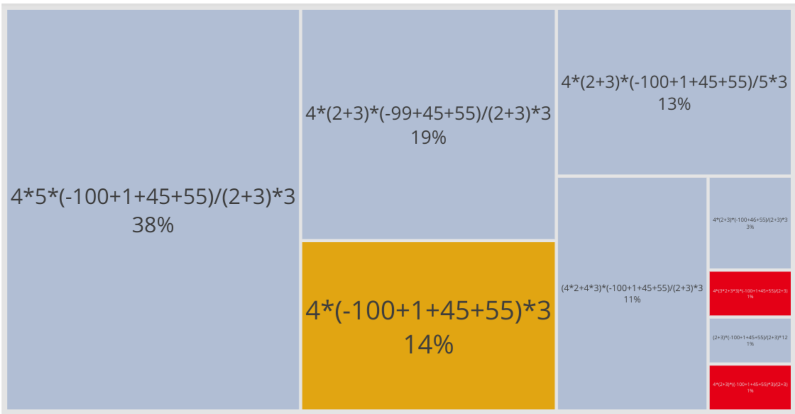


Figure 8. Initial and Final Design of the Treemap Component

Given that some first steps were used infrequently, the sizes of such rectangles would be small and hard to interpret. As a potential solution, we added an interactive hovering feature that displays the expression, percentage, and number of attempts (see Figure 9). Teachers may also choose to click the rectangle to expand it to make it easier to view.

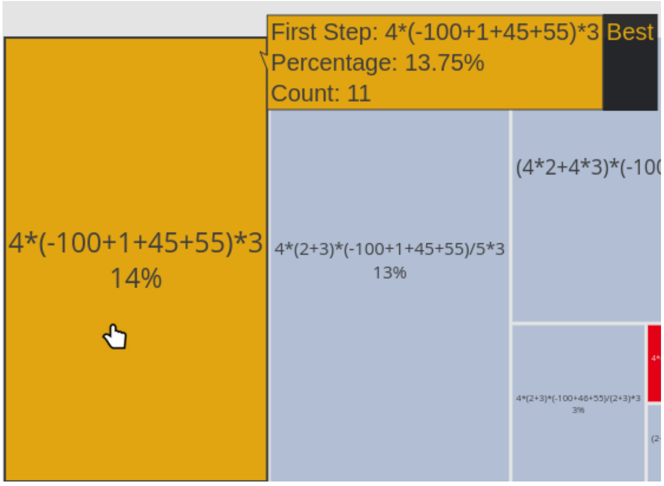


Figure 9. Interactive hover feature of the Treemap Component

Attempt Type Percentage Component

To provide a broad overview of the variation in the class' problem-solving strategies, we presented teachers with a table denoting the percentages of each of the classifications derived from the MathFlowLens tool (see Figure 10) to include in the dashboard. Based on the strong positive feedback of teachers suggesting that this would be especially useful for them in decision making, this was integrated into the Examples Tables component.

Completed Attempt Type	Percentage
Best	22.22 %
Sub Optimal	62.22 %
Dead End	15.56 %

Figure 10. Initial Design for Attempt Type Percentage Component

Examples Table Component

The initial needs analysis revealed that teachers wanted to view and compare different types of student problem-solving strategies and attempts. We suggested and designed an Example Table component that would display different attempts chosen at random for each attempt classification.

The initial design consisted of a table with step numbers, and a column for each attempt classification: optimal, sub-optimal, and dead-end (see Figure 11a). Based on the feedback from teachers, we integrated the attempt type percentage information in the table header and added color coding highlighting the classification of the step in the final design (see Figure 11b). The color coding would help teachers identify common misconceptions in the case of dead-end attempts and steps that lead to inefficiencies in sub-optimal attempts.

(a) Initial Prototype

Step Number	Best	Sub Optimal	Dead End
1	4*(2+3)*(-100+1+45+55)/(2+3)*3	4*(2+3)*(-100+1+45+55)/(2+3)*3	4*(2+3)*(-100+1+45+55)/(2+3)*3
2	4*(-100+1+45+55)*3	4*5*(-100+1+45+55)/(2+3)*3	4*5*(-100+1+45+55)/(2+3)*3
3	4*(-100+1+100)*3	4*(5*(-100+1+45+55))/(2+3)*3	20*(-100+1+45+55)/(2+3)*3
4	4*(-99+100)*3	(4*5*(-100+1+45+55))/(2+3)*3	20*(-99+45+55)/(2+3)*3
5	4*1*3	(20*(-100+1+45+55))/(2+3)*3	20*(-54+55)/(2+3)*3
6	4*3	(20*(-99+45+55))/(2+3)*3	20*1/(2+3)*3
7	12	(20*(-54+55))/(2+3)*3	20*1/5*3
8	6+6	(20*1)/(2+3)*3	60*1/5
9		20/(2+3)*3	12*5*1/5
10		20/5*3	
11		4*3	
12		12	
13		6+6	

(b) Final Design

Step Number	Optimal (22.73%)	Suboptimal (61.36%)	Deadend (15.91%)
Start State.	$4\cdot(2+3)\cdot\frac{-100+1+45+55}{2+3}\cdot 3$	$4\cdot(2+3)\cdot\frac{-100+1+45+55}{2+3}\cdot 3$	$4\cdot(2+3)\cdot\frac{-100+1+45+55}{2+3}\cdot 3$
1.	$4\cdot(-100+1+45+55)\cdot 3$	$4\cdot 5\cdot\frac{-100+1+45+55}{2+3}\cdot 3$	$4\cdot(3\cdot 2+3\cdot 3)\cdot\frac{-100+1+45+55}{2+3}$
2.	$4\cdot(-100+1+100)\cdot 3$	$20\cdot\frac{-100+1+45+55}{2+3}\cdot 3$	$4\cdot(6+3\cdot 3)\cdot\frac{-100+1+45+55}{2+3}$
3.	$4\cdot(-99+100)\cdot 3$	$20\cdot\frac{-99+45+55}{2+3}\cdot 3$	$4\cdot(6+9)\cdot\frac{-100+1+45+55}{2+3}$
4.	$4\cdot 1\cdot 3$	$20\cdot\frac{-54+55}{2+3}\cdot 3$	$4\cdot 15\cdot\frac{-100+1+45+55}{2+3}$
5.	$4\cdot 3$	$20\cdot\frac{1}{2+3}\cdot 3$	
6.	12	$20\cdot\frac{1}{5}\cdot 3$	
7.	$6+6$	$20\cdot\frac{1\cdot 3}{5}$	
8.		$20\cdot\frac{3}{5}$	
9.		$\frac{20\cdot 3}{5}$	
10.		$\frac{60}{5}$	
11.		12	
12.		$6+6$	

Figure 11. Initial and Final Design of the Examples Table Component

Performance Metrics Table Component

As mentioned above, teachers emphasized the need to see growth over time in key performance metrics such as the number of steps, time taken to solve a problem, and total number of errors. Our team decided to present key student performance metrics indicating the overall class performance. This could be used by teachers to compare students' performance on similar problems at different points in time (e.g. start and end of the year). While the contents of the table did not change across the initial and final design (see Figure 12), we added an interactive hover feature that provides a short description of the statistic. All measures included in this component and their respective descriptions have been provided in Appendix B.

(a) Initial Prototype

Measure	Class Mean
Completions	93.94%
Number of Attempts	2.36
Number of Hints	0.55
Number of Resets	2.18
Number of steps in First Attempt	10.87
Number of steps in Last Attempt	11.06
Efficiency First Attempt	0.68
Efficiency Last Attempt	0.67
Time Taken (sec)	79.07
Pause Time First Attempt (sec)	11.47
Pause Time Last Attempt (sec)	11.22
Number of Total Errors	4.7
Number of Keypad Errors	0.03
Number of Shaking Errors	3.29
Number of Snapping Errors	1.19

(b) Final Design

Measure	Average
Completions	93.94%
Average number of hints requested by students	2.36
Number of Hints	0.55
Number of Resets	2.18
Number of steps in First Attempt	10.87
Number of steps in Last Attempt	11.06
Efficiency First Attempt	0.68

Figure 12. Initial and Final Design of the Sample Means Component

Discussion

The iterative codesign session allowed us to design a teacher dashboard that was teacher centric and truly supportive for the needs of teachers. Their input was paramount to the initial conceptualization, the iterative improvement of individual components and the placement of these components in the layout of the web-based dashboard. The final MathFlowLens Dashboard (accessible through: mathflowlens.com) uses the Vue.js and Vuetify JavaScript libraries for the main frontend user interface, d3.js and plotly.js for the visualizations, and utilizes the Django Python web framework for the backend. Data processing is done in the backend using the Pandas and NetworkX library (same as the MathFlowLens tool).

Currently, when users visit this dashboard they are presented with an interface to explore historical FH2T data from the RCT. However, as mentioned above, a primary goal was to make this dashboard generalizable to external learning platforms and their data. Therefore, by allowing such external platforms to send POST requests containing their raw data, our system can efficiently preprocess and present the created dashboard in real-time. Depending on the use case, this dashboard can then be displayed internally using an iFrame, or in a separate tab or window. This generalizable dashboard can be used by any learning platform that, like GM and FH2T, logs moment-by-moment student transformations. In doing so, we create a dashboard that is applicable, teacher-centric, and supportive for real world pedagogical usage.

Thesis Conclusion

In this multi-study thesis, we iteratively built upon the findings and developed tools to explore students' procedural pathways to ultimately bolster and support the acquisition of algebraic knowledge. First, we developed a novel tool called MathFlowLens for the analysis of students' procedural pathways while solving problems. This tool is capable of classifying students' pathways and subsequently representing them using sequential network visualizations. Second, to test the applicability of the classifications in research and predicting long-term learning gains such as students' scores measuring the individual constructs of algebraic knowledge, we estimated three linear models to seek correlations between the identified classifications and the conceptual, procedural and flexibility scores. Surprisingly, the results indicated that frequent use of dead-end pathways, rather than efficient and optimal pathways, resulted in higher conceptual and procedural scores. In fact, the frequent use of optimal and efficient pathways led to lower flexibility scores.

Next, we further investigated the surprising results regarding dead-end and optimal attempts, by creating linear and mixed-effects models predicting students' overall algebraic performance. This analysis revealed that students who exhibit what we call 'Regular Dead-ending' behavior have a significantly higher algebraic knowledge score. Finally, we deploy the developed tool, along with other visualizations and student performance metrics, by following a co-design methodology with teachers to develop a data and research driven interactive dashboard (mathflowlens.com). A vital and important feature of the created dashboard and technological infrastructure is its generalizability. In other words, any learning platform that logs detailed student transactional data may communicate and display the dashboard. In doing so, we create a dashboard that is applicable, teacher-centric, and supportive for real world pedagogical usage.

Future Directions

This work can be extended in many areas. First, it would be of academic interest to further explore the relations between the generated attempt classifications and algebraic learning outcomes. In particular, future work can focus on the sequence of pathways a student takes (e.g. dead-end → sub-optimal → sub-optimal → optimal) which may provide additional insight into students problem-solving strategies or exhibited behavior in comparison to using the frequencies of the classifications. Additionally, investigating the interplay between the detected behaviors or problem-solving strategies with demographic features (e.g. race and sex) and evaluative measures such as the levels of math anxiety can lead to the development of more personalized and student-centric learning environments.

Secondly, the developed dashboard should undergo vigorous user-testing, preferably by teachers, to test its applicability and usefulness in real world pedagogical scenarios. Consequently, the feedback from the user-testing should be incorporated, with particular focus on interpretability and cognitive load of the various dashboard components.

Finally, future work can extend the MathFlowLens tool by applying heuristic search algorithms or rule-based solvers to populate the graph rather than relying on historical raw data.

This would allow digital systems to provide contextual hints and allow teachers to monitor students in real-time while solving problems.

References

Aleven, V., Mclaren, B., Roll, I., & Koedinger, K. (2006). Toward meta-cognitive tutoring: A model of help seeking with a Cognitive Tutor. International Journal of Artificial Intelligence in Education, 16(2), 101-128.

Alonso-Fernandez, C., Calvo-Morata, A., Freire, M., Martinez-Ortiz, I., & Fernández-Manjón, B. (2019). Applications of data science to game learning analytics data: A systematic literature review. Computers & Education, 141, 103612. https://doi.org/10.1016/j.compedu.2019.103612

Anderson, J. R. (2014). *Rules of the mind*. Psychology Press.

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. The journal of the learning sciences, 4(2), 167-207.

Asquith, P., Stephens, A. C., Knuth, E. J., & Alibali, M. W. (2007). Middle school mathematics teachers' knowledge of students' understanding of core algebraic concepts: Equal sign and variable. Mathematical Thinking and Learning, 9(3), 249-272.

Australian Education Systems Officials Committee. (2006). *Statements of learning for mathematics*. Curriculum Corporations.

Baral, S., Botelho, A. F., Santhanam, A., Gurung, A., Erickson, J., & Heffernan, N. T. (2023a, June). Investigating patterns of tone and sentiment in teacher written feedback messages. In International Conference on Artificial Intelligence in Education (pp. 341-346). Cham: Springer Nature Switzerland.

Baral, S., Santhanam, A., Botelho, A., Gurung, A., & Heffernan, N. (2023b, July). Automated Scoring of Image-based responses to Open-ended mathematics question. In The Proceedings of the 16th International Conference on Educational Data Mining.

Baroody, A. J. (2013). The development of adaptive expertise and flexibility: The integration of conceptual and procedural knowledge. In *The development of arithmetic concepts and skills* (pp. 1-33). Routledge.

Cano, A. R., Fernández-Manjón, B., & García-Tejedor, Á. J. (2018). Using game learning analytics for validating the design of a learning game for adults with intellectual disabilities. *British Journal of Educational Technology*, *49*(4), 659-672.

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

Chan, J.Y.-C., Ottmar, E. R., & Lee, J.-E. (2022b). 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

Chan, J. Y. C., Byrne, C., Jerusal, J., Liu, A. S., Roberts, J., & Ottmar, E. (2023). Keep DRAGging ON: Is solving more problems in DragonBox 12+ associated with higher mathematical performance during the COVID-19 pandemic?. British Journal of Educational Technology.

Chen, X. (2013). STEM attrition: College students' paths into and out of STEM fields. Statistical Analysis Report. NCES 2014–001.

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. https://doi.org/10.1007/s40593-020-00202-6.

Chui, K. T., Fung, D. C. L., Lytras, M. D., & Lam, T. M. (2020). Predicting at-risk university students in a virtual learning environment via a machine learning algorithm. *Computers in Human behavior*, *107*, 105584.

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

Common Core. (2010). *Common core state standards for mathematics*. Retrieved from https://www.thecorestandards.org/assets/CCSSI_Math%20Standards.pdf.

Corbett, A. T., Koedinger, K. R., & Anderson, J. R. (1997). Intelligent tutoring systems. In Handbook of human-computer interaction (pp. 849-874). North-Holland.

Crossley, S., Karumbaiah, S., Ocumpaugh, J., Labrum, M. J., & Baker, R. S. (2019). Predicting math success in an online tutoring system using language data and click-stream variables: A longitudinal analysis. In *2nd Conference on Language, Data and Knowledge (LDK 2019)*.

Decker-Woodrow, L. E., Mason, C. A., Lee, J. E., Chan, J. Y. C., Sales, A., Liu, A., & Tu, S. (2023). The impacts of three educational technologies on algebraic understanding in the context of COVID-19. *AERA open*, *9*, 23328584231165919.

Dede, C. (2009). Immersive interfaces for engagement and learning. *Science*, 323(5910), 66–69.

Feldon, D. F. (2007). Cognitive load and classroom teaching: The double-edged sword of automaticity. Educational psychologist, 42(3), 123-137.

Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment (CIE), 1*(1), 20

Gobert, J. D., Sao Pedro, M., Raziuddin, J., & Baker, R. S. (2013). From log files to assessment metrics: Measuring students' science inquiry skills using educational data mining. *Journal of the Learning Sciences, 22*(4), 521–563.

Gurung, A., Botelho, A. F., & Heffernan, N. T. (2021, April). Examining student effort on help through response time decomposition. In LAK21: 11th international learning analytics and knowledge conference (pp. 292-301).

Gurung, A., Botelho, A., Thompson, R., Sales, A., Baral, S., & Heffernan, N. (2022, November). Considerate, unfair, or just fatigued? examining factors that impact teachers. In Proceedings of the 30th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education.

Gurung, A., Baral, S., Lee, M. P., Sales, A. C., Haim, A., Vanacore, K. P., ... & Heffernan, N. T. (2023a, July). How Common are Common Wrong Answers? Crowdsourcing Remediation at Scale. In Proceedings of the Tenth ACM Conference on Learning@ Scale (pp. 70-80).

Gurung, A., Baral, S., Vanacore, K. P., Mcreynolds, A. A., Kreisberg, H., Botelho, A. F., ... & Hefferna, N. T. (2023b, March). Identification, Exploration, and Remediation: Can

Teachers Predict Common Wrong Answers? In LAK23: 13th International Learning Analytics and Knowledge Conference (pp. 399-410).

Hästö, P., Palkki, R., Tuomela, D., & Star, J. R. (2019). Relationship between mathematical flexibility and success in national examinations. *European Journal of Science and Mathematics Education, 7*(1), 1–13. https://doi.org/10.30935/scimath/9530

Hagberg, A., Swart, P., & S Chult, D. (2008). *Exploring network structure, dynamics, and function using NetworkX* (No. LA-UR-08-05495; LA-UR-08-5495). Los Alamos National Lab.(LANL), Los Alamos, NM (United States).

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, 470-497.

Hodges, J., & Mohan, S. (2019). Machine learning in gifted education: A demonstration using neural networks. *Gifted Child Quarterly*, *63*(4), 243-252.

Holstein, K., McLaren, B. M., & Aleven, V. (2017, March). Intelligent tutors as teachers' aides: exploring teacher needs for real-time analytics in blended classrooms. In *Proceedings of the seventh international learning analytics & knowledge conference* (pp. 257-266).

Hong, W., Star, J.R., Liu, RD. *et al.* A Systematic Review of Mathematical Flexibility: Concepts, Measurements, and Related Research. *Educ Psychol Rev* **35**, 104 (2023). https://doi.org/10.1007/s10648-023-09825-2

Hopfenbeck, T. N. (2020). Making feedback effective?. Assessment in Education: Principles, Policy & Practice, 27(1), 1-5.

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.

Jere-Folotiya, J., Chansa-Kabali, T., Munachaka, J. C., Sampa, F., Yalukanda, C., Westerholm, J., et al. (2014). The effect of using a mobile literacy game to improve literacy levels of grade one students in zambian schools. *Educational Technology Research and Development, 62*(4), 417–436.

Johnson, B., & Shneiderman, B. (1998). *Tree-maps: A space filling approach to the visualization of hierarchical information structures*. UM Computer Science Department; CS-TR-2657.

Robinson, K. M., Ninowski, J. E., & Gray, M. L. (2006). Children's understanding of the arithmetic concepts of inversion and associativity. *Journal of experimental child psychology*, *94*(4), 349-362.

Koedinger, K. R., & Nathan, M. J. (2004). The real story behind story problems: Effects of representations on quantitative reasoning. *The journal of the learning sciences*, *13*(2), 129-164.

Koedinger, K. R., McLaughlin, E. A., & Heffernan, N. T. (2010). A quasi-experimental evaluation of an on-line formative assessment and tutoring system. Journal of Educational Computing Research, 43(4), 489-510.

Kiili, K., Devlin, K., Perttula, T., Tuomi, P., & Lindstedt, A. (2015). Using video games to combine learning and assessment in mathematics education. *International Journal of Serious Games, 2*(4), 37–55.

Knuth, E. J., Stephens, A. C., McNeil, N. M., & Alibali, M. W. (2006). Does understanding the equal sign matter? Evidence from solving equations. Journal for Research in Mathematics Education, 37(4), 297–312.

Kumar, K., & Thakur, G. S. M. (2012). Advanced applications of neural networks and artificial intelligence: A review. *International journal of information technology and computer science*, *4*(6), 57. https://doi.org/10.5815/ijitcs.2012.06.08.

Lee, J. E., Chan, J. Y. C., Botelho, A., & Ottmar, E. (2022a). Does slow and steady win the race?: Clustering patterns of students' behaviors in an interactive online mathematics game. E*ducational technology research and development, 70(5)*, 1575-1599.

Lee, J. E., Hornburg, C. B., Chan, J. Y. C., & Ottmar, E. (2022b). Perceptual and number effects on students' initial solution strategies in an interactive online mathematics game. *Journal of Numerical Cognition*, *8*(1), 166-182.

Lee, J. E., Stalin, A., Ngo, V., Drzewiecki, K., Trac, C., & Ottmar, E. (2022c). Show the flow: Visualizing students' problem-solving processes in a dynamic algebraic notation tool. *Journal of Interactive Learning Research*, *33*(2), 97-126.

Martin, T., Petrick Smith, C., Forsgren, N., Aghababyan, A., Janisiewicz, P., & Baker, S. (2015). Learning fractions by splitting: Using learning analytics to illuminate the development of mathematical understanding. *Journal of the Learning Sciences, 24(4)*, 593–637. https://doi.org/10.1080/10508406.2015.1078244.

McKinney, W. (2010). Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference* (Vol. 445, No. 1, pp. 51-56).

Mehlhorn, K., & Sanders, P. (2007). Algorithms and data structures: The Basic Toolbox. https://doi.org/10.1007/978-3-540-77978-0

Mendicino, M., Razzaq, L., & Heffernan, N. T. (2009). A comparison of traditional homework to computer-supported homework. Journal of Research on Technology in Education, 41(3), 331-359.

Ministry of Education of Singapore. (2006). *Secondary mathematics syllabuses*.

Ministry of Education of the People's Republic of China. (2022). *Mathematics curriculum standards for compulsory education.*

Moses, R. P., & Cobb, C. E. (2002). *Radical equations: Math literacy and civil rights* (No. 19868). Beacon Press.

Murphy, R., Roschelle, J., Feng, M., & Mason, C. A. (2020). Investigating efficacy, moderators and mediators for an online mathematics homework intervention. Journal of Research on Educational Effectiveness, 13(2), 235-270.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. US Department of Education.

Ocumpaugh, J., San Pedro, M. O., Lai, H. Y., Baker, R. S., & Borgen, F. (2016). Middle school engagement with mathematics software and later interest and self-efficacy for STEM careers. *Journal of Science Education and Technology*, *25*, 877-887.

Ottmar, E., Landy, D., & Goldstone, R. L. (2012). Teaching the perceptual structure of algebraic expressions: Preliminary findings from the pushing symbols intervention.

The Proceedings of the Thirty-Fourth Annual Conference of the Cognitive Science Society, 2156–2161.

Ottmar, E., Landy, D., Goldstone, R. L., & Weitnauer, E. (2015). Getting From Here to There!: Testing the Effectiveness of an Interactive Mathematics Intervention Embedding Perceptual Learning. In *CogSci*.

Ottmar, E., Lee, J. E., Vanacore, K., Pradhan, S., Decker-Woodrow, L., & Mason, C. A. (2023). Data from the Efficacy Study of From Here to There! A Dynamic Technology for Improving Algebraic Understanding. *Journal of Open Psychology Data*, *11*, 5.

Owen, V. E., Anton, G., & Baker, R. (2016, July). Modeling user exploration and boundary testing in digital learning games. In *Proceedings of the 2016 conference on user modeling adaptation and personalization* (pp. 301-302).

Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014). Effectiveness of cognitive tutor algebra I at scale. Educational Evaluation and Policy Analysis, 36(2), 127-144.

Pradhan, S., Gurung, A., & Ottmar, E. (2024a, March). Gamification and Dead-ending: Unpacking Performance Impacts in Algebraic Learning. In *Proceedings of the 14th Learning Analytics and Knowledge Conference* (pp. 899-906).

Pradhan, S., Lee, J., Egorova, A., Jerusal, J., & Ottmar, E. (2024b). An Application of Data Mining Methods on In-Game Behaviors: Predicting Student Math Performance. In *Proceedings of the International Conference of the Learning Sciences (ICLS) 2024*. New York, USA: International Society of the Learning Sciences.

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. Journal of Educational Psychology, 93(2), 346–362. https://doi.org/10.1037/0022-0663.93.2.346.

Rittle-Johnson, B., Schneider, M., & Star, J. R. (2015). Not a one-way street: Bidirectional relations between procedural and conceptual knowledge of mathematics. *Educational Psychology Review, 27(4)*, 587–597. https://doi.org/10.1007/s10648-015-9302-x.

Ritter, S., & Fancsali, S. (2016). MATHia X: The Next Generation Cognitive Tutor. In EDM (pp. 624-625).

Robinson, K. M., Ninowski, J. E., & Gray, M. L. (2006). Children's understanding of the arithmetic concepts of inversion and associativity. Journal of experimental child psychology, 94(4), 349-362.

Schneider, M., Rittle-Johnson, B., & Star, J. R. (2011). Relations among conceptual knowledge, procedural knowledge, and procedural flexibility in two samples differing in prior knowledge. *Developmental Psychology, 47(6)*, 1525–1538. https://doi.org/10.1037/a0024997.

Schoenfeld, A. H. (2007). Problem solving in the United States, 1970–2008: Research and theory, practice and politics. *ZDM*, *39(5–6)*, 537–551. https://doi.org/10.1007/s11858-007-0038-z.

Schwarz, B. B., Swidan, O., Prusak, N., & Palatnik, A. (2021). Collaborative learning in mathematics classrooms: Can teachers understand progress of concurrent collaborating groups?. Computers & Education, 165, 104151.

Shute, V. J., D'Mello, S., Baker, R., Cho, K., Bosch, N., Ocumpaugh, J., ... & Almeda, V. (2015). Modeling how incoming knowledge, persistence, affective states, and in-game progress influence student learning from an educational game. *Computers & Education*, *86*, 224-235.

Spector, J. M., Polson, M. C., & Muraida, D. J. (Eds.). (1993). Automating instructional design: Concepts and issues.

Star, J. R., & Rittle-Johnson, B. (2009). Making algebra work: Instructional strategies that deepen student understanding, within and between representations. *ERS Spectrum*.

Star, J. R., Pollack, C., Durkin, K., Rittle-Johnson, B., Lynch, K., Newton, K., & Gogolen, C. (2015). Learning from comparison in algebra. Contemporary Educational Psychology, 40, 41–54. https://doi.org/10.1016/j.cedpsych.2014.05.005.

Steenbergen-Hu, S., & Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. Journal of educational psychology, 105(4), 970.

Threlfall, J. (2009). Strategies and flexibility in mental calculation. *ZDM*, *41*, 541-555.

Vanacore, K., Sales, A., Liu, A., & Ottmar, E. (2023). Benefit of gamification for persistent learners: Propensity to replay problems moderates algebra-game effectiveness. In *Proceedings of the Tenth ACM Conference on Learning@ Scale* (pp. 164-173).

Venkat, H., Askew, M., Watson, A., & Mason, J. (2019). Architecture of mathematical structure. *For the Learning of Mathematics*, *39*(1), 13-17.

Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Van Assche, F., Parra, G., & Klerkx, J. (2014). Learning dashboards: an overview and future research opportunities. *Personal and Ubiquitous Computing*, *18*, 1499-1514.

Verschaffel, L., Luwel, K., Torbeyns, J., & Van Dooren, W. (2009). Conceptualizing, investigating, and enhancing adaptive expertise in elementary mathematics education. *European Journal of Psychology of Education*, *24*, 335-359.

Walkoe, J., Wilkerson, M., & Elby, A. (2017). Technology-mediated teacher noticing: A goal for classroom practice, tool design, and professional development. Philadelphia, PA: International Society of the Learning Sciences.

Welder, R. M. (2012). Improving algebra preparation: Implications from research on student misconceptions and difficulties. *School Science and Mathematics, 112*(4), 255–264.

Wouters, P., Van Nimwegen, C., Van Oostendorp, H., & Van Der Spek, Erik D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *American Psychological Association.*

Wouters, P., Van Oostendorp, H., ter Vrugte, J., Vandercruysse, S., de Jong, T., & Elen, J. (2017). The effect of surprising events in a serious game on learning mathematics. *British journal of educational technology*, *48*(3), 860–877. https://doi.org/10.1111/bjet.12458.

Xu, L., Liu, R. D., Star, J. R., Wang, J., Liu, Y., & Zhen, R. (2017). Measures of potential flexibility and practical flexibility in equation solving. Frontiers in Psychology, 8, 1368.

Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., et al. (2012). Our princess is in another castle: A review of trends in serious gaming for education. *Review of Educational Research, 82*(1), 61–89. https://doi.org/10.3102/0034654312436980.

Zhang, K., & Aslan, A. B. (2021). AI technologies for education: Recent research & future directions. *Computers and Education: Artificial Intelligence*, *2*, 100025.

Appendix A

Item	Problem Type	Question	Correct Answer
1	Procedural	Solve the equation for y. 5 (y - 2) = 3 (y - 2) + 4 a. 1 b. 5/2 c. 4 d. 10	b
2	Flexibility	Kim solved the problem: 1/3(x + 5) = 4 Kim's first step was: (3) 1/3(x + 5) = 4(3) x + 5 = 12 What step did Kim use to get from the first line to the second line? a. combine like terms b. distribute across parentheses c. add or subtract the same quantity on both sides d. multiply or divide the same quantity on both sides	d
3	Conceptual	3+4=7 ↑ What does this symbol mean? a. the total b. two quantities on either side have the same value c. what the answer is d. the problem has been solved	b
4	Flexibility	Imagine you are taking a timed test. You want to use fast (and correct) ways to solve the problems so you can finish as many as possible. Which would be the best way to start the problem 3(x + 2) = 14? a. distribute b. divide both sides c. multiply both sides d. subtract both sides	a
5	Procedural	Solve the equation below for x. 3 = (8 - 6x)/2 a. -1/3 b. 0 c. 8 d. 1/3	d
6	Conceptual	If 10x + 12 = 17, which of the following must also be true? a. 10x + 12 - 12 = 17 - 12 b. 10x - 10 + 12 - 10 = 17	a

		c. -10x - 12 = 17 d. 5x + 6 = 17	
7	Conceptual	Which of the following is equivalent to (the same as) (n + 3) + (n + 3) + (n + 3) + (n + 3)? a. n + 12 b. 4n + 3 c. n4 + 12 d. 4(n + 3)	d
8	Procedural	Solve the equation below for n. 12n + 3 = 14n + 15 - 8n a. 0 b. 2/3 c. 2 d. 3	c
9	Flexibility	Imagine you are taking a timed test. You want to use fast (and correct) ways to solve the problems so you can finish as many as possible. Which would be the best way to start the problem? (Choose the letter for the best way to start) 1/4 (5x + 2) = 8 a. distribute first b. subtract 8 from both sides c. multiply by 4 on both sides first d. divide by 4 on both sides first	c
10	Conceptual	Which of the following is NOT equivalent to 19(73 - 15)? a. 19(58) b. 19(73) - 19(15) c. 19(15 + 73) d. 19(73) - 15	d

Appendix A. The items included in the pretest algebraic knowledge assessment (adapted from Star et al., 2015; 4 conceptual items, 3 procedural items, 3 flexibility items)

Appendix B

Measure	Description
Completions	Percentage of students that completed the problem
Number of Attempts	Average number of attempts for this problem
Number of Hints	Average number of hints requested by students
Number of Resets	Average number of resets (i.e. restart problem to initial start state)
Number of steps in First Attempt	The average number of steps in students' first completed attempt
Number of steps in Last Attempt	The average number of steps in students' final completed attempt
Efficiency First Attempt	The average efficiency score in the first completed attempt
Efficiency Last Attempt	The average efficiency score in the final completed attempt
Time Taken (sec)	The average time taken in seconds to complete the problem
Pause Time First Attempt (sec)	The average time students spent thinking about the problem before making a first step in the first completed attempt
Pause Time Last Attempt (sec)	The average time students spent thinking about the problem before making a first step in the final completed attempt
Number of Total Errors	The average number of errors students made
Number of Keypad Errors	The average number of errors resulting from the incorrect decomposition of a number
Number of Shaking Errors	The average number of errors resulting from the combining of unlike terms
Number of Snapping Errors	The average number of errors resulting from incorrect 'drag-and-drop' action

Appendix B. All measures and their respective descriptions included in the Performance Metrics Table