UNDERSTANDING AND ENGAGEMENT THROUGH DYNAMIC TECHNOLOGY AND GAMIFIED LEARNING ENVIRONMENTS

by

Daniel Vito Manzo

A Dissertation

Submitted to the Faculty
of
WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Doctor of Philosophy
in
Learning Sciences & Technologies
April 2020

APPROVED:

Erin Ottmar, Ph.D.

Neil T. Heffernan III, Ph.D.

Ivon Arroyo, Ph.D.

David Landy, Ph.D., Indiana University Bloomington

ABSTRACT

As technology becomes more integrated in the classroom, more research is needed to examine its effects on engagement and learning. It is important that we fully explore how students interact with learning technologies and the affordances that these technologies bring to improve engagement and student learning. In this dissertation, I explored the benefits and drawbacks of using dynamic technology in the classroom as an instructional system, support structure, and assessment tool. Iterative design cycles were used to improve the accessibility and user experience of several dynamic technologies in the classroom. Additionally, the incorporation of gamified elements such as points and leaderboards were explored. Preliminary data suggests that gamified elements could lead to higher engagement and elicit behaviors associated with learning. As a result, a series of 4 randomized controlled trials were conducted that explored the intersection of gamification, engagement, and learning. This dissertation is a compilation of those studies with a focus on the development and improvement of learning platforms through an iterative design process and the incorporation of gamified elements. Based on the findings and implications of these studies, several new technologies were designed, developed, and implemented to include these gamification techniques and provide data for both educators and researchers. Recommendations for potential usage and future research are discussed.

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank those who have supported me throughout my graduate career. First and foremost, I would like to thank God for guiding me through this and keeping me hidden as much as possible. I would also like to thank my incredibly patient wife, Robin, who willingly decided to put up with my nonsense permanently. If nothing else, tricking her into marrying me is my greatest achievement. I would not be here without the endless support from her, Zoey, Ruth, and Amelia.

Next, I would like to thank my committee for their help and support as I progressed through the doctoral program. I am very grateful for the faculty of the Learning Sciences and Technologies program for their availability and feedback when needed.

To the ATA of DATA, I consider myself blessed for having three amazing individuals to work beside. I'm incredibly grateful for our time together. I have no doubt that you all will do amazing things in life, I'm just happy to have been a small part of it.

To my friend, John Barron. Thank you for your constant reminders to keep moving forward regardless of the circumstance. You'll never know what your encouragement means to me.

To my family, friends, and colleagues past and present, thank you for being there or understanding when I wasn't. We'll hang out eventually.

Finally, I would like to acknowledge my math teacher, Mrs. B., to whom this dissertation is dedicated. Her lifetime of devotion and dedication keeps me moving forward. She accompanied me to my first college interview and it is only right that she is honored at the end of my graduate career. "An hour of wolves and shattered shields when the Age of Men comes crashing down, but it is not this day." She would have enjoyed this work more than anyone.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FOREWORD

The current work consists of two main parts. Part I provides the foundation for why this work matters as well as the current state of this area of educational research as well as a foundation for the technical achievements that I have developed either by myself or alongside the Graspable Inc. development team at Indiana University Bloomington. Chapters 1 and 2 detail the evolution of the technical achievements that I've developed during my time in the Ph.D. program as well as the research and design that I've expanded with two applications: Graspable Math and From Here to There. This work has played a large part in the development and expansion of these tools in classrooms, the way these tools are deployed at scale for research, and ultimately the learning sciences community. These chapters have built the foundation of my work and are comprised mostly of submitted and accepted papers containing modifications that better explain the overall theme of this work. Part II, Gamification and Games for Education, introduces the concept of gamification as a catalyst for motivation and engagement and explores various research projects at the intersection of gamification and learning environments. Furthermore, this section introduces an original work that further demonstrates some of the technical proofs and resulting behaviors that are explored earlier in the chapters on games in education.

Before I present this work, it is imperative that I acknowledge the path behind me to better describe my graduate experience and explain how the current work came to be. I first began my graduate career in January 2014 where I was baptized by fire into the world of research during Neil Heffernan's Artificial Intelligence in Education course. After introducing myself to the class as "Dan from New Jersey", Professor Heffernan asked me if I knew of a particular town where they were just greenlighted to conduct a longitudinal study (coincidentally, it was the next town over from where I grew up) and asked if I would be interested in taking the lead on this study. Being completely new to the university, ASSISTments, and empirical research at the time, I naturally agreed. It was during these next few months that I began to sit in on several weekly phone calls, meetings, and Google Hangouts and develop an understanding for how randomized controlled trials were executed. Shortly before finishing my Master's program, I was introduced to my future advisor, Erin Ottmar, by Professor Heffernan. The following January, I began the Ph.D. program and was introduced to Graspable Math and the amazing colleagues at Indiana University Bloomington. It was in these early months that I expanded on the initial work of Graspable Math to allow teachers and researchers to design their own problems and assignments. Having worked on ASSISTments in my previous program, I developed a working integration that allowed Graspable Math goal state problems to be deployed from within ASSISTments and facilitated user data to be logged in both ASSISTments and in Graspable Math.

As Graspable Math grew as a notation tool, our research began to expand. One of the undergraduate students in the lab, Lindsay, had expressed interest in early algebra introduction research as well as the native iPad app and predecessor of Graspable Math, From Here to There!. Unfortunately, we found that the updates, deployment, and data retrieval using the iPad proved to be difficult in itself, not to mention the unfortunate fact that very few schools had iPads. This resulted in the urgent need for a browser-based implementation of From Here to There! that can be used on regular computers as well as on iPads. After a series of design iterations, user-testing, pilot studies, and data analyses, I now had a suitable testbed for running studies in and out of classrooms regardless of the age range or content. During the following semester, I began to explore more metacognitive and affective topics in my studies and became interested in motivating students to engage with interventions. Having come from a discipline that revolved around serious games for education, I became interested in whether or not the gamified aspect of education was worthwhile. My initial designs for Graspable Math problems featured a minimalist design with monochromatic colors which clearly differed from the bright colorful backgrounds of From Here to There! Furthermore, there were gamified elements in From Here to There! that served no purpose other than its extrinsic value as an achievement. This interest ultimately lead to the research that is featured in Part II. Shortly after the end of that study, I wanted to explore the effects of other gamified elements in a learning environment. PBL's (points, badges, and leaderboards) are often the most common elements of gamification as they are easy to implement and can be implemented into most activities. I, however, have seen very little research as to the effect of leaderboards when it comes to engagement (and more specifically disengagement and discouragement) in an educational environment. I have often wondered whether or not this type of competitive display would discourage those on the bottom and disengage those on the top out of boredom or lack of challenge. In my review, I point out that a lot of the current literature on gamification and gamified elements often discusses the effects of gamification as a whole, but fails to inspect the effects of individual aspects or components such as leaderboards in an isolated setting.

The second last chapter of this work features an original work entitled Treasure Hunter. I feel strongly that this project is the end result of its preceding chapters. During the creation of Treasure Hunter, I carefully focused on various aspects of technical game design to ensure that it maintained a fun and engaging interface that incorporated gamification techniques such as pointification (that visually animates as additional points are attained) as well as the technical necessities discussed in the earlier chapters for proper logging and recording of data. Lastly, I had to ensure that the cognitive properties of my educational work were included as well so that it was not just an aesthetically-pleasing recording application, but one that also targeted cognitive principles as well. As you will see in its design, the tasks themselves pose as algebraic puzzles that are not complicated to understand, but perplexing to solve. While there were other projects that were to be included in addition to Treasure Hunter that focused on more complex algebraic tasks, it was decided that it was better to put my best foot forward for this dissertation, and I feel confident that Treasure Hunter is some of my best work to date.

Lastly, this work concludes with a synthesis chapter that summarizes and integrates Parts I & II. This chapter reflects on the intersection of dynamic technology, gamification, and games in education and further promotes my contributions both in technical achievement as well as educational research. Additionally, it discusses the future of Treasure Hunter as an online resource for parents and teachers. It is my honor to present this dissertation work as a reflection of my graduate work and thank you for reading.

1 INTRODUCTION

1.1 Overview

Understanding the student perspective of mathematical tasks and identifying prior misconceptions are two of the more difficult tasks in education and research (Ysseldyke & Bolt, 2007). Within mathematics education, a common approach to understanding student knowledge is to ask students to explain their problem solving process on paper or out loud in class (Chapin, O'Connor, & Anderson, 2009). While asking students to "show their work" can be effective on a small scale, it is not feasible or scalable for teachers to compile, organize, and analyze students' step-by-step processing on paper-based assessments. Using educational platforms in the classroom can help make similar data collection more feasible at a larger scale and may help collect additional information that would not be available using paper assessments (Manzo et al., 2016). Some of our prior work demonstrates the feasibility of using dynamic technology to reveal new information about student problem solving strategies and mathematical flexibility. While it is now generally accepted that technology in the classroom has positive effects on the classroom, it is important that we fully explore how students interact with learning technology. Although there are many data advantages of using technology, there are also some drawbacks to using technology in the math classroom. For example, like most online systems, the student and teacher are both required to be connected to the Internet while using the technology which can create socioeconomic barriers and serious accessibility issues for schools that are not readily equipped with Internet-capabilities. In some of our prior work, we have been successful in implementation and feasibility. However, in-school scalability demonstrated a need for technology that does not add a socio-economic barrier, but rather utilizes existing available resources that are accessible and available to all. Just as it was with initiatives such as Apple's Classrooms of Tomorrow and "Computers in the Classroom", positive outcomes which are primarily correlated to exposure would only pertain to more affluent groups (Attewell & Battle, 1999; Becker & Ravitz, 1998; Milone & Salpeter, 1996). Thus, increased levels of student-to-student interaction in computer learning environments appear to provide positive levels of student achievement. Unfortunately, just as home computers were a luxury for wealthy students in the 1990's and 2000's, learning environments that depend on touchscreen iPad tablets and other modern technologies are simply not available in many schools, thus creating a socio-economic barrier in classrooms. Additionally with the rise of Chromebooks as affordable and accessible devices, dynamic technologies need to be able to run in virtually any modern browser without having to install or store anything locally was necessary. Furthermore, as the overall experience of the app was crucial to its effectiveness, platforms need to be adaptive and responsive to provide as much of a similar experience as possible when being used in various environments, devices, screen sizes, input peripherals, browsers, and operating systems. In addition to designing and implementing effective dynamic technology, the overall experience should be engaging for students. A goal for most teachers and researchers is to not only teach a student so that they can learn the material, but also motivate them to care about the material they are learning and understand its value. Durik & Harackiewicz (2007) demonstrates the concept of "catch and hold" as a method that "entices" students in mathematics and then once they are in the door, demonstrates the utility of mathematics in an effort to keep them engaged and see the intrinsic value. Their findings suggest that the engagement levels of students with a higher interest in mathematics were either not affected or negatively affected when presented with a mathematical notebook filled with cartoons and colorful images, while students with a lower interest in mathematics were more engaged and positively affected by the colorful notebook. Mitchell (1993) also suggests that explaining the value or utility of an educational concept piques interest. Mitchell argues that teachers must “sell” an interest to their students in order for them to “buy” into the idea of being interested in mathematics. This can be done through various approaches such as lab activities that involve mathematical procedures or more culturally relevant mediums such as video games. As video games have become rapidly popular and are quickly to be engaged by young students, games can be used as the “sugar pill of learning” (Falstein, 2005; Warren 2009). Research suggests that video games can motivate and interest learners (Dempsey et al., 1994; Jacobs & Dempsey, 1993; Lepper & Malone, 1987; Malone, 1981, 1983; Malone & Lepper, 1987; Malouf, 1988), increase retention rates (Dempsey et al., 1994; Jacobs & Dempsey, 1993; Pierfy, 1977), and improve reasoning skills and higher-order thinking (Mayland, 1990; Rieber, 1996; Wood & Stewart, 1987; Hogle, 1996). Despite this success, educational video games are often not embraced as willingly as mainstream entertainment video games.

1.2 Research Goals

The current research presented explores the intersection of math education, the development and implementation of dynamic technology, and gamification in education. Though some chapters are solely focused on one of these topics, this work as a whole is intended to explore how the integration and combination of all three of these areas can be applied both in and out of the classroom. While each of the following chapters are written to supplement the main research question of the dissertation, some of the chapters will feature their own research questions that are answered in the subsequent studies. To reach the goals of further expanding the knowledge of the scientific community and improving development of learning technologies, our main research questions are as follows:

1) **What are the benefits and disadvantages of using dynamic technology in math education?**
2) **How can we better design learning technologies to be more efficient for research and integration in classrooms?**
3) **What are the effects and affordances of gamification in dynamic learning technologies on engagement and learning?**

PART 1: DYNAMIC TECHNOLOGY AND THE STUDENT PERSPECTIVE

2 DYNAMIC TECHNOLOGY

The following chapter was originally published as a conference paper for the National Council for Teachers in Mathematics (NCTM). This chapter focuses on feedback given to teachers and researchers when students solve algebra problems using an intelligent tutoring system (ITS). When students use an ITS, teachers are typically able to see information about student progress. While this information is helpful to assess whether or not the student solves the problem correctly, it gives little insight into *how* the student solved the problem as well as their conception when reaching their solution. This paper explores two approaches in an ITS – a dynamic technology and static text – and the feedback that is reported. Comparisons of the data output are made between the two approaches that teachers and researchers can use to assess both student and classroom understandings.

2.1 Introduction

2.1.1 Overview

Teachers must be able to understand student perspectives and assess a student's prior knowledge before they can properly build and expand on student knowledge (Popham, 1999; Airasian, 1991). Assessing each student's understanding of tasks and identifying their misconceptions is one of the more difficult tasks in education and research (Ysseldyke & Bolt, 2007). Within mathematics education, a common approach to understanding student knowledge is to ask students to explain their problem solving process on paper or out loud in class (Chapin, O'Connor, & Anderson, 2009). While asking students to "show their work" can be effective on a small scale, it is not feasible or scalable for teachers to compile, organize, and analyze students' step-by-step processing on paper-based assessments. Furthermore, there is no additional meta-knowledge about each process such as the amount of time that a student spent on each problem. Using educational platforms in the classroom can help make similar data collection more feasible at a larger scale and may help collect additional information that would not be available using paper assessments. While most educational platforms report student data such as correctness, time spent, or hints used (Feng & Heffernan, 2006), very little information is collected on the student strategies used or reasoning applied during the problem solving process. More specifically, we do not know *how* they solved the problem. This chapter focuses on the use of dynamic technology in mathematics education, which has the potential to record more information on the student problem solving process and support teachers in identifying student understanding and misconceptions in innovative and efficient ways.

2.1.2 Theoretical Framework

The National Research Council (2001), Common Core Standards for Mathematical Proficiency (2010), and the NCTM Process standards (2000), all highlight that mathematical proficiency requires a set of mutually dependent skills that can be represented by five strands: *procedural fluency*, *strategic competence*, *adaptive reasoning, conceptual understanding, and productive disposition* (Figure 2.1). Having ***procedural fluency*** refers to executing algorithms or strategies efficiently and flexibly (Star, 2005; Hiebert, 2013). ***Strategic competence,*** or the ability to formulate, represent, and solve math problems, and ***adaptive reasoning***, the ability to consider alternative strategies, compare the usefulness of options, and explain specific strategy choices, both make up what is commonly referred to as mathematical flexibility (Baroody, 2003). Students that can recognize, reason about, and apply mathematical processes when problem solving have a strong ***conceptual understanding***. The final strand of mathematical proficiency, ***productive disposition,*** focuses on motivation and the ability to see the utility and logic of mathematics.

Figure 2.1. An illustration of the constructs and indicators that make up mathematical proficiency

Despite significant research on the five components of mathematical proficiency and the importance of each component in the overall goal of fluency, efficiency still remains the primary form of assessment in K-12 classrooms (Schoenfeld, 2007). These assessments typically apply a binary value of student accuracy - correct or incorrect - and neglect the other equally important components of student problem solving. Instead of focusing primarily on efficiency, assessments of mathematical proficiency should focus also on measuring the other important components as well. These measures may help teachers in identifying early indicators of student struggle and may give insight into student understanding when solving problems. In this chapter, we highlight some of the benefits (and disadvantages) of using educational technology as an aid and form of assessment for mathematics teachers.

2.1.3 Research Goals

The goal of this chapter is to examine how teachers and researchers can gain a better data-driven understanding of student problem solving processes by using dynamic technology. We conducted a study where students used a dynamic mathematical learning technology (Graspable Math) in an online tutoring system. We compare the data obtained through this technology to data obtained from using static text in an online platform. We aim to: 1) examine the different types of knowledge of student perspective that can be gained from using dynamic technology, 2) highlight the benefits and disadvantages of using dynamic technology, and 3) discuss how the data report obtained through dynamic technology could be useful to teachers and researchers.

2.2 Methods

2.2.1 ASSISTments

ASSISTments is an online tutoring system that can provide scaffolding and feedback to assist students (Heffernan & Heffernan, 2014). The platform, hosted by Worcester Polytechnic Institute, allows teachers and researchers to create individual assignments composed of questions with answers and associated hints. ASSISTments has been found to improve student learning when used as a homework tool in combination with teacher training (Roschelle, Feng, Murphy, & Mason, 2016). For this chapter, we used ASSISTments as the main platform to run our study as it supports both static content and embedded dynamic technology while supplying detailed reports for both conditions that are consistent with many learning platforms.

2.2.2 Graspable Math

Graspable Math (GM) is a web-based dynamic algebra notation system with a powerful user interface for manipulating algebraic expressions (Weitnauer, Landy, & Ottmar, 2016). Instead of solving problems on paper and typing in an answer, students physically manipulate terms in an equation using specific gestures to perform step-by-step problem solving. With the support of immediate feedback and scaffolding, GM is designed so that students can use motion to learn the underlying structure of algebra (Weitnauer, Landy, & Ottmar, 2016; 2017). As users manipulate algebraic expressions and equations, clickstream data is recorded that allows teachers and researchers to fully capture the problem solving process. Graspable Math runs on most modern browsers and devices including desktops, tablets, and smart boards. GM is a free service developed in conjunction with Indiana University Bloomington.

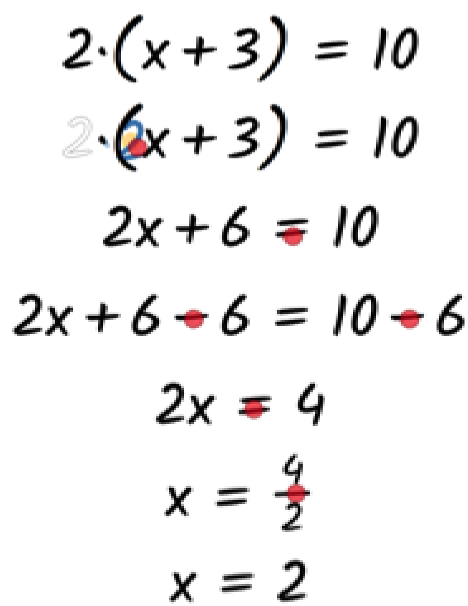


Figure 2.2. A visualization of a student's performance and strategic solution when solving the problem as well as the point of interaction highlighted in red. To solve the problem 2(x+3)=10 in Graspable Math, the student can use a mouse to drag the "2" to the right and distribute across to the "x" and "3" to make 2x+6=10. The student could then subtract 6 from both sides and then divide both sides by 2 to solve for the variable.

2.2.3 Study Procedures

Approximately 80 middle school students from 4 mathematics classes within a northeast public school district participated in the study for approximately 45 minutes. Participants within classes were randomly assigned to one of two conditions: the dynamic technology condition using Graspable Math or the static text condition with immediate correctness feedback. Data from both conditions were gathered online using the ASSISTments online platform, which sorted students into conditions to answer problems using their assigned technology.

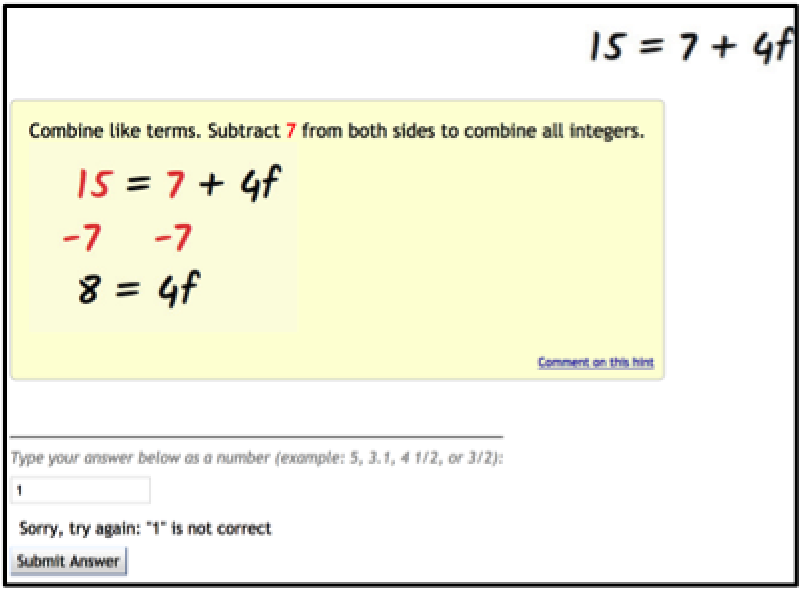


Figure 2.3. An example of requesting a hint and submitting an answer in ASSISTments.

Our goal is to examine the usefulness of the data reports generated from each condition. Students in the dynamic technology condition received a 10-minute training of how to use Graspable Math to demonstrate how to transform equations using the system gestures, but gave no additional practice or advantage to solving problems. The static text condition received no additional training. Participants in both conditions were then assigned the same 12 problems. Students in the static text condition solved problems on a sheet of paper and submitted it by typing in an answer, while the students in the dynamic technology condition solved problems using dynamic transformations in Graspable Math on the screen and submitted it by typing in an answer. During this intervention, both groups were able to receive identical hints using the native hints system in ASSISTments.

14, Participant	81%	× 3240 50%	× 0,1,2,3 66%	✓ B = A² + 1 100%	× 902 33%	00:09:04
15, Participant	81%	✓ 30240 100%	✓ 2 100%	× B = 2A + 1 50%	✓ 762,902 100%	00:08:51
16, Participant	58%	✓ 30240 100%	× 1,3 66%	✓ B = A² + 1 100%	× 902 0%	00:17:23

Figure 2.4. Example of ASSISTments class report. Green checks indicate correct answer, green X's indicate correct answer with some errors, and red X's indicate incorrect answers, skips, or out of time.

2.2.4 Data Sources

When generating the reports for our two conditions in ASSISTments, the system gives immediate correctness feedback to students while they are working and instantly provides student-level data reports to teachers on any assignment. The information reported to teachers contains either a red X for incorrect; a green X for a wrong, but later corrected answer; and a green check for a correct answer as well as the number of hints students requested and the total time spent per problem. With the dynamic technology, Graspable Math, every mouse click, drag, and interaction is reported in addition to the ASSISTments report. In addition to this data, Graspable Math also generates summary data such as the number of steps taken, resets of the problem, and submitted answers (Figure 2.4).

Time	Action	Object ID / Input text
Thu Mar 23 22:51:04 -0400 2017	Started a problem	PRABC2MW
0 mins 31 secs	Answered	PCS, Mrs. Doherty, 8th grade
Thu Mar 23 22:51:42 -0400 2017	Started a problem	PRA9SP2
0 mins 25 secs	Answered incorrectly	12 + 8t
Thu Mar 23 22:52:09 -0400 2017	Started a problem	PRA9SP3
0 mins 27 secs	Answered incorrectly	c=5
Thu Mar 23 22:52:37 -0400 2017	Started a problem	PRA9SP4
0 mins 14 secs	Answered incorrectly	x=3

Figure 2.5. Example of ASSISTments report on a student level. The green rows indicate correct answers, red rows indicate incorrect answers, and blue rows indicate open or ungraded answers

2.3 Results

We obtained teacher-reports of the student level data for both conditions through the ASSISTments control panel. ASSISTments class reporting gives an overall view of how students performed on each problem (Figure 2.4), whereas the student-level reports show a more in-depth view on each individual student's progress (Figure 2.5). Graspable Math reports on conventional measures at the class and student levels (Figure 2.6), but also provides student strategy data.

0 mins 19 secs	Answered correctly	Reset at 5 steps. Solved in 2 steps and 2 attempts.
Wed Mar 22 12:41:36 -0400 2017	Started a problem	PRABCX3N
0 mins 38 secs	Answered correctly	Solved in 2 steps and 1 attempts.
Wed Mar 22 12:42:16 -0400 2017	Started a problem	PRABCX3P
Wed Mar 22 12:46:14 -0400 2017	Resumed a problem	PRABCX3P
Wed Mar 22 12:47:16 -0400 2017	Resumed a problem	PRABCX3P
0 mins 19 secs	Answered correctly	Solved in 1 steps and 1 attempts.
Wed Mar 22 12:47:36 -0400 2017	Started a problem	PRABCX3Q
0 mins 13 secs	Answered correctly	Solved in 3 steps and 1 attempts.

Figure 2.6. An example of Graspable Math summary data in the ASSISTments individual assignment report

As part of the strategy reports, Graspable Math provides a visualization tool that can show teachers and researchers the types of strategies individual students use when solving problems, as well as which strategies were most popular for the entire classroom. Figure 2.7 shows the different strategies implemented by a group of students when solving a problem using Graspable Math.

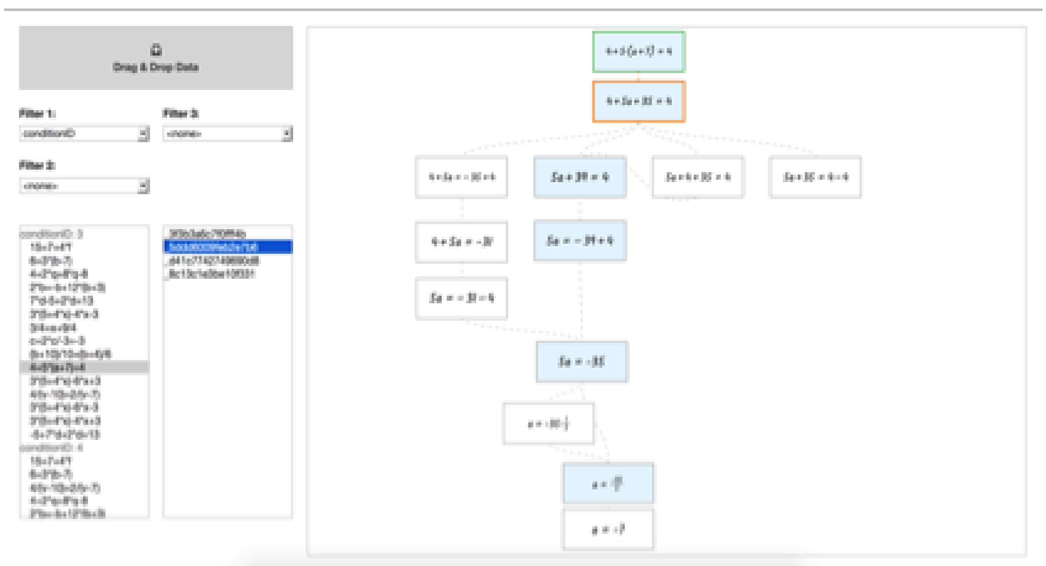


Figure 2.7. An example of the Graspable Math data visualization tool illustrating the different solutions that students used to solve the problem with a specific student's strategy highlighted in blue.

In Figure 2.8, we can see an isolated visualization of a student's problem attempt including key assessment factors such as the number of steps, amount of time taken, time and step averages, number of attempts, among others. We can also see a visual progression of the strategy itself with events such as resetting the problem, making errors, or guessing an answer illustrated in a colored-coded box. Providing this information allows teachers to gain a better understanding at *how* a student solved a problem and help devise a plan to correct any errors of student struggle or misconceptions. Early indicators of misconceptions such as an unusually high number of steps or highly inefficient strategy can then be highlighted when a teacher is viewing the report thus allowing for earlier teacher interventions.

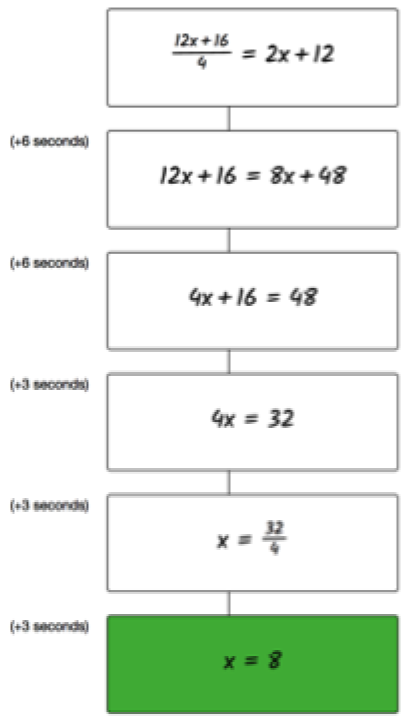


Figure 2.8. An example of the Graspable Math data visualization tool illustrating an individual student's strategy when solving a problem. Events like resetting a problem or guessing an answer are illustrated in color.

2.4 Discussion

2.4.1 Benefits of Using Dynamic Technology

Both the static text and dynamic technology conditions reported student accuracy when solving mathematical problems. This type of correctness data is on par with other kinds of computer-assisted educational environments and can be used to make the grading process easier. Beyond correctness, however, the benefits of using dynamic technology are evident. With each gesture, mouse trajectory, and step that a student executes using the dynamic technology, a data log is recorded that can easily be analyzed using visualization tools and statistical software. This provides teachers with fine-grained data that gives insight on a student's perspective and mathematical understanding. For teachers, this data has the potential to identify different strategy types which could then be used to create student profiles, predict learning, and adapt content to suit students' learning needs. An additional advantage of solving math problems on a computer is the alleviation of public embarrassment when solving math problems in front of a classroom. Such embarrassing experiences can contribute to an increase in math anxiety (Newstead, 1998; Bekdemir, 2010; Rubinsten & Tannock, 2010).

	Static Content	**Dynamic Technology**
Reports accuracy of students' answers	✔	✔
Quick grading for class	✔	✔
Allows hints	✔	✔
Readily available content	✔	✖
Requires training to use properly	✖	✔
Allows for in-depth analysis of strategy	✖	✔

Table 2.1. A comparison of static content and dynamic technology

2.4.2 Disadvantages of Using Technology in the Mathematics Classroom

Although there are many data advantages of using technology, there are also some drawbacks to using technology in the math classroom. For example, like most online systems, the student and teacher are both required to be connected to the Internet while using the technology which can create serious accessibility issues for schools that are not readily equipped with Internet-capabilities. There are also few curricula resources available that use these dynamic technologies, so content that matches the teacher's needs would have to be created independently. However, technology design is an iterative process, so it is possible that content could be developed and shared over time by dedicated users.

One of our most significant lessons learned from running this study is in the learning curve of the educational technology. One of the biggest drawbacks to using dynamic technology is that there is a learning curve for students when learning how to use the system and for teachers when navigating student data and building their own content. Although we provided a 10-minute tutorial for learning the gestures, it may prove to be difficult for students who have yet to master the gestures of the technology. This is something that should decrease significantly over time as students become more comfortable with the gestures and the software. As technology improves in both processing speed and innovation, it is likely that free educational platforms such as GM or ASSISTments will become more widely used in schools. However, the transition to using such platforms is often difficult and may be discouraging and confusing for both teachers and students alike.

2.4.3 Educational Importance

In this chapter, we explored the benefits and drawbacks of using dynamic technology in the classroom as an instructional system, support structure, and assessment tool. In addition to measurements of correctness, the GM technology performs well as an assessment tool because it requires students to externalize their problem solving strategies and procedures on screen that they would normally internalize or “do in their head” with write-in or type-in your answer problems. Teachers can also see summarizing data of the dynamic technology in the reports that highlight key assessment data without even having to look at any of the more complex visualization data. This summarizing data also provides more perspective in determining when students are struggling and may allow teachers to plan an intervention earlier than if they only saw that the student got the problem right or wrong. This will be further expanded on in the next chapter when we focus solely on the information that teachers and researchers can extract when using dynamic technology, in particular the variation that can occur when analyzing student strategies when attempting complex algebraic problems.

References

Airasian, P. W. (1991). Classroom assessment. New York: McGraw-Hill.

Anderson, J. R. (2007). How can the human mind exist in the physical world? Oxford, England: Oxford University Press.

Baroody, A. J. (2003). The development of adaptive expertise and flexibility: The integration of conceptual and procedural knowledge. The development of arithmetic concepts and skills: Constructing Adaptive Expertise, 1-33.

Bekdemir, M. (2010). The pre-service teachers' mathematics anxiety related to depth of negative experiences in mathematics classroom while they were students. Educational Studies in Mathematics, 75, 311–328, http://dx.doi.org/10.1007/s10649-010-9260-7.

Booth, J. L., & Koedinger, K. R. (2008). Key misconceptions in algebraic problem solving. In *Proceedings of the 30th annual conference of the cognitive science society (pp. 571-576)*

Chapin, S. H., O'Connor, C., & Anderson, N. C. (2009). Classroom discussions: Using math talk to help students learn, Grades K-6. Math Solutions.

Common Core State Standards Initiative (2010). Common Core State Standards for Mathematics (CCSSM).Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers.

Council, N. R. (2001). Adding it up: Helping children learn mathematics. Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education.

Feng, M., & Heffernan, N. T. (2006). Informing teachers live about student learning: Reporting in the ASSISTment system. *Technology Instruction Cognition and Learning*, *3*(1/2), 63.

Goldstone, R. L., Landy, D. H., & Son, J. Y. (2010). The education of perception. Topics in Cognitive Science, 2(2), 265-284.

Goldstone, R. L., Weitnauer, E., Ottmar, E. R., Marghetis, T., & Landy, D. H. (2016). –Modeling Mathematical Reasoning as Trained Perception-Action Procedures. Design Recommendations for Intelligent Tutoring Systems: Volume 4-Domain Modeling, 4, 213.

Heffernan, N. T., & Koedinger, K. R. (1998). A developmental model for algebra symbolization: The results of a difficulty factors assessment. In *Proceedings of the twentieth annual conference of the cognitive science society* (pp. 484-489).

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

Hiebert, J. (2013). Conceptual and procedural knowledge: The case of mathematics: Routledge.

Kellman, P. J., Massey, C. M., & Son, J. Y. (2010). Perceptual learning modules in mathematics: Enhancing students' pattern recognition, structure extraction, and fluency. Topics in Cognitive Science, 2(2), 285-305.

Popham, W. J. (1999). *Why standardized tests don't measure educational quality.* Educational leadership, 56, 8-16.

Principles and standards for school mathematics. (2000). Reston, VA: National Council of Teachers of Mathematics.

National Research Council Mathematical Sciences Education Board (Ed.). (1993). Measuring what counts: A conceptual guide for assessment . Washington, DC: National Academy Press.

Newstead, K. (1998). Aspects of children's mathematics anxiety. Educational Studies in Mathematics, 36, 53–71, http://dx.doi.org/10.1023/A:1003177809664.

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

Rubinsten, O., & Tannock, R. (2010). Mathematics anxiety in children with developmental dyscalculia. Behavioral and Brain Functions, 6, 1–13, http://dx.doi.org/10.1186/1744-9081-6-46.

Russell, S. J. (2000). Developing computational fluency with whole numbers. Teaching Children Mathematics, 7(3), 154.

Schoenfeld, A. H. (2007). Assessing mathematical proficiency (Vol. 53): Cambridge University Press.

Star, J. R. (2005). Reconceptualizing procedural knowledge. Journal for Research in Mathematics Education, 404-411.

Weitnauer, E., Landy, D., & Ottmar, E. (2016). Graspable math: Towards dynamic algebra notations that support learners better than paper. In Future Technologies Conference.

Weitnauer, E., Landy, D., & Ottmar, E. (2017). [Graspable Math]. Retrieved from www.graspablemath.com

Ysseldyke, J., & Bolt, D. M. (2007). Effect of technology-enhanced continuous progress monitoring on math achievement. School Psychology Review, 36(3), 453.

3 ASSESSING VARIATION AT SCALE

The following chapter was originally published as a conference paper for the American Educational Research Association (AERA) that has since been modified for the current work. Computational fluency is composed of three components: efficiency, mathematical flexibility, and conceptual understanding. Although research highlights the importance of all three of these components in mathematical understanding, a majority of assessments primarily focus on efficiency. In this section, we propose a novel way of measuring and exploring student's mathematical flexibility using dynamic technology. We introduce Graspable Math, a dynamic mathematics notation tool, and present data from a goal-based puzzle task given to experts in mathematics. We then use the GM tools to present evidence demonstrating significant variability in problem solving approaches, efficiency, flexibility, and thinking among participants considered to be experts in mathematics.

3.1 Introduction

3.1.1 Overview

As today's society becomes increasingly intertwined in technology, mathematics is one of the most important subjects that affects future career opportunities (O'Brien, 2015). Despite how the subject bears an increasing amount of weight in today's society, mathematics is widely disliked by students and even causes a separate form of anxiety (Meece, 1990). Math invokes working memory to identify problems and implement processes and solutions that go beyond simple memory retrieval. It is for this reason that it is imperative for students to fully understand the underlying logic behind mathematics as a whole and cognitively separate it from the "listen, memorize, and repeat" retrieval of other subjects (Passolunghi and Siegel, 2001). A crucial conceptual leap for students is to transition from grounded arithmetic problems to a more abstract world of algebra problems with variables (Heffernan and Koedinger, 1998). Students with prior misconceptions of algebra and mathematics solve fewer equations correctly and have difficulty learning new procedures and problems (Booth and Koedinger, 2008). Ideally, the overall goal for teachers is to have students attain computational fluency. Fluency, in terms of a natural language, is defined as the ability to read with speed, accuracy, and proper expression. Much like native languages, mathematics has its own values in what it means to be mathematically or computationally fluent. In the *Principles and Standards for School Mathematics,* the National Council of Teachers of Mathematics (NCTM, 2000) states that computational fluency is made of 3 components: efficiency, flexibility, and conceptual understanding. In order to be successful in mathematics, students need to apply efficient and accurate methods for computing and solving problems (Russell, 2000), demonstrate flexibility in the computational methods they choose, and understand and explain the methods that they use (NCTM, 2000).

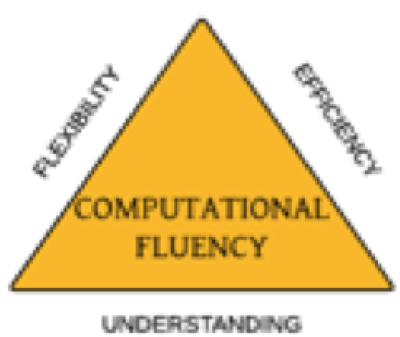


Figure 3.1. An illustration of the conceptual model of the three main components of computational fluency

All three components of this framework have been found in prior research to be necessary and effective for promoting mathematical understanding (Calhoon et al, 2007; Bass, 2003) and are often used to guide standards-based classroom practice and instruction. Building on NCTM's principles, the Common Core Standards of Mathematics emphasize the importance of not just performing mathematical procedures and algorithms, but being able to conceptually explain what they are doing and understand how and why numbers relate to each other (CCSS, 2010). While it is useful and beneficial to have students show their work and explain their thinking to understand a student's knowledge of mathematics content (Chapin, O'Connor, and Anderson, 2009), it is difficult to assess these components at scale. While asking students to show their work is often useful to identify errors or student misconceptions, one of the most daunting tasks in the effort of understanding student perspectives is the ability to see the progression of work when a student solves a problem correctly (Ysseldyke and Bolt, 2007).

While the three components of computational fluency are valued, many standardized assessments of math performance predominantly focus on efficiency, where students are often evaluated solely on whether their answer is correct or incorrect. This focus on the correct answer in assessments often brings blindness toward measuring and valuing the equally important traits of flexibility and conceptual understanding. Furthermore, under binary assessments, it is impossible to distinguish between the student that completes 99% of the problem correctly, but makes a minor mistake and a student that did not even attempt the problem. More work needs to be conducted to explore *how* we can begin to measure mathematical flexibility and begin unpack the variability of mathematical strategies when solving problems. This is the motivation of this study.

3.1.2 Graspable Math: A Dynamic Perceptual Learning Technology

Algebra represents one of the strongest cases of symbolic reasoning in human cognition (Anderson, 2007). Despite this, a considerable body of research emphasizes the importance of perceptual aspects of learning even in abstract domains like algebra (Kellman, Massey, Son, 2010; Goldstone, Landy, Son, 2010; Weitnauer, Landy, & Ottmar, 2016). When students engage in algebraic reasoning, they often attend to perceptual patterns and spatially manipulate algebraic notations before transforming such notations into abstract symbols representations (Goldstone et al, 2016).

In our prior work, we have developed Graspable Math, a web-based dynamic algebra notation system with a focus on a consistent, efficient and powerful user interface for manipulating algebraic expressions. The development of GM was guided by the concept of direct manipulation interfaces (Hutchins, Hollan, & Norman, 1985) and their advantages in terms of ease of learning and use. Although GM is a tool that is still in development, preliminary testing of earlier versions of dynamic approaches in classroom and informal learning contexts suggests that students are generally highly enthusiastic and engaged when interacting with dynamic systems (Weitnauer, Landy, & Ottmar, 2016). GM records each mouse click, trajectory, and interaction. Through these spatial transformations and fine-grained data logs, we as researchers can observe, record, analyze, and replicate the process of problem solving as it directly occurs. The rendering of this data and the use of a data visualization tool that our team built (see Figures 3.2 and 3.4) allows us to traverse through a hierarchy of pathways taken by students when solving problems.

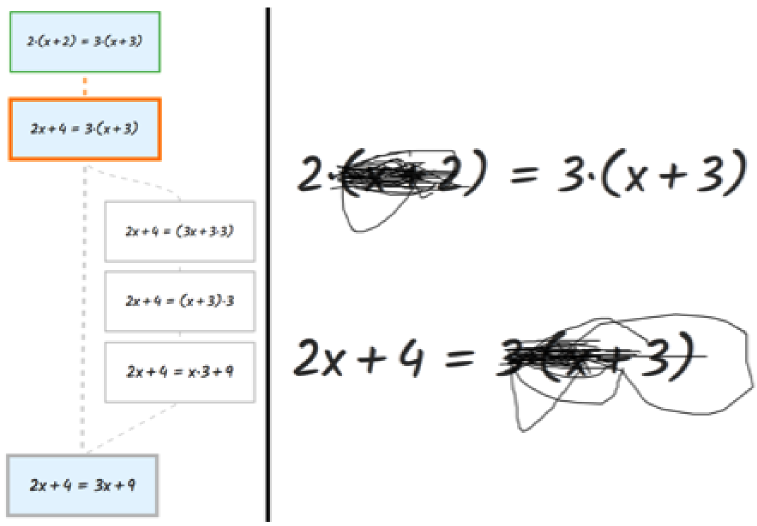


Figure 3.2. On the left, an example of the GM visualization tool that shows all solutions per group as well as the highlighted path of an individual student. On the right, the mouse trajectory recordings of students at two problem states. Note: this is not the problem given in this study but is instead meant to be a simple example to demonstrate the complex and rich data visualized in the system.

3.3 Research Goals

Given that GM and these dynamic tools are relatively new, we are just beginning to explore the potential of using dynamic technology as a vehicle to better understand mathematical flexibility at scale. The primary goal of this study is to explore the feasibility of using Graspable Math as a method of assessing problem solving strategy at scale to gain a better understanding of student perspective and mathematical flexibility. In this study, we will examine the variation in the use of problem solving strategies among a group of experts in mathematics when completing a mathematical task to better understand and reveal different components of mathematical thinking.

3.2 Methods

3.2.1 Participants

35 undergraduate students from a selective engineering and mathematics university in the northeast participated in a study assessing strategies when solving simple math problems in Graspable Math. Students received credit for a psychology course in exchange for participating in the study. Each student was trained to use Graspable Math with a 10-minute interactive tutorial that demonstrated basic interface gestures and explained the mechanics of how to solve a problem using GM. They were then asked to solve 8 math problems. The first 4 problems consisted of sample GRE math questions (to assess general math knowledge). The next 3 problems were intended as training problems and helped ensure that they could use the GM system effectively. This focus of the study was on the 8th problem presented to students. This task was set up as a *goal state problem*, where students were not asked to solve the problem, but were asked to manipulate and transform the starting equation into a specified equivalent, yet visually different state (see Figure 3.3). While the goal state approach is a relatively new approach for examining mathematical understanding, prior research in dynamic algebras has often used this approach (Ottmar, et al., 2012; 2015).

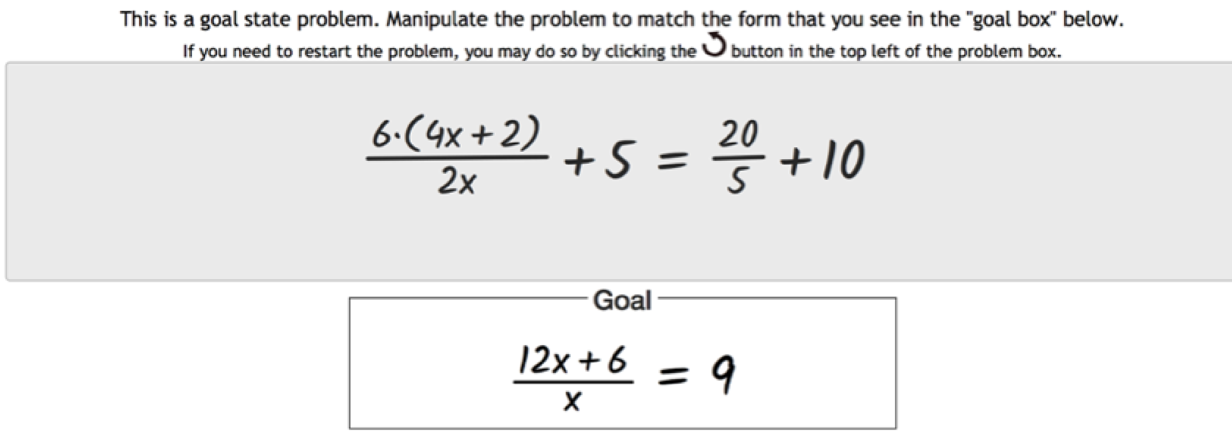


Figure 3.3. The goal state task that students were asked to complete in this study

This specific mathematical task was chosen to discriminate between student strategies and examine students' ability to identify more efficient paths to solving the problem. While students were not required to solve a problem in any specific way or with any emphasis on efficiency, there is one identifiable solution path that would get students to the goal most efficiently in 4 steps (Figure 3.4). Students were allowed to reset the problem as many times as they wanted. In addition, there was the option to skip and give up on the problem. After 2 minutes, a skip button would appear and students were told that they could give up if they wanted.

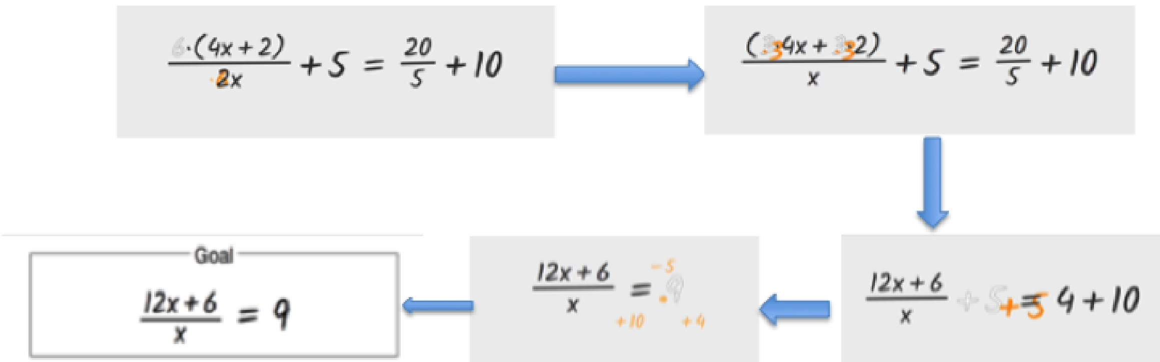


Figure 3.4. Step by step example of an efficient solution to the problem using Graspable Math. First, students can simplify the left term by canceling out the least common factor (the 2 in the denominator and the 6 in the numerator to get 3 (4x+2)/x, distribute the 3 into the parentheses to turn it into 12x+6/x, subtract both sides by 5, and then simplify and add the constants on the right side of the equation (to get 9 on the right).

3.2.2 Data Sources and Analysis

All of the data was collected using logs within the Graspable Math technology. The measures and constructs are presented in Table 3.1. Every time a student moves a term or performs an action, a step is recorded. In addition to steps, the number of attempts a student makes is also recorded. A new attempt is logged when a student hits the reset button on the canvas, which will clear all prior work and return the problem to its initial form. Total time spent solving the problem (in seconds) was also recorded. When students reset a problem, the time did not reset. We also recorded whether or not the student skipped the problem. After two minutes of attempting a problem, a skip button appears that allows students to give up.

Construct	**Description**	**Data Metrics**

Efficiency and Accuracy	Solving a problem both quickly and accurately	Did student complete the problem; Number of steps taken in a solution; amount of time taken (in seconds)
Perseverance and Effort	Self-motivation to solving a problem	Problem skips; amount of additional time spent after the initial two minutes; total number of steps taken; Number of attempts/resets
Flexibility Strategy	How students solved a problem	The visualizer shows all steps.

Table 3.1. Defined constructs and metrics used when analyzing data

A number of descriptive statistics (means and SD) were run to explore the variation in students' efficiency, perseverance, and problem solving steps. Next, to examine the problem solving pathways that people took, we used the GM data visualization tool to create an aggregated map of all the pathways and steps that people took (Figure 3.6).

3.3 Results

The results from the study are presented in Table 3.2. Although 91% of students were able to successfully complete the task, the results from the process data demonstrate enormous variability with regards to student efficiency, persistence, and strategy.

Table 3.2

Descriptives of gamified vs. non-gamified

Metric	Minimum	Maximum	Mean	Std. Dev.
Total number of steps	9	113	42.36	27.607
Number of attempts	1	5	1.97	1.085
Total amount of time	28.958 seconds	330.976 seconds	139.288 seconds	6.499 13.268
Completed	Skipped	Completed > 2 min		Mean time > 2 min
91.43% students	8.57% students	40%		80 additional secs

Table 3.2. Descriptive results from the goal state problem (n=35)

It took people on average 42 steps and two attempts to complete. Furthermore, the amount of time taken to solve this problem varied significantly, ranging from roughly 28 seconds to 330 seconds. Of the 17 who worked on the problem for longer than two minutes, only 3 skipped the problem. The remaining 14 had the option to skip, yet chose to persist on average for an additional 80 seconds. Figure 3.5 shows that although total time and total number of steps are highly correlated, the number of discrete attempts (that is, the number of times restarting the problem) is not tightly related to either.

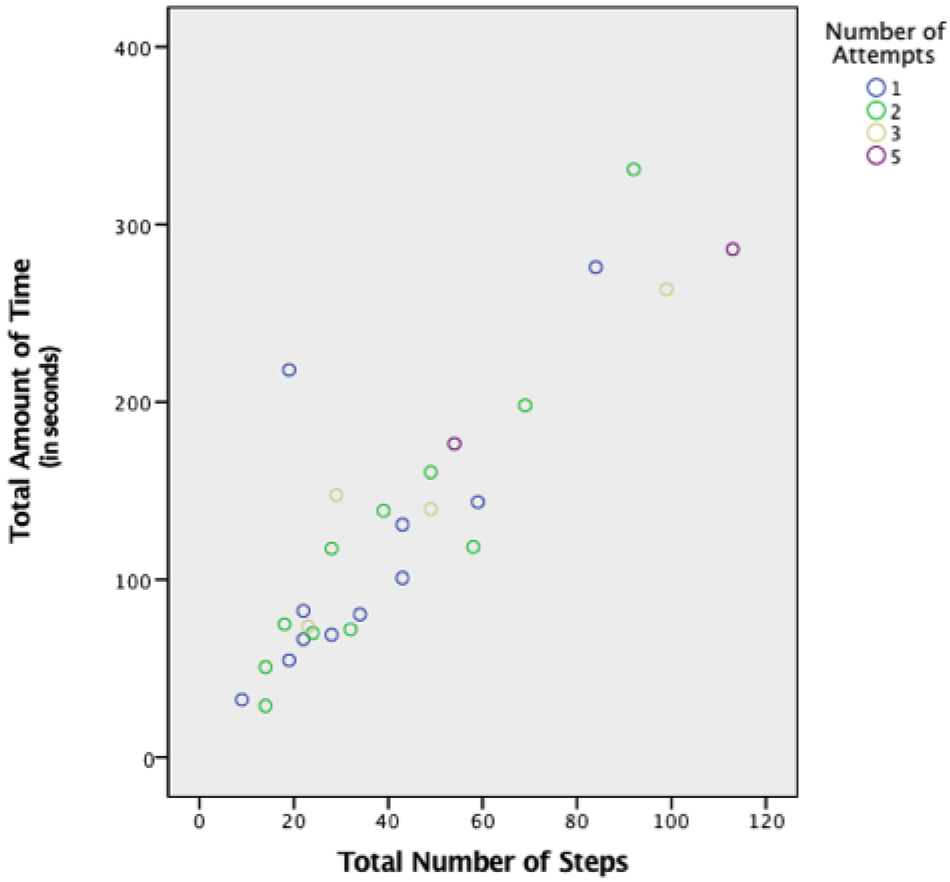


Figure 3.5. Relations between the number of steps, the amount of time to complete the problem, and the number of attempts (resets)

With regards to flexibility, we used the GM visualizer (Figure 3.6) to examine the participant's individual problem solving strategies and pathways to gain a better understanding of student perspective, flexibility, and thinking.

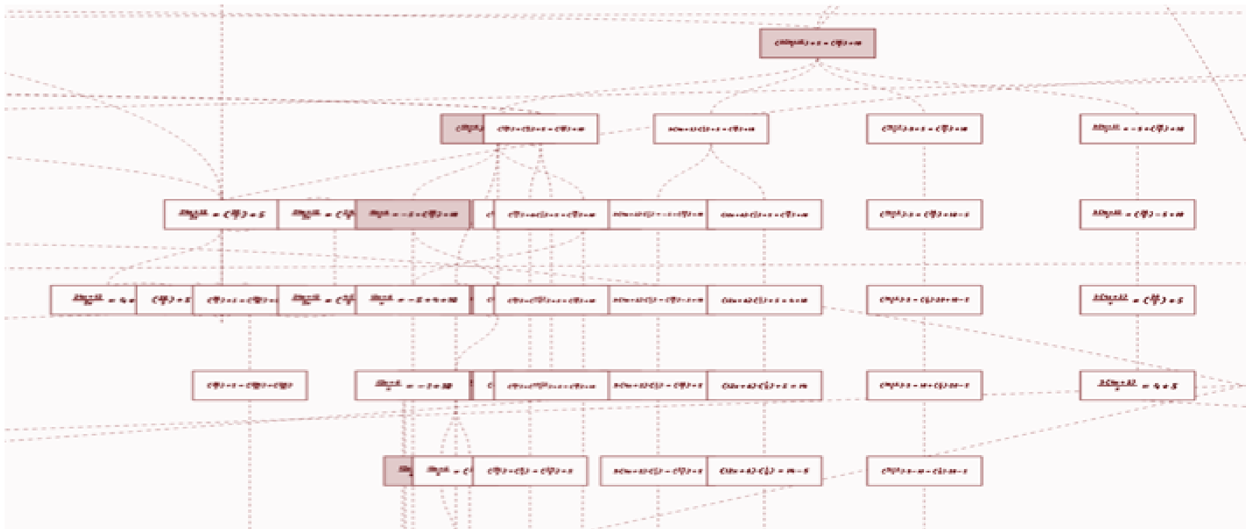


Figure 3.6. A group of nodes displaying some of the many pathways students explored when solving the problem. The shaded boxes highlight the path of a particular student. (Note: Since some people used more than 100 steps to solve the problem, it is difficult to visualize the complete sets of pathways for all participants. However, this is intended to demonstrate the incredible variability and complexity of mathematical strategy, cognition, and flexibility).

3.4 Discussion

This study demonstrates the feasibility of using dynamic technology to reveal new information about student problem solving strategies and mathematical flexibility. There was enormous variability in their approaches and only 2 people were able to complete the task in less than 10 steps. Approaches like this could be adapted to be used in the classroom by teachers to gain insight into the mathematical thinking of all students. This has enormous potential for informing teacher instruction, identifying gaps, and meeting students where they are. It could also be used as a way to promote discourse and instruction in the classroom. For example, a teacher could use the arrangement of student strategies to identify 2 or 3 unique pathways to arriving at the answer, ask those students to explain their thinking to the class, and facilitate a discussion about the different ways to approach a problem.

While GM provides several features that aid in the process of student assessment, there are some limitations that need to be addressed as well. The learning curve for mastering a new technology is difficult regardless of age or mathematical ability. Graspable math requires significant training before users are fully comfortable with the system. Furthermore, because it is a relatively new technology it currently does not support many operators that a teacher might expect from a math tool such as square roots, an intuitive decomposition function, and mixed numbers. Despite these shortcomings, we believe that the advantage the system offers far outweighs its currently unsupported features.

We recognize that this study brings up more questions than answers. Largely, it warrants future studies in mathematics education that explore the utility of goal state problems to identify gaps in student knowledge. Further, demonstrating variation in solutions in this way can help inform future studies that emphasize student strategy among struggling learners. It also suggests studies investigating why particular students are likely to reset a problem or give up. In related work, we are currently exploring eye tracking and mouse trajectory data to attempt to identify attentional anchors (Abrahamson et al, 2016) — or locations on the screen where people tend to focus their attention when solving particular types of problems. It is our hope that by using this dynamic technology as both an instructional tool and an assessment tool, we can begin to reveal aspects of mathematical cognition that have previously been invisible and nearly impossible to assess at scale. By using dynamic technology to record detailed problem-solving behavior at scale, we can begin to unpack more knowledge about the student *process.*

References

Abrahamson, D., Shayan, S., Bakker, A., & van der Schaaf, M. (2016). Eye-tracking Piaget: Capturing the emergence of attentional anchors in the coordination of proportional motor action. *Human Development*, *58*(4-5), 218-244.

Anderson, J. R. (2007). How can the human mind exist in the physical world? Oxford, England: Oxford University Press.

Bass, H. (2003). Computational fluency, algorithms, and mathematical proficiency: One mathematician's perspective. Teaching Children Mathematics, 9(6), 322.

Calhoon, M. B., Emerson, R. W., Flores, M., & Houchins, D. E. (2007). Computational fluency performance profile of high school students with mathematics disabilities. Remedial and Special Education, 28(5), 292-303.

Chapin, S. H., O'Connor, C., & Anderson, N. C. (2009). Classroom discussions: Using math talk to help students learn, Grades K-6. Math Solutions.

Common Core State Standards Initiative. Operations & Algebraic Thinking. (n.d.). Retrieved July, 2016, from http://www.corestandards.org/Math/Content/OA/

Hutchins, E. L., Hollan, J. D., & Norman, D. A. (1985). Direct manipulation interfaces. Human–Computer Interaction, 1(4), 311-338.

Kellman, P. J., Massey, C. M., & Son, J. Y. (2010). Perceptual learning modules in mathematics: Enhancing students' pattern recognition, structure extraction, and fluency. Topics in Cognitive Science, 2(2), 285-305.

Principles and standards for school mathematics. (2000). Reston, VA: National Council of Teachers of Mathematics.

R. L. Goldstone, D. H. Landy, and J. Y. Son, "The education of perception," Topics in Cognitive Science, vol. 2, no. 2, pp. 265–284, 2010.

Russell, S. J. (2000). Developing computational fluency with whole numbers. Teaching Children Mathematics, 7(3), 154.

Ysseldyke, J., & Bolt, D. M. (2007). Effect of technology-enhanced continuous progress monitoring on math achievement. School Psychology Review, 36(3), 453.

4 IMPROVING ACCESSIBILITY IN DYNAMIC TECHNOLOGY

This chapter highlights some of the differences when introducing an intervention across varying school systems. The introduction of the "computer in the classroom" has transformed the learning experience in schools. Today, computers are becoming smaller, faster, and more accessible than ever. With the introduction of groundbreaking devices such as the touchscreen tablet and smartphone, the user's interactive experience has changed significantly. Prior research has explored the variations in performance and preference of various device inputs across use-cases and users. However, there is very little work exploring the effects of input devices in learning technologies in the classroom. This chapter explores first graders' learning gains and reactions to using touchscreen tablets and computer mice in an interactive learning game. A study conducted failed to reveal significant differences between computer and iPad groups in first graders when using a dynamic math notation tool. This chapter then segues into a technical design segment that reviews the history of the formation of From Here to There! (FH2T) and explores the design and development of the web implementation of the app. The technical details of how dynamic content is loaded and unloaded for each problem as well as the iterative design cycle that ultimately led to the nuances that contribute to a more robust research tool. Finally, the structure and feasibility of its implementation is explored in a large-scale elementary study.

Part 1. Assessing Feasibility of Hybrid Alternatives in Development

4.1 Introduction

4.1.1 Overview

As technology becomes more prevalent in the classroom, a student's learning experience will ultimately revolve more around computers. While it is now generally accepted that technology in the classroom has positive effects on the classroom, it is important that we fully explore how students interact with learning technology through native or peripheral inputs. Though the basic computer mouse has reigned as the more popular form of input over the last few decades, there has been an increase in use of touchscreen devices in the classroom with the introduction and accessibility of the portable tablet, touchscreen computer, and smartphone. Research has shown that when factoring speed, accuracy, and preference, the touchscreen was the preferred method for short tasks with larger icons or graphical interfaces (Milner, 1988). The flexibility and utility of input devices vary between devices and use cases. For example, the touchscreen input involves naturally reaching out and touching the screen with one's fingers to manipulate or interact with an application and requires very little training. The computer mouse, however, often involves using an optical laser to track movement of the mouse as the user moves it across a flat surface requiring more hand-eye coordination. Prior research comparing the two found that user's proficiency with each device depended on their prior experience using computers with experts were more efficient with a mouse while beginners were more efficient with a touchscreen (Thomas & Milan, 1987). Furthermore, though estimating the performance of each input device in terms of how each will alter the student's learning experience, it is also worthwhile to examine the feasibility of both modalities in classrooms. As touchscreen devices such as iPads are still relatively new and expensive technology, the availability of such resources is predominantly limited for public schools, thus it has come into question whether or not porting to these devices are necessary or even feasible in a typical classroom. This chapter explores how students interact with a mathematical notation app that features symbolic representations and whether or not preference and learning gains may be affected when using a touchscreen versus a traditional mouse as well as the redesign of an iOS app in an effort to make learning technology more readily available to schools with a low socioeconomic status (SES).

4.1.2 From Here To There!

From Here to There! (FH2T) is an online gamified learning environment where students transform and manipulate mathematical expressions and equations from a given state to another goal state and uses perceptual based interventions to introduce foundational algebraic concepts (Ottmar et al., 2015A). Originally developed as a native app for iOS, From Here to There! consists of over 270 levels across 14 worlds that introduces gesture-based manipulations covering a variety of mathematical concepts from addition and subtraction through factoring and distribution. The spatial expression transformation engine is based on the Graspable Math library that was developed by researchers from Indiana University Bloomington. Aligned with the progression in the Common Core Standards, the problems in the game allow students to slowly increase in complexity at an individual level. From Here to There! relies on self-paced interaction through spatial transformations that engage perceptual-motor systems (Ottmar et al., 2015A). This innovative game displayed on a touch-screen interface allows both physical and dynamic manipulations of the expressions by students, providing a powerful source of perceptual-motor experiences, which in turn lead to increased acquisition of appropriate operation parameters (Ottmar et al., 2015B). From Here to There meets the criteria for a high quality mathematical program with appropriate expectations and assessments (NCTM, 2016).

4.2 Methodology

4.2.1 Study

The study included twenty-three first graders (6-7 years old; 11 males, 12 females) from one classroom in the northeast of the United States. The study utilized a between-participants experimental design, meaning that half the students were randomly assigned to play the From Here to There! app on an iPad, while the other half played the same game on a laptop computer with a mouse. Therefore, conditions were based on modality: iPad or Computer - and addressed the following goals:

- Examine improvements in mathematics performance scores after the intervention
- Determine differences, if any, in the effectiveness between iPad and computer versions

All devices used in the study were laptops and/or iPads purchased by the schools or provided by the researchers. In both conditions of the study, the From Here to There! intervention was played through either an iPad or using Google Chrome on a laptop with a mouse. Students played through the game From Here to There! for 1 week.

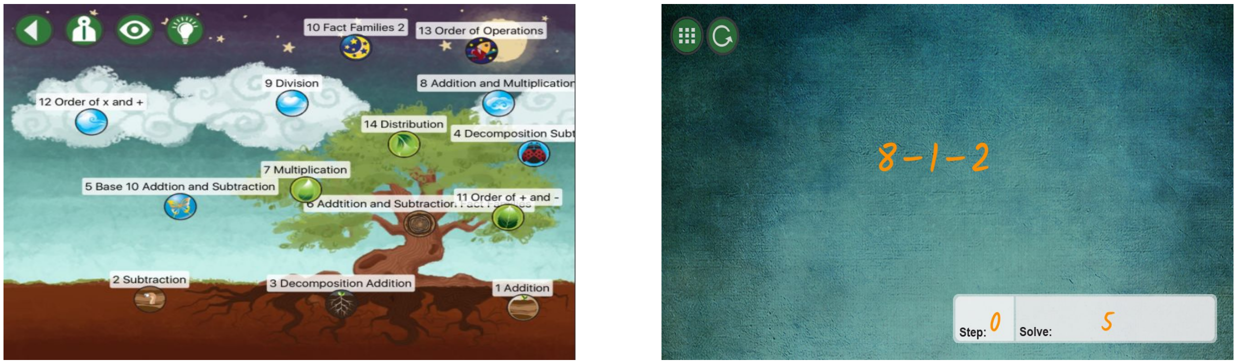


Figure 4.1. The images above display the home screen layout of the FH2T! game (left) featuring a world menu including the 14 worlds present within the game. On the right, an example of one problem where participants must simplify the expression to match the goal state in the bottom right corner.

4.2.2 Measures

Data collection for this study included a combination of student scores on pre and post-study worksheets, researcher observations of the students interacting with the game, and teacher interviews to gain perspective regarding student activity during the study. The assessments contain 15 questions and have been developed to mirror first and second grade math standards set forth by the Common Core. To obtain quantitative data relative to student's perceptions of the game and how the overall game was received within the classroom, an informal interview was conducted with the teacher, and student observations were taken by the student investigator. In the interview, the teacher was asked to describe the experience, student perceptions, strengths and weaknesses of the intervention, as well as general comments relating to the feasibility and rationale supporting the game.

4.3 Results

The FH2T program was given to 23 first grade students, all within a single classroom with a single teacher. Of the 23 first graders, 11 students used a mouse on a computer and 12 students used an iPad tablet. We examined whether there was improvement in math performance using a two-tailed t-test on the pre and post assessment mean scores. Test scores, vindicated by the total percentage of items correct, improved significantly from 63.6% to 71.8%, comparing pre- (M= 0.636, SD=.26) to post- worksheets (M= 0.718, SD= .24). There is an approximate difference of 8.1% (SD=.33, ɲp^2= 0.33, t (1) = 1.717, p = 0.05).

A linear regression was also used to examine the contributions of pretest performance, gender, and input device on mathematics performance. Pretest scores (M=0.636, SD=0.256, β=-0.189) for individuals significantly predicted posttest scores (M =0.718, SD=0.243), (β=0.607, t (22) =14.21, partial eta^2 = 0.33, R^2= 0.036, p=0.05, two-tailed test). No gender differences were observed; girls and boys performed similarly on pre/post-assessments. An independent samples t-test failed to reveal significant differences in pre and post-test assessment scores between computer and iPad groups. Overall, the intervention was feasible for implementation by the students and the teacher, both of whom offered positive comments at the end of the intervention. In addition to the quantitative in-app data collected from using FH2T, the teacher and in-class researchers made observations across the five days of using the intervention. The following observations recorded in regards to the students' experience using each input device. Some of the more interesting observations included: "Students were not used to the mouse movements and had difficulty clicking and selecting", "Computers pose the biggest challenge/Children struggle with using the mouse", "Out of frustration, kids will try to physically touch the computer screen and move numbers manually with their hands", and "Students who achieved higher levels were on iPads".

4.4 Discussion

Our pilot data suggests that FH2T! seemed to be effective in increasing learning gains during the 1-week session. There were no significant learning differences between the traditional computer mouse and touchscreen groups. Additionally, the feedback and teacher observation indicated that students overall preferred the touchscreen interface and struggled using the computer mouse. This is consistent with previous research linking the touch screen input to natural intuition, but, as mentioned earlier, does not affect learning gains and will most likely reverse as the students get older. From a technical development standpoint, this could indicate that deployment on a regular computer is just as effective as on a touchscreen device. We believe this is an important finding as the accessibility of touchscreen devices is much less than traditional computers or laptops in many classrooms. Though the initial efforts of From Here to There have been successful in implementation and feasibility, in-school scalability demonstrated a need for a non-native version of the program that could be used across multiple devices instead of strictly for iPads. Furthermore, designing for a web-based intervention, for example, that would be used with a computer mouse on a browser such as Google Chrome, would be compatible using the touchscreen input on an iPad as well in most cases. This would ultimately allow classrooms with iPads to engage in such apps using their touchscreen devices, but would also not hinder classrooms that did not have iPads readily available. We feel that this is important as it helps to balance the playing field for schools with a low SES where technology such as iPads are not available due to limited resources.

Part 2. Iterative Design of From Here to There Web!

4.5 Introduction

4.5.1 Background

From Here to There! (FH2T) is an online gamified learning environment where students transform and manipulate mathematical expressions and equations from a given state to another goal state. Originally developed as a native app for iOS, From Here to There! consists of over 270 levels across 14 worlds that introduces gesture-based manipulations covering a variety of mathematical concepts from addition and subtraction through factoring and distribution. The spatial expression transformation engine is based on the Graspable Math library that was developed by researchers from Indiana University Bloomington.

Starting in the Addition world, players are given step-by-step tutorial problems that demonstrate the gestures they must use to manipulate and transform each problem into the given goal state. Each problem has a specified beginning and end (goal) state. A problem is completed when a player successfully transforms a problem from the initial state to its goal state. Players must progress through each problem in a linear order (World 1 Problem 1, World 1 Problem 2, and so on) until they have completed the first 14 problems of the world, thus unlocking the next world. Players are then presented with a choice for continuing in the current world to complete the remaining extra problems or moving on to the newly unlocked next world. With each transformation that a student makes when solving a problem, the step counter increases until they solve the problem correctly or reset the problem. Upon completion, players are rewarded with up to 3 clovers if they can solve the problem using a limited amount of steps (best step). If a player solves the problem using as many or fewer steps than the "best step" of the problem, they are awarded three clovers. If they go over the "best step" limit by two steps or less, they are awarded two clovers. If they go over the "best step" limit by more than two steps, they are awarded one clover. Players can also go back and retry a problem for more clovers after they initially complete it (a measure referred to as "go-backs"). When attempting to solve a problem, players may reset a problem back to its original state and start over, thus also resetting the step counter back to zero. Furthemore, should a player become stuck when solving a problem, they may request hints, when applicable.

4.5.2 Original Development

The development of the native iOS iPad app of From Here to There! originated from a team of researchers collaborating with University of Richmond as well as Indiana University Bloomington. The app consisted of 14 worlds and 270+ levels of math problems for players to solve using proprietary gestures that are introduced as a player progresses through the games. In Ottmar et al. (2015), an initial pilot study was conducted to determine whether FH2T contributed to learning gains. 110 6th-8th grade students (41% male, 59% female) from six classes in a large suburban middle school participated in a 4-hour study during six of their regular math periods. A significant main effect was found for exposure: for every additional world that the students completed, their posttest accuracy scores increased by 0.76 problems (effect size=0.48). However, several difficulties became apparent when using the iOS app for research in schools. At the time of development, the infrastructure for how reporting was retrieved relied on third-party standards which unfortunately became obsolete over the years. This made any real attempt at data retrieval incredibly difficult for most studies. Furthermore, as Graspable Math became the replacement engine for the original dynamic tool used in the iOS app, the burden to readapt code and support them both became too much for the development team and ultimately resulted in the iOS app receiving very little updates and bug fixes. As mentioned in Chapter 3, a socio-economic difficulty became apparent when trying to find schools that had iPads that were compatible with the software. At the time (and even now), schools were extremely limited in their technology resources and finding schools that had 1-to-1 iPads proved to be difficult. Typically, schools had a limited number of mobile devices such as iPads or tablets, but most had computer labs, albeit rather outdated. It was then decided that the way forward was through a web app that would be able to run on most browsers and not require any type of administrative approval or installation to run.

Web App

Though the initial efforts of From Here to There have been successful in implementation and feasibility, in-school scalability demonstrated a need for a non-native version of the program that could be used across multiple devices instead of strictly for iPads. As iPads are still not readily available in most public schools, we viewed this hurdle particularly as a socio-economic barrier rather than a technological barrier. Just as it was with initiatives such as Apple's Classrooms of Tomorrow and "Computers in the Classroom", positive outcomes which are primarily correlated to exposure would only pertain to more affluent groups (Attewell & Battle, 1999; Becker & Ravitz, 1998; Milone & Salpeter, 1996). Thus, increased levels of student-to-student interaction in computer learning environments appear to provide positive levels of student achievement. Unfortunately, just as home computers were a luxury for wealthy students in the 1990's and 2000's, learning environments that depend on touchscreen iPad tablets and other modern technologies are simply not available in many schools, thus creating a socio-economic barrier in classrooms. To counteract this SES barrier, a responsive web app was developed that would be able to run on most tablets, laptops, and computers that had an Internet browser. Among other concerns, compatibility for operating systems, browsers, and devices became a top priority as many schools that we planned to run studies in were not properly equipped with state-of-the-art tablets nor had administrative access to install proprietary software or additional browsers. Designing a system that could be run in virtually any modern browser without having to install or store anything locally was necessary. Furthermore, as the overall experience of the app was crucial to its effectiveness, it was evident that it would need to be adaptive and responsive to provide as much of a similar experience as possible when being used in various environments, devices, screen sizes, input peripherals, browsers, and operating systems.

4.6 Design Techniques

4.6.1 Dynamic Content Loading

The Graspable Math library served as the dynamic notation tool that is foundational to the mechanics of *From Here to There!*. Upon the initial development of FH2T web as a research tool, it was evident that this new web app should be made to be modular in both function and appearance. In particular, various research cases arose that would require the same set of problems to be administered to participants while varying in the look and feel of the intervention and in some cases, the functions of the intervention itself. To structurally accommodate various research needs, a dynamic content loading system was implemented that would load specified cascading style sheets (CSS) as well as any required JavaScript files for runtime based on condition. Upon load of a problem, a request would be called to the server to retrieve the appropriate files to load in based on the participant's condition and the problem number. These files would then be dynamically loaded into the page before finally initializing the problem-specific data. This resulted in the ability to run specific functions and alterations in one condition without having to create duplicate pages of the same code with minor alterations for each additional condition. Furthermore, this also allowed for cosmetic changes to be applied to one condition without it being affected for other conditions. This became useful in studies featured in Part II of this dissertation when developing with gamification versus non-gamification and the leaderboard studies. Additionally, the ability to load in dynamic files was also implemented on a per-problem basis. This addition allowed for the various color schemes, backgrounds, and styles of each world to be dynamically loaded in as well as any specific functions required for a specific problem.

4.6.2 Data Logging

The iOS app version of From Here to There! proved to be difficult in data retrieval. Each iPad must be assigned a unique ID on the iPad before manually uploading data and whether or not the iPad was successful was not evident after upload. When it was successfully uploaded, the data structure was not designed according to the current research practices and was deemed unable to update by the developers due to the obsoletion of the third-party infrastructure it used. Because of this, a new data structure was developed in both Graspable Math and in the From Here to There! web app. Graspable Math implemented an event-level data logging system that reported back to the Graspable Math server. The Graspable Math data-logging system allows for custom events to be reported as well including resets, problem solves, number decomposition, etc. This data is then stored in the Graspable Math server and is able to be downloaded in JSON format.

Though the event-level would be stored when users played From Here to There!, the game itself is not using this data and requires its own logging system for in-game data. Additional data is recorded for activities in FH2T such as player progress, individual attempt measures, etc and is stored in its own server separate from the Graspable Math server. The majority of in-app data that is recorded when students use FH2T is stored on a per attempt basis. These “attempts” are recorded every time a player loads a problem and finishes when they quit or successfully complete the problem. During that attempt if a player makes any transformations such as resetting the problem (using the reset button not by reloading the page), transforms or decomposes a number, or solves the problem, summary data is recorded in one row on the FH2T server in real-time (there’s a 5 second interval that pushes the data to the server). This allows for real-time analysis of a participant's problem-solving strategy, if needed. More importantly, it records activity and attempts regardless of whether or not the user actually completes the problem.

This record contains summary data such as:

- User ID
- Problem ID
- GM Canvas ID
- GM Trial ID
- Number of clovers earned, if any
- Start time and stop time (or last updated time in event of closing window) in Unix timestamp
- Number of Steps (actions that change the equation)
- Number of Errors (actions that do not change the equation)
- The Unix timestamp that a hint was requested (if applicable)
- Was the problem completed? (1 = yes, 0 = no)

It is worth noting that FH2T data can then easily be linked to GM event-level data through the canvas ID or trial ID using a RESTful API. Thus, this infrastructure allows teachers and researchers a full spectrum of data to further analyze with both in-game app data and event-level data. The From Here to There data can be expanded into hundreds of measures across each player using the attempt data. For example, we can use aggregate student data to answer questions such as how many problems were completed in a particular world, how many worlds did a player complete, how much time is spent in a particular world, how many resets a player made overall, etc. and compare this across students, conditions, schools, and other groups.

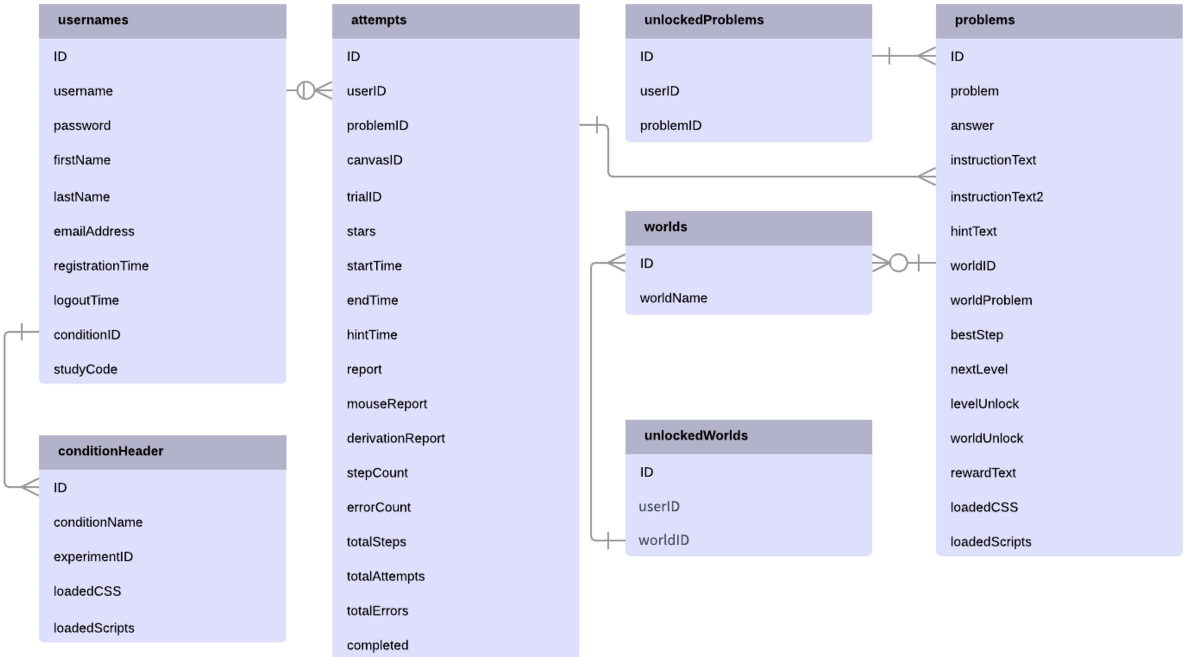


Figure 4.2. An entity-relationship diagram for From Here to There!

4.7 Feasibility

To further demonstrate the technical feasibility of the FH2T web version in a research environment, a study was conducted using a modified version From Here to There! designed for Elementary (FH2T-E) school students. Approximately 185 out of 229 students from 9 elementary classrooms participated in a 4-day study using two versions of FH2T-E, while the remaining 44 students were given traditional instruction as a control condition. Results showed feasibility success and high student engagement with the FH2T technology, as well as significant learning improvement from pre- to post-tests in relation to a traditional teaching control. Student performance improvement was observed regardless of modality usage or presence of gamelike features. This may indicate that the learning effect was due to the FH2T-E conceptual design and goal-state experience. Overall, students and teachers enjoyed the game leaving positive comments regarding fun and engagement. The goal-state and decomposition approaches embedded in FH2T may promote greater learning, engagement, and motivation. One possible explanation for these findings may be that playing with FH2T prepares students for future learning of more abstract mathematical ideas. In terms of feasibility, the study demonstrated that the web app version of From Here to There! was an effective tool for research. During the six week period that this study was conducted, the server provided quick and reliable service during peak hours without any downtime. Furthermore, the data pipeline that was established proved to be successful for deeper analysis of in-game activity. Through the updated data logging, we were able to record various measures of a user's problem solving process including the number of steps taken, the number of resets, completion rates, time between steps, overall time spent on a problem, and whether or not the user goes back to reattempt a problem. This data is able to be visualized on an individual scale to show a detailed view into event-level actions and provide timestamps between each action. Additionally, we can aggregate problem-level data on a classroom scale and observe common strategies or trends among students through an interactive visualization as described in Chapter 1. Our research team was then able to use this clickstream data in a principal components analysis (PCA) to determine 7 factors that could significantly predict learning gains between pre- and post-test scores. Significant interaction effects between Completion and Prior Knowledge and between Go-Backs and Prior Knowledge were found. In both cases, students who started with low or medium-level knowledge have an added benefit compared to high knowledge students when they complete more problems (Completion) or revisit problems to try to improve their score after they have already completed it successfully once.

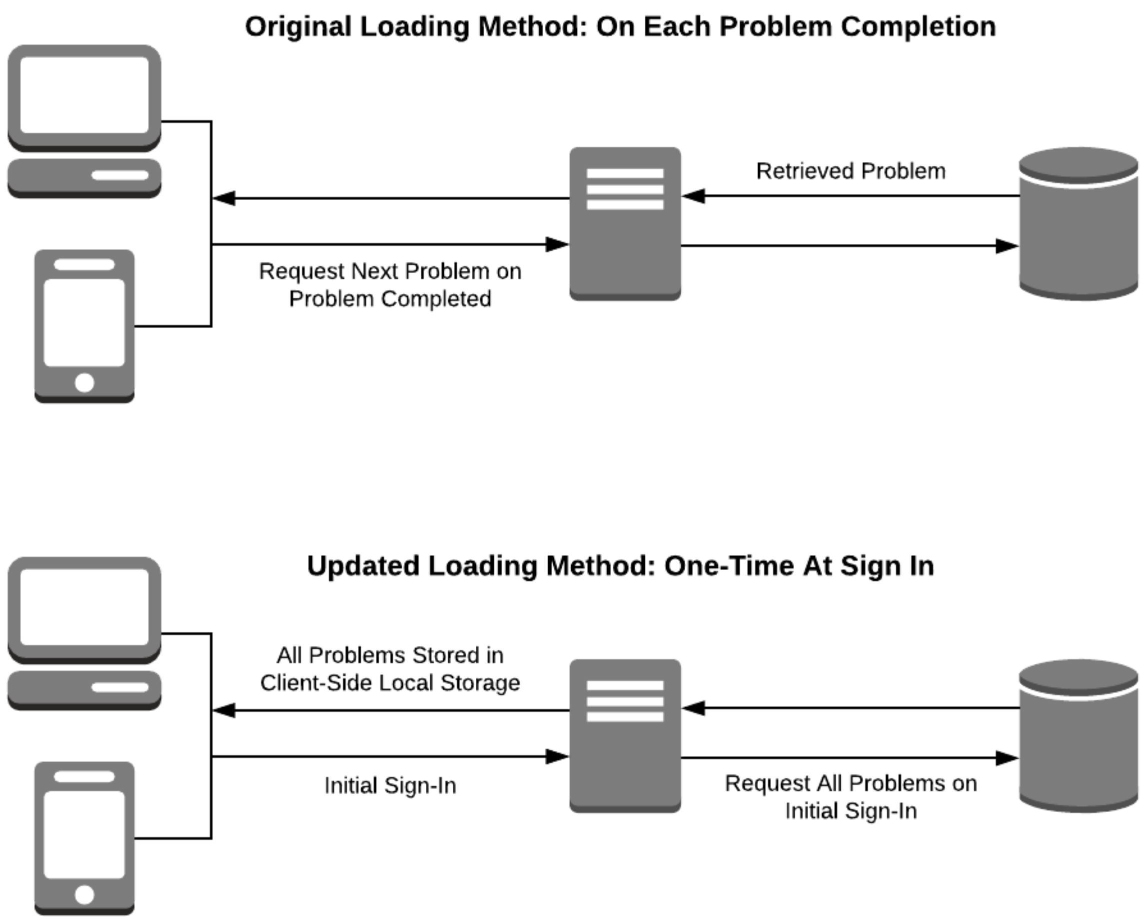


Figure 4.3. An improved workflow for loading and retrieving problems that reduces the number of calls to the server from occurring after each problem completion to only occurring once on sign-in

4.8 Discussion

Upon several iterative cycles of design, we found that there were some significant improvements that could be made to the design of the web app. Based on SES implications, it was found that many schools were not equipped with high-speed Internet connections and would thus have to wait for an extended period of time based on our current designs. To counteract this, we implemented two major changes to the problem loading process. The first change affected how the next problem would load into a page. Our initial design loaded one problem per page and would simply redirect the user to a new page with the next problem upon completion. We improved on this design by dynamically unloading each problem after completion and then retrieving and loading the next problems information into the originally loaded page. This allows for much fewer calls to the server as well as eliminates the need for constantly reloading the same page over and over again when a user completes the problem. This results in faster loading times and ultimately an overall smoother experience, especially when running on a computer with a suboptimal Internet connection. With these new designs and optimizations, we feel that From Here to There! will be much more accessible when based in a web-browser, but also breaks through the socio-economic barriers that were in place when it was solely designed as an iOS app due to the limited availability of iPads.

References

Battle, P. A. J. (1999). Home computers and school performance. *The information society*, *15*(1), 1-10.

Becker, H. J., & Ravitz, J. L. (1998). The equity threat of promising innovations: Pioneering internet-connected schools. *Journal of Educational Computing Research*, *19*(1), 1-26.

Milner, N. (1988). A review of human performance and preferences with different input devices to computer systems. In D. M. Jones & R. Winder (Eds.), People and computers IV. Proceedings of human-computer interaction ’88 (pp. 341-362). Cambridge, UK: Cambridge University Press.

Milone Jr, M. N., & Salpeter, J. (1996). Technology and equity issues. *Technology & Learning*, *16*(4).

Thomas, C., & Milan, S. (1987). Which input device should be used with interactive video?. In *Human–Computer Interaction–INTERACT'87* (pp. 587-592).

National Council Teachers of Mathematics (NCTM) (2016). https://www.nctm.org

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., Weitnauer, E., Landy, D., & Goldstone, R. (2015). Graspable mathematics: Using perceptual learning technology. *Integrating touch-enabled and mobile devices into contemporary mathematics education*, *24*.

PART 2: GAMES IN EDUCATION

5 GAMES IN EDUCATION, GAMIFICATION, AND THE OCTALYSIS FRAMEWORK

The following chapter provides a literature review of games as it relates to culture and education. The chapter then segues to gamification and introduces the Octalysis framework, an eight-part framework that maps various aspects of gamification. This chapter focuses on the fundamental characteristics of games and how they have and continue to motivate people to perform difficult or less desirable tasks in all aspects of life. Furthermore, this chapter will discuss various attributes of games and game-like features that can then be used as an augmentation for educational instruction.

5.1 Background

Video games have become a key element in today's youth culture (Aarsand, 2007; Gee, 2007). The current generation of children is raised on the immediate feedback and immersive, adaptive environments found in video games and other technologies. Games are more popular now than they have ever been, and the video game industry is on the verge of hitting its Golden Age (Diver, 2015). In 2015 alone, the video game industry was predicted to reach nearly $100 billion in sales revenue (Nunnely, 2015) and has been increasing each year. With high yields as this, some may question why the gaming industry continues to outperform with each passing year. An obvious answer is that as technology improves, so does the quality and quantity of gaming systems, peripherals, and titles. Some experts speculate that the rise in popularity is also due to the shift in culture from generation to generation. In Beck and Wade's book *The Kids Are Alright: How the Gamer Generation is Changing the Workplace* (2006), they argue that because the millennial generation has grown up with technology that it has become a central part of their lives (Beck & Wade, 2006). Research suggests that video games can motivate and interest learners (Dempsey et al., 1994; Jacobs & Dempsey, 1993; Lepper & Malone, 1987; Malone, 1981, 1983; Malone & Lepper, 1987; Malouf, 1988), increase retention rates (Dempsey et al., 1994; Jacobs & Dempsey, 1993; Pierfy, 1977), and improve reasoning skills and higher-order thinking (Mayland, 1990; Rieber, 1996; Wood & Stewart, 1987; Hogle, 1996). Despite this success, educational video games are often not embraced as willingly as mainstream entertainment video games. Like other mediums such as film, video games have the ability to immerse the end-user into an environment and stimulate thought. Films and videogames have generated fanatic obsessions and studious efforts that sometimes supersede a normal viewing or interacting experience. From a research perspective, it can be perplexing to understand why some people are more motivated to learn the history of the kingdom of Gondor from Lord of the Rings, yet difficult for that same person to feel motivated to learn about scholarly topics such as the American Revolutionary War or the United States Constitution. Experts in topics such as immersion and flow suggest that this has to do with the captivating environment and storylines that are experienced while exploring each microcosm (Brown and Cairns, 2004) as well as the belongingness of the community of fans that evolves around such franchises. Once engaged, end-users may sometimes become fanatics and take it upon themselves to explore the given work in great detail. The end-result is a large community or following that has fully embraced and digested everything that the author has administered. Furthermore, the fandom of such franchises expands with highly-attended social events such as Comic Con and E3 where hundreds of thousands of fans gather from all over the world to dress in cosplay, discuss theories and canon, and overall socialize with others that have common interests. These events are often considered as the biggest event of the year for fans and ultimately highlight the importance of a well-established community. Such gatherings further demonstrate a sense of community discussed in Sarason (1977) and expanded on in McMillan and Chavis (1986). Sarason outlines the psychological sense of community (PSOC) as a person's sense of being part of a group of people with whom they have something in common (Sarason, 1977). McMillan and Chavis (1986) expanded on the idea of PSOC hypothesizing four dimensions: belonging, fulfillment of needs, influence, and shared connections. This sense of community is clearly a driving force to many to stay engaged in these subcultural environments. Obst, Zinkiewicz, & Smith (2002) suggest that PSOC can be a strong facet of communities of interest. This may be because members choose to belong to such communities and are drawn together for a common interest. Another interesting finding of Obst, Zinkiewicz, & Smith (2002) is that face-to-face interaction is not essential for PSOC as there were no significant differences of PSOC between participants with contact over text-to-text versus face-to-face. Thus, strong PSOC can exist regardless of geo-proximity, thus rendering online communities as a valid source of PSOC.

The idea of engaged collaboration and dedicated sense of community is what some educators and researchers aspire to see in education. However, attaining that goal often requires appealing to the common threads of such an experience. Unlike movies and television which remain constant regardless of any participation from the user, video games allow and often require interaction from the user to direct and guide the experience. In this regard, users must be active and processing each stage of their experience as they do when playing traditional games or board games. Video games have the unique advantage of being able to then process the users' decisions and adapt the experience accordingly. While there is often little narrative or overarching storyline to grab onto in our projects as they would be in film, we try to captivate them using similar techniques so that they feel that there is more to their given activity than just solving math problems. In our work, we use interactive puzzles and video games to hopefully achieve the same type of piqued interest. While there may never be as large of a following for educational games as there would be in non-educational works, there are components. Some of these techniques as well as many examples of successful and unsuccessful implementations in educational environments are covered in this chapter.

5.2. History of Educational Games

5.2.1 "Chocolate Covered Broccoli"

The concept of having a bridge between video games and education has been around since the 1970s. Even early games such as *Lemonade Stand* in 1979 taught real-world skills and lessons about business and economics. While the game itself seems to have a trivial appearance, the critical thinking that goes into decision making and planning is all but trivial. One of the most well known video games of all time is *Tetris*. Developed in 1984, the goal of the game is to arrange the falling geometric block patterns in complete rows to gain a high score. The patterns of the geometric shapes are randomly picked and often need to be rotated properly to complete a row. While *Tetris* has been named one of the greatest games of all time, *Tetris* has also been found to have some real-world benefits. Studies show that students who played *Tetris* improved performance on tasks associated with spatial visualization and two-dimensional mental rotation (Okagaki and Frensch, 1994; De Lisi and Wolford, 2002).

Figure 5.1. An example of a successful educational game, Atari's Lemonade Tycoon required players to keep track of sales, inventory, and customer feedback while running their own business.

Although there have been games that have successfully designed educational games with the right balance of both entertainment and education (otherwise known as edutainment), many games have not been able to master this balance and are otherwise referred to as “chocolate covered broccoli”. Coined in 2001 by Brenda Laurel, Chocolate covered broccoli is a colloquial phrase in the area of research involving games for learning and refers to when an educational topic or concept attempts to make it more appealing by wrapping it in a gamified context, but fails to actually make it more palatable for the student. Chocolate covered broccoli occurs most often when “games” are successful at educating the students, but not entertaining them. Though there has been a strong push for education-based games, the hurdle of chocolate covered broccoli may be difficult to overcome for several reasons. More often than not, the teams behind such educational games are educators or researchers that are experts in their instructional field. While the pedagogical aspect of these games tend to be high quality, the game design is sometimes an afterthought or poorly executed by programmers who may not be familiar with modern game design theory. Conversely, a push for “serious games” in the computer science field has yielded very entertaining and high quality games, but in actuality contains only watered down, surface-level pedagogical content and is based very little on any didactic theory. In this sense, the experience is not so much “chocolate covered broccoli” as it is “broccoli flavored chocolate”. Thus, balancing the educational and entertainment levels of an educational game is very difficult and requires high-level collaboration from both pedagogical experts as well as knowledgeable game-designers to achieve excellent results.

In the early 2000's, researchers began designing more robust educational video games that would both apply direct research and engage students in real-world studies in unique ways. *Immune Attack*, for example, is a first-person serious game that teaches complex biology and immunology to students through a video game. It is funded by the National Science Foundation and jointly developed by the FAS, the University of Southern California, Brown University, and Escape Hatch Entertainment. In *Immune Attack*, players interact with objects to train the body's immune system how to function properly to prevent it from dying. As the game progresses, more biological threats and pathogens attack the body and the player must identify the threat and teach the body to handle it accordingly (FAS, 2006). While players use strategies and critical thinking that is present in other video games, they are learning more about biology and putting their theories into practice in a relevant and fun way. Examples like these demonstrate that engaging educational games are possible to create, however, they require expertise in a plethora of areas including domain-specific knowledge, pedagogical theory, and game design theory and are unfortunately significantly outnumbered by a world of "chocolate covered broccoli".

5.3 The Octalysis Framework

5.3.1 Overview

While serious games are beneficial, the amount of work that goes into developing an engaging yet educational experience makes it difficult to scale. Serious games typically require building an engaging game first and then adding the more pedagogical content afterward. Gamification, the use of game elements in a non-game context, is a popular alternative that is being used more in the classroom (Deterding, Sicart, Nacke, O'Hara, & Dixon, 2011).

Gamification occurs when ordinary tasks incorporate game-elements such as engaging contexts and themes, peer competitions, or incentivization. Since gamification is an addition to and not a replacement of the task or activity, the gamified elements are often adapted to the original pedagogical curriculum. Ragupathi (2012) highlighted some of the key components of using gamification in an educational setting using a system called JFDI Academy. The platform allows students to complete typical course assignments, but incorporates gamified elements such as experience points, leveling up, and leaderboards. Prior to using the platform with gamified elements, most students completed only the minimum required amount of assignments, attempted few to no practice problems, and submitted their assignments only a few hours before the deadline (Ragupathi, 2012).

It is crucial to our development and design to build the foundation of our research against a human-centered design (HCD). According to ISO 9241-210:2010 standards, human-centered design is an approach to interactive systems development that aims to make systems usable and useful by focusing on the users, their needs and requirements, and by applying human factors/ergonomics, usability knowledge, and techniques. Several frameworks have emerged that attempt to map the various components on the spectrum of gamification using human-centered design, one of those being the Octalysis framework. We have applied the Octalysis framework to our work as it elegantly maps dozens of gamification components along eight core drives for human motivation. Developed by Yu-kai Chou, the Octalysis framework was developed over the past 10 years and has been implemented as a motivational framework by companies such as Google, Lego, and Tesla. A Stanford University lecturer, Yu-kai Chou was an early pioneer in gamification research and behavioral design as well as the author of the book *Actionable Gamification: Beyond Points, Badges, and Leaderboards*. The Octalysis framework is the byproduct of Chou's research in motivation and gamification. According to Chou's book *Actionable Gamification: Beyond Points, Badges, and Leaderboards*, the Octalysis framework is comprised of eight core drives: epic meaning and calling, development and accomplishment, empowerment of creativity and feedback, ownership and possession, social influence and relatedness, scarcity and impatience, unpredictability and curiosity, and loss and avoidance.

5.3.2 Epic Meaning and Calling

Epic meaning and calling is the core drive that is in action when a person is motivated to do something that they believe has a greater purpose or that they are performing a task that they were "chosen" to perform. Two examples of this drive are Wikipedia and Waze. Wikipedia editors are unpaid volunteers that dedicate hours of their time in editing and reviewing changes and additions to Wikipedia articles. This motivation to spend their free time moderating Wikipedia content without any compensation indicates their belief in the power of knowledge and its need of protection. Waze is an Israeli-based GPS app that collects user data to formulate and update their traffic patterns and directions based on the crowdsourced data it acquires. In addition to using the collected GPS data of users, Waze also allows users to report accidents, traffic jams, speed and police traps. The user-reported data is extremely accurate and allows users to also thumbs-up or thumbs-down any user-reported data as another form of crowdsourced data. This also exemplifies the idea of the epic meaning and calling core drive as it further promotes a better virtual environment where drivers are well-informed of any events during their commute. Despite there being no extrinsic benefit to reporting for a user, users consistently report information as an act of benevolence to other drivers. In terms of education, this core drive can be applied through clearly explaining *why* their participation and efforts matter. This is often accomplished in educational and non-educational video games alike through the power of literary narratives that compel the player to reach their goal or destiny for altruistic purposes. In the classic video game *The Oregon Trail*, players take on the role of a wagon leader that must guide a party of settlers through Missouri to Oregon in a covered wagon. This call-to-action gives meaning to players' performances and ultimately deals them a great responsibility in keeping the settlers alive (despite the inevitable dysentery-related deaths). *Development and accomplishment* is the core drive that focuses on making progress, developing skills, achieving mastery, and ultimately overcoming challenges. This is also the core drive that naturally motivates through its design and is the drive in focus in our research on leaderboards (see Part II, Chapter 3). The concept of rewards through mastery was incorporated in the original additions to From Here to There! in the form of starbursts. Players of the original FH2T! were encouraged to complete the remaining problems of each world in order to learn a fun math fact and earn a golden starburst ring around the world icon on the main screen. *Social influence and relatedness* incorporates motivating factors such as competition, social acceptance, and companionship. In addition to influencing academic performance (Christy and Fox, 2014), there is also an overwhelming amount of pressure in social comparisons to either maintain a high standing or ascend to an acceptable standing (Wells & Skowronski, 2012). *Scarcity and impatience* is the core drive of wanting something simply because it is extremely rare, exclusive, or immediately unattainable such as turn-based strategies and timed operations. When design prolongs or prevents users from getting or accessing something immediately, they are often motivated to return to the product more often. This drive is implemented often in turn-based games such as Farmville that produce countdown timers before players are able to harvest crops or access new areas. Furthermore, players often pay real money to get around imposed time limits.

5.3.3 Creativity and Feedback

Empowerment of creativity and feedback is expressed when users are engaged in a creative process where they repeatedly figure new things out and try different combinations. People not only need ways to express their creativity, but they need to see the results of their creativity, receive feedback, and adjust in turn. This is why playing with Legos and the Sims is intrinsically fun. Parker and Lepper (1992) suggests that students are more likely to engage with a learning environment if they are able to be creative with their activity and personalize their learning experience. *Ownership and possession* is the core drive that deals with users customizing and controlling their experience. In most technologies, this is implemented through avatars, profiles, and signatures. This also occurs in games when players can accumulate virtual wealth to purchase goods to further customize their playable characters or homes. *Unpredictability* is the core drive focusing on motivation through surprise and chance. When users are unaware of the outcome to their actions, they often pay more attention to the unexpected. The Skinner box experiments, where an animal presses a lever in a box frequently for unpredictable results, are exclusively referring to the core drive of unpredictability. Though this drive focuses on uncertainty in outcome, it has also been misunderstood as the drive behind points, badges, and leaderboard mechanics in general. *Loss and avoidance* is the motivation to avoid an unwanted outcome from occurring. Some examples of such negative outcomes include losing all of your work/progress, having to do additional work, or feeling compelled to act on a limited offer in fear of missing out after it expires. When brought together, these eight core drives comprise an octogonal map of motivating factors that can be applied to various processes and projects. This framework can be further illustrated to highlight how strong or weak a particular project is in any or all of the core drives of the Octalysis framework, thus providing a clearer representation of the process or project and how balanced it is overall according to the framework. Figure 5.2 demonstrates the conforming illustration of applying the Octalysis framework to Twitter.

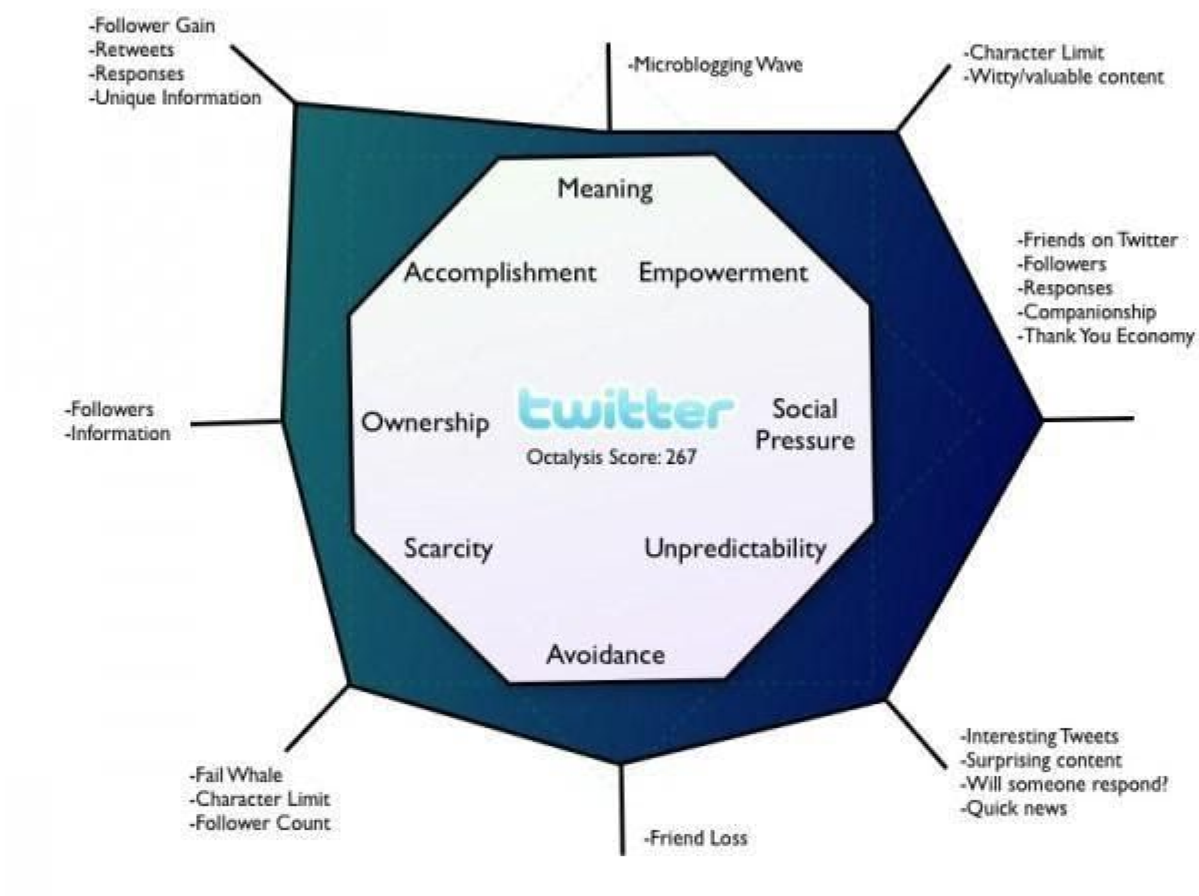


Figure 5.2. An example of the Octalysis framework being applied to the social media platform Twitter

Though research can be done to fully explore the positive and negative effects of each of the core drives in tandem to education and learning gains, our research projects discussed in Part II focus mainly on *epic meaning and calling, development and accomplishment, social influence and relatedness,* and *scarcity and avoidance.*

5.4 Discussion

Though gamification is relatively new as an academic area of research, it has been a hot topic in both education and business. The idea of an easy way to motivate students to be more engaged in the classroom seems appealing, and though a review of over 20 peer-reviewed papers on gamification suggest that there are positive results in the outcome of using gamified interventions overall, there were some limitations in the studies themselves which suggest a need to be revisited. Some of these limitations include a small sample-size, lack of transferability especially in an educational setting, lack of clarity due to multiple affordances being investigated simultaneously, and results relying solely on user observation or a lack of control condition (Hamari, Koivisto, & Sarsa, 2014). Furthermore, although technology and intelligent tutoring systems have become more predominant in the classroom as a primary method of instruction, a secondary tool to assist in teaching curricula, and as a take-home tool to assist with homework, there has been less integration with gamification and gamified elements in these tools. Though several studies show value for gamified components such as points, badges, and leaderboards in non-game contexts, there are very few ITS that actually implement these types of motivational features. Ideally, studies similar those conducted in Chapters 2-5 would encourage intelligent tutors to incorporate minor gamification elements just to further enhance the user experience particularly for younger students. As featured in Chapter 3, some of these features do not necessarily need to demonstrate true competition between peers, yet use the power of social competition to further engage less motivated students. The intended outcome for our work is to ultimately help motivate students to want to learn and understand mathematics. While this is not an easy task, we find that adding gamification and gamified elements to our interventions may alleviate some of the more distasteful aspects of math education at least until they fully understand the value of mathematics themselves. In Ottmar et al. (2015), we found that learning gains when using the From Here to There! app were positively linked to exposure, thus the more worlds that a user played through the higher they would perform on the posttest. Additionally, other research also suggests that increased exposure to STEM initiatives and activities positively impacts perceptions and dispositions in mathematics (Bybee, & Fuchs, 2006). Thus among other use cases, we are exploring gamification and gamified elements as a means to increase prolonged exposure to math-related activities to pique interest and motivate students to further explore mathematics.

References

Aarsand, P. A. (2007). Computer and video games in family life: The digital divide as a resource in intergenerational interactions. *Childhood*, *14*(2), 235-256.

Beck, J. C., & Wade, M. (2006). *The kids are alright: How the gamer generation is changing the workplace*. Harvard Business Press.

Brown, E., & Cairns, P. (2004). A grounded investigation of game immersion. In *CHI'04 extended abstracts on Human factors in computing systems* (pp. 1297-1300). ACM.

Bybee, R. W., & Fuchs, B. (2006). Preparing the 21st century workforce: A new reform in science and technology education. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *43*(4), 349-352.

Chou, Y. K. (2015). *Actionable gamification: Beyond points, badges, and leaderboards*. Octalysis Group.

Christy, K. R., & Fox, J. (2014). Leaderboards in a virtual classroom: A test of stereotype threat and social comparison explanations for women's math performance. *Computers & Education*, *78*, 66-77.

De Lisi, R., & Wolford, J. L. (2002). Improving children's mental rotation accuracy with computer game playing. *The Journal of genetic psychology*, *163*(3), 272-282.

Dempsey, J., Lucassen, B., Gilley, W., & Rasmussen, K. (1993). Since Malone's theory of intrinsically motivating instruction: What's the score in the gaming literature? Journal of Educational Technology Systems, 22(2), 173-183.

Dempsey, J. V., Rasmussen, K., & Lucassen, B. (1994, February). Instructional gaming: Implications for instructional technology. Paper presented at the Annual Meeting of Association for Educational Communications and Technology, Nashville.

Deterding, S., Sicart, M., Nacke, L., O'Hara, K., & Dixon, D. (2011, May). Gamification. using game-design elements in non-gaming contexts. In *CHI'11 extended abstracts on human factors in computing systems* (pp. 2425-2428). ACM.

Diver, Mike. "We're in the New Golden Age of Video Games." *VICE*. VICE Media, 8 Jan. 2015. Web. 21 Apr. 2015.

Federation of American Scientists. (2006). Summit on educational games: Harnessing the power of video games for learning.

Gee, J. P. (2007). *Good video games+ good learning: Collected essays on video games, learning, and literacy* (Vol. 27). Peter Lang.

Hamari, J., Koivisto, J., & Sarsa, H. (2014, January). Does gamification work?--a literature review of empirical studies on gamification. In *2014 47th Hawaii international conference on system sciences (HICSS)* (pp. 3025-3034). IEEE.

Hogle, Jan G. "Considering Games as Cognitive Tools: In Search of Effective" Edutainment."." (1996).

Jacobs, J. W., & Dempsey, J. V. (1993). Simulation and gaming: Fidelity, feedback, and motivation. In J. V. Dempsey & G. C. Sales (Eds.), Interactive instruction and feedback (pp. 197-227). Englewood Hills, NJ: Educational Technology Publications.

Malone, T. (1980) . What makes things fun to learn? A study of intrinsically motivating computer games. Unpublished dissertation, Stanford University.

Malone, T. (1983). Guidelines for designing educational computer programs. Childhood Education, 59(4), 241-47.

Malone, T., & Lepper, M. (1987) . Making learning fun: A taxonomy of intrinsic motivations for learning. In R. Snow & M. Farr (Eds.), Aptitude, learning, and instruction (Vol. 3, pp. 223-253) . Hillsdale, NJ: Earlbaum.

Malouf, D. B. (1988) . The effect of instructional computer games on continuing student motivation. Journal of Special Education, 21(4), 27- 38.

Mayland, V. (1990) . Carmen Sandiego is in your classroom. Miami, Florida: Dade Public Education Fund.

McMillan, D. W., & Chavis, D. M. (1986). Sense of community: A definition and theory. *Journal of community psychology*, *14*(1), 6-23.

Nunneley, Stephany. "Global Industry Analysts Predicts Gaming Market to Reach $91 Billion by 2015." VG247 Global Industry Analysts Predicts Gaming Market to Reach 91 Billion by 2015 Comments. Global Industry Analysts, 23 June 2009. Web. 21 Apr. 2015.

Parker, L. E., & Lepper, M. R. (1992). Effects of fantasy contexts on children's learning and motivation: Making learning more fun. *Journal of personality and social psychology*, *62*(4), 625.

Pierfy, D. A. (1977) . Comparative simulation game research: Stumbling blocks and stepping stones. Simulation and games, 8(2), 255-268.

Obst, P., Smith, S. G., & Zinkiewicz, L. (2002). An exploration of sense of community, Part 3: Dimensions and predictors of psychological sense of community in geographical communities. *Journal of community psychology*, *30*(1), 119-133.

Okagaki, L., & Frensch, P. A. (1994). Effects of video game playing on measures of spatial performance: Gender effects in late adolescence. *Journal of applied developmental psychology*, *15*(1), 33-58.

Ragupathi, K. (2012). Gamification: How to do it Right and Why it is No Good. *Technology in Pedagogy series*, 4-5.

Rieber, L. P. (1996). Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games. *Educational technology research and development*, *44*(2), 43-58.

Sarason, S. B. (1977). *Work, aging, and social change: Professionals and the one life-one career imperative*. Free Press.

Wells, B. M., & Skowronski, J. J. (2012). Evidence of choking under pressure on the PGA tour. *Basic and Applied Social Psychology*, *34*(2), 175-182.

Wood, L. E., & Stewart, P. W. (1987). Improvement of practical reasoning skills with a computer game. Journal of Computer-Based Instruction, 14(2), 49-53.

6 STUDY 1: GAMIFICATION IN EDUCATIONAL LEARNING ENVIRONMENTS

The present work explores student engagement when using a learning technology with gamified elements. A study was conducted examining engagement between gamified and non-gamified versions of a game-based perceptual learning intervention. One group was given a dynamic technology with gamification elements while the other was given the same program with a plain design and no gamification elements. Findings suggest that the students who were in the gamification condition spent more time engaging with the technology and made more attempts than the non-gamified condition regardless of their reported prior interest of engagement in math games and puzzles. Further, findings suggest that high interest learners in the gamified condition spent the most time engaging in the intervention, while high interest learners in the non-gamified condition spent the least time engaged, even when compared to low interest learners.

6.1 Introduction

6.1.1 Overview

"A spoonful of sugar helps the medicine go down", this infamous approach used by Mary Poppins motivated children to perform less desirable tasks. While it might be more ideal to have the children take their medicine without needing a spoonful of sugar, this idea of tempting children with something they enjoy certainly is effective in motivating them to do something they are not interested in such as taking medicine (or doing homework). A goal for most teachers and researchers is to not only teach a student so that they can learn the material, but also motivate them to care about the material they are learning and understand its value. Durik &

Harackiewicz (2007) demonstrates the concept of “catch and hold” as a method that “entices” students in mathematics and then once they are in the door, demonstrates the utility of mathematics in an effort to keep them engaged and see the intrinsic value. Their findings suggest that the engagement levels of students with a higher interest in mathematics were either not affected or negatively affected when presented with a mathematical notebook filled with cartoons and colorful images, while students with a lower interest in mathematics were more engaged and positively affected by the colorful notebook. Mitchell (1993) also suggests that explaining the value or utility of an educational concept piques interest. Mitchell argues that teachers must “sell” an interest to their students in order for them to “buy” into the idea of being interested in mathematics. This can be done through various approaches such as lab activities that involve mathematical procedures or more culturally relevant mediums such as video games.

As video games have become rapidly popular and are quickly to be engaged by young students, games can be used as the “sugar pill of learning” (Falstein, 2005; Warren 2009). Research suggests that video games can motivate and interest learners (Dempsey et al., 1994; Jacobs & Dempsey, 1993; Lepper & Malone, 1987; Malone, 1981, 1983; Malone & Lepper, 1987; Malouf, 1988), increase retention rates (Dempsey et al., 1994; Jacobs & Dempsey, 1993; Pierfy, 1977), and improve reasoning skills and higher-order thinking (Mayland, 1990; Rieber, 1996; Wood & Stewart, 1987; Hogle, 1996). “Serious games”, video games designed with a purpose other than pure entertainment, are becoming popular additions to instruction. However, they are often used in addition to classroom curricula and not used solely in classrooms. Research has shown that more children paid attention in class while using a serious game tool than while not using it (Rosas et al., 2003). Similar studies have also found that serious games can be used in a formal learning environment to learn about non-mathematical subjects such as geography. These studies showed statistically significant student learning gains when learning about world continents and countries through a video game (Tüzün et al., 2009). Thus, when educational games can efficiently engage students through its fun nature, the prolonged exposure to the pedagogical content may often lead to increased learning gains (Ottmar et al., 2015).

While serious games are beneficial, the amount of work that goes into developing an engaging yet educational experience makes it difficult to scale. Serious games typically require building an engaging game first and then adding the more pedagogical content afterward. Gamification, the use of game elements in a non-game context, is a popular alternative that is being used more in the classroom (Deterding, Sicart, Nacke, O'Hara, & Dixon, 2011). Gamification occurs when ordinary tasks incorporate game-elements such as engaging contexts and themes, peer competitions, or incentivization. Since gamification is an addition to and not a replacement of the task or activity, the gamified elements are often adapted to the original pedagogical curriculum. Ragupathi (2012) highlighted some of the key components of using gamification in an educational setting using a system called JFDI Academy. The platform allows students to complete typical course assignments, while incorporating gamified elements such as experience points, leveling up, and leaderboards. Prior to using the platform with gamified elements, most students completed only the minimum required amount of assignments, attempted few-to-no practice problems, and submitted their assignments only a few hours before the deadline. After incorporating the gamified elements, students completed more optional practice assignments and the average submission time increased to more than 2 days before the deadline (Ragupathi, 2012). While these studies observe the effects of various components of gamification combined, there are very few studies that examine the effects of a single-component such as rewards or points in a learning environment against a controlled condition (Hamari, Koivisto, & Sarsa, 2014).

6.1.2 Research Goals

We conducted a study using a similar methodology as Durik and Harackiewicz (2007) for this chapter with a focus on motivation and engagement in a puzzle-based mathematics learning environment. In this study, we observe engagement using a dynamic learning technology, where participants are given either gamified or non-gamified interventions, to answer the following research questions:

4) Do gamification elements relate to indicators of higher engagement compared to a non-gamified version?
5) Does higher interest in math puzzles and activities predict higher engagement?
6) Do students' levels of interest in math puzzles interact with conditions to predict engagement?

6.2 Methods

6.2.1 From Here to There!

From Here to There! (FH2T) is a web-based application for the exploration of arithmetic, symbolic algebra, and logic (Figure 6.1). This learning technology allows students to physically and dynamically interact with algebraic expressions and equations rather than having to write and rewrite processes as they would using a pen and paper. FH2T uses the *Graspable Math* library, a dynamic mathematical notation tool developed at Indiana University Bloomington. The web-based application was developed at Worcester Polytechnic Institute and is available online in both elementary and middle school versions.

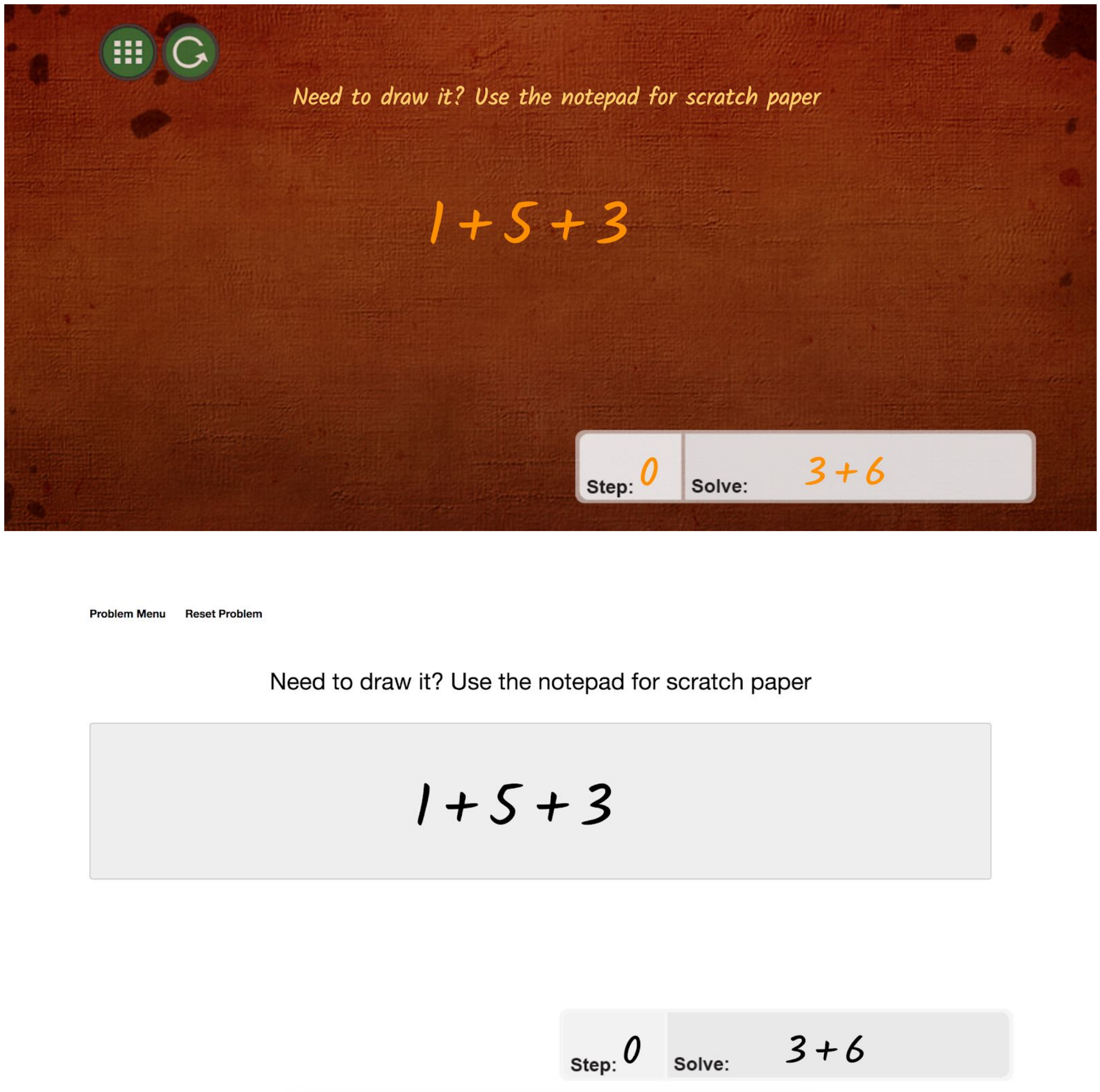


Figure 6.1. Screenshots of the gamified and non-gamified versions of From Here to There!

In FH2T, students solve math problems by transforming an equation or expression from the original state to a different end state through spatial transformations. These "goal states" often require the user to perform various mathematical procedures such as commutation and decomposition in order to solve the problem in the fewest number of steps. More often than not, these problems do not simply require the user to solve for x, but actually transform and manipulate the expression into a completely different form in an effort to display conceptual understanding. With each successful completion, players are awarded up to three clovers depending on their performance. Three clovers are awarded when students complete the problem in the fewest number of steps, two clovers are awarded when students complete the problem within two steps of the lowest possible amount, and one clover is awarded when students complete the problem with any additional steps. There are over a dozen worlds to complete in FH2T with around 20 problems per world. With each world, the user is introduced to more difficult problems and more advanced mathematical procedures ranging from commuting terms to factoring and distribution. Prior research has shown that FH2T improves learning gains and has tremendous success in introducing early algebra concepts to students on an elementary level (Braith et al., 2017).

In addition to being a successful learning technology, FH2T can also be used as a tool for teachers and researchers to gain a better understanding of the student perspective. As students solve problems in FH2T, interaction data is logged on the backend to allow researchers to recreate the problem-solving process for students' attempts as well as identify strategies in how they play the game. The system records the start and end time of each attempt, the number of resets a player makes, the number of steps in each attempt, the number of clovers gained with each completion, and the mouse trajectories used when interacting with the software for playback.

6.2.2 Participants

A study was conducted that explored the motivational and engagement factors of gamification in interactive learning environments. Forty seven (47) undergraduates from a private, engineering university in the Northeastern United States participated in a study where they solved math problems using a modified version of the math game, *From Here to There!*. Each participant was given credit for an unrelated course as compensation for taking part in the study. College students were selected because of their prior knowledge of mathematics required to solve the problems. Because the study outcome is focused on engagement and motivation rather than learning gains, students were not given a pretest to assess prior knowledge.

6.2.3 Procedures

Upon entering the room, students were seated at a computer and given an outline of the study. Each participant was to complete a questionnaire, a short tutorial page (Figure 6.2), and two main tasks. They were allowed to spend as much time or as little time on the first task (FH2T intervention) as they would like up to one hour, but were required to spend the remainder of the hour completing the second task. Unbeknownst to the participants, the second task was merely walking out of the room and they were free to go as soon as they left. The study was designed this way so the students felt the freedom to explore the intervention as long as they were engaged, but also dedicate the full hour toward the study so that they were not tempted to leave early if they knew there was no second task.

Before using the system, participants were asked to complete an online questionnaire assessing their interest in mathematics. The first question asked how many math courses they have taken ranging from Algebra I to Calculus IV and higher with an option to add additional courses. These questions targeted their overall exposure to mathematics and indicates whether or not they are in a math-related program in their current studies. They were also asked to assess their general standing in their math classes as worst in class, below average, average, above average, or best in class. Lastly, the questionnaire gave 11 statements regarding math and math activities that participants were to agree or disagree with on a 7-degree scale from strongly disagree (1) to strongly agree (7). This targeted participant interest in math using positively inflective statements such as “I find math enjoyable”, “Puzzles like Sudoku are a fun way to pass the time” and negatively inflective statements such as “I rarely try math problems not required by the teacher”. The feedback of this questionnaire was used to assess student interest in mathematics to see if high-interest students performed differently than low-interest students depending on conditions (RQ2, RQ3).

After completing the questionnaire, each participant was directed to a tutorial web-page that showed how to use the proprietary gestures and movements of the application that will be needed to play *From Here to There!*. There are 7 pages of tutorials with roughly 4 examples on each page demonstrating how to manipulate terms and perform mathematical operations to solve math problems. The tutorial goes through basic interactions of the software, but does not promote any particular strategy in solving problems nor does it provide any additional practice in solving problems. On completing the tutorial, the participant was redirected to either the gamified or non-gamified intervention based on the participant’s ID code. Both groups were then automatically logged into the web version of *From Here to There!* and asked to solve the same problems in the same order. A timestamp was logged from the moment they started the application. The gamification condition's version of FH2T included the colorful backgrounds, achievements, reward messages (Figure 6.3), and levels while the plain condition's version of FH2T consisted of a directory of the problems themselves without any gamified elements, designs, rewards, or levels (Figure 6.4). Both versions of the program had a link on the top of each screen that read "I don't want to continue" that participants were instructed to click when they were finished. When the link was clicked, the user was logged out and a timestamped event was logged indicating that the user finished. This corresponds to the timestamped event that was created when a user initially logged in.

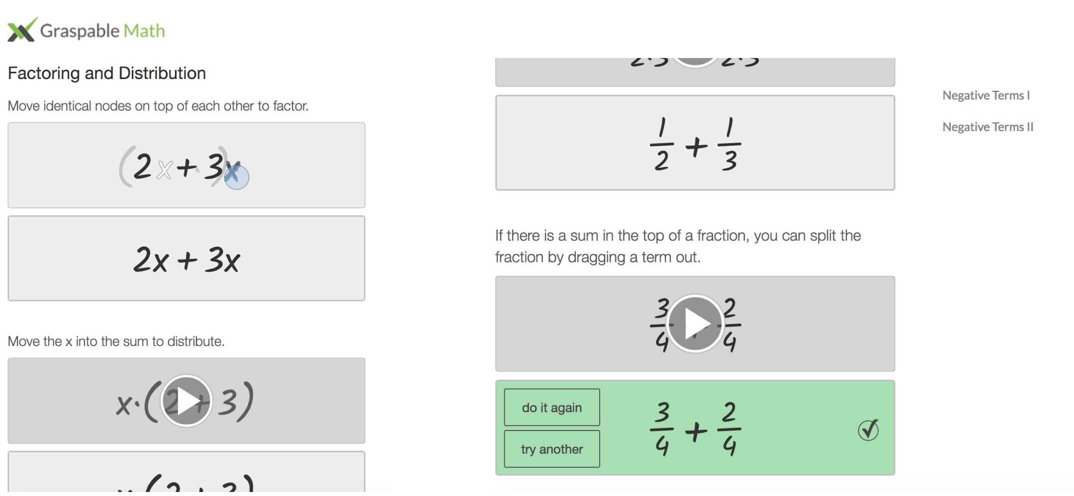


Figure 6.2. Screenshots of the gesture tutorial pages

Both groups were given the same 42 problems to solve that increased in overall difficulty. Each problem featured an initial state and a goal state. Participants were instructed to manipulate the problem to make it look like the goal state using the gestures that they learned in the tutorial. The Gamification condition's saw each of the 42 problems or "levels" into one of four worlds - Easy, Hard, Insane, and Impossible. Each level was unlocked upon completion of the previous one. After completing the 8th level of a world, the next world was unlocked and a reward text was displayed to the player. The player could still continue to finish the remaining levels of that world, if they chose to. The plain condition presented each problem in a directory on the main page that increased when players solved the previous problem. To keep the exposure to available problems consistent, the problem that would have been unlocked when solving the 8th level of a world in the gamification condition's also became unlocked when solving that same problem in the plain condition.

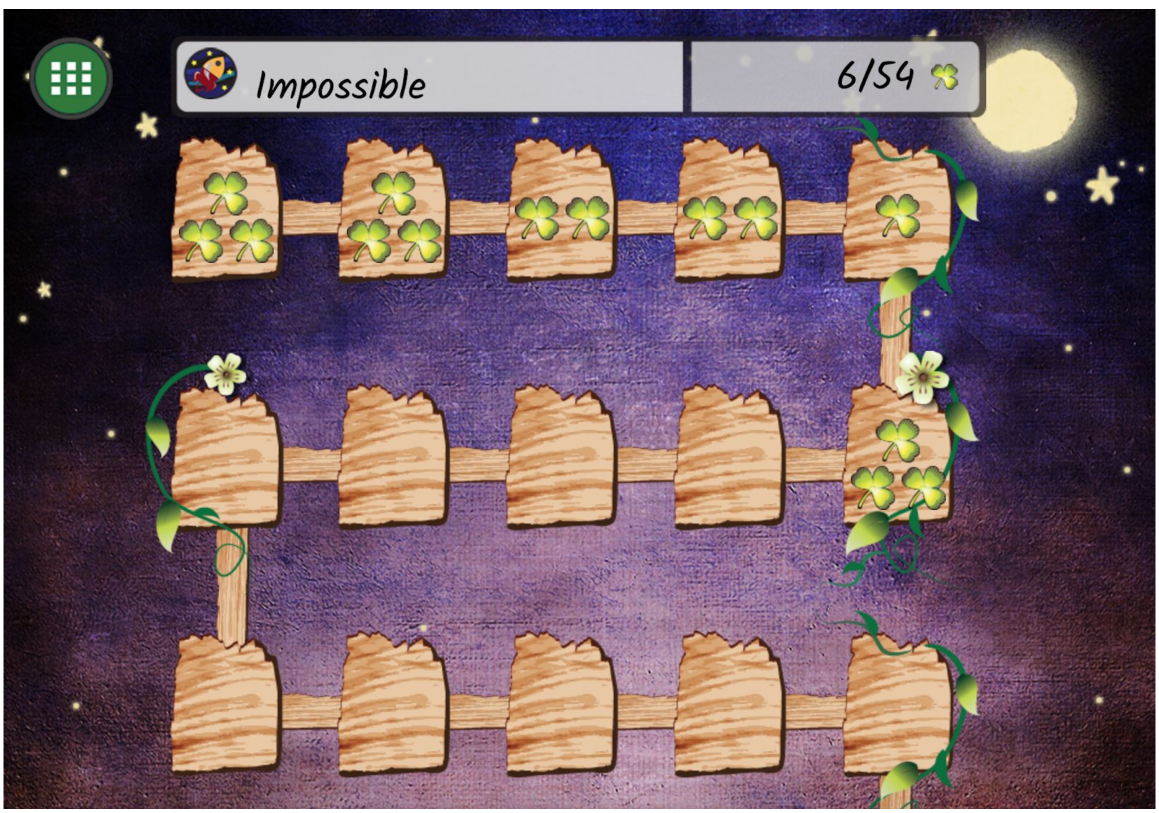


Figure 6.3. An example of the gamification condition's version of FH2T!

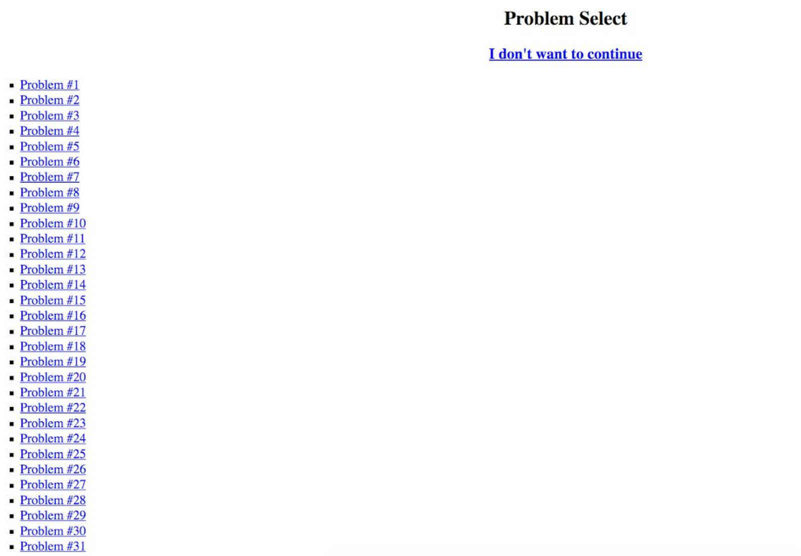


Figure 6.4. An example of the plain version of FH2T! without gamification

Participants were then given a post-survey after completing the intervention. These statements were intended to evaluate the enjoyment, challenge, and interest that each participant experienced when using their version of the program. The questionnaire included statements such as “The application really held my attention” and “I would prefer solving math problems on paper instead of using the application” and was also answered on a 1-7 scale. After completing the survey, participants were taken outside of the room and told that they were free to leave and that the second task was simply walking out of the room. This was designed so that participants would not rush through the first task with the intention of leaving early, but instead allowing them to freely choose when they no longer wanted to complete this task given their allotted timeslot.

6.2.4 Measures

1) **Interest in math puzzle questionnaire.** Participants completed a 13-item questionnaire that helped to measure their likely interest and engagement in math puzzles and activity. The first two questions asked the number of math courses they have taken as well as their self-assessment of their ranking in class (i.e. below average, best in class, etc.). This remainder of the questionnaire was modeled after the questionnaire used in Durik & Harackiewicz (2007) and consisted of 11 statements participants would agree with on a scale of 1-7. These statements were designed to gain an understanding of the participants' valuation of mathematics and included statements such as "I find math enjoyable" or "Math just doesn't really appeal to me". For measurement, these statements were summed with a lowest possible score of 11 and highest possible score of 77. For the negatively inflective statements such as "Math just doesn't really appeal to me", these were reverse scored in the total sum. The participants' scores were then averaged for data analysis as the primary factor of interest in math puzzles.

2) **Engagement.** Engagement is the main focus of the study and was measured using a combination of data recorded in the From Here to There web application.

    a) **Time spent in game.** Once the participant finished the training tutorial, they were automatically logged in to the FH2T intervention and a timestamp was recorded for that individual participant. When the participant finished the intervention by clicking the "I don't want to continue" link at the top of the screen, the system recorded a second timestamp indicating their logout time. Once the participant logged out, they were not able to solve any more problems and were redirected to a post-survey. We used these two timestamps to measure the amount of time spent in the app. Because the study was designed so that a student was free to stop using the application at any point,The participants that willingly chose to continue playing the intervention would have a higher amount of time spent in the intervention, demonstrating more engagement than a participant that stopped after only a few minutes.

b) **Number of distinct problems completed.** Similar to time spent, the number of problems that a participant completed in the intervention is useful as it refines the perspective of the "time spent" measure. The intervention had 42 problems increasing in difficulty. While the amount of time spent in the intervention can show us one facet of engagement, the number of problems can be used as an indicator of engagement to show how far a participant progressed in the intervention.

c) **Number of attempts.** When solving a problem, a participant could reset the problem as many times as they liked before completing it. When the participant started a problem, they were on attempt #1 and with each reset, the number of attempts increased. We can use the overall number of attempts a participant makes in the intervention as a form of engagement because it shows effort or what?.

d) **Number of "go-backs".** The gamification version of the intervention included a feature that awards clovers based on the user's performance. A participant was awarded clovers when they completed a problem. Completing a problem in the fewest number of moves awarded 3 clovers, completing a problem in two moves more than the fewest number awarded 2 clovers, and completing a problem in any additional moves awarded 1 clover. The number of clovers that a participant achieved in the game does not affect the study in any way and the goal of maximizing the number of clovers earned serves only as an intrinsic reward. However, this intrinsic motivation to go back and complete a problem for 3 clovers can be used as a measure of engagement. We can calculate the number of "go-back attempts" by calculating the total number of problems completed and then subtracting it by the number of distinct problems completed.

While these variables will distinguish specific in-game activity, we will focus our analysis and how we measure engagement using the amount of time spent using the technology (elapsed time).

6.3 Results

RQ1, "Do gamification elements relate to indicators of higher engagement compared to a non-gamified version?"

We found that engagement, as defined by the amount of time spent in the intervention, did increase by approximately 8 minutes in the gamified condition when compared to the non-gamified condition. Our initial hypothesis of the study was that the engagement and completion level of the group with the gamified version would be significantly higher than the group without any gaming elements when participants have a lower interest in mathematics. Additionally, it was expected that the opposite effects would occur for students with a high reported interest in mathematics in that they will have a much lower engagement level and completion rate in the gamified version and a slightly higher engagement level and completion rate in the gamified version (Table 6.1).

Table 6.1

Descriptives of gamified vs. non-gamified conditions

Measure	Group	N	Mean	Std. Dev.	Std. Error
Elapsed Time	Non-Gamified	20	1809.75	444.730	99.445
	Gamified	19	2281.37	696.921	159.885
Distinct Completed	Non-Gamified	20	36.850	6.499	1.453
	Gamified	19	30.895	10.413	2.389
Total Completed	Non-Gamified	20	36.850	6.499	1.453
	Gamified	19	33.526	13.268	3.044
Go Backs	Non-Gamified	20	0.000	0.000	0.000
	Gamified	19	2.632	5.529	1.269
Highest Completed	Non-Gamified	20	37.050	6.452	1.443
	Gamified	19	33.684	11.334	2.600
Total Attempts	Non-Gamified	20	48.750	10.269	2.296
	Gamified	19	65.789	38.084	8.737
Math Courses	Non-Gamified	20	7.65	1.785	0.399
	Gamified	19	6.53	2.525	0.579

Out of 47 students that were recruited to be participants for the study, 8 students were removed from the analysis for reasons such as spending less than 10 minutes in the session, showing up late to the session, and other external issues (technical issues, fire drills, etc.). These removals were made because they reduced the maximum amount of time allowed for the session. In the cases of leaving early, our pool of participants was awarded credit on arrival and unfortunately sometimes left immediately after showing up. Thus, if they spent less than 10 minutes in the session, they were not even completing the training sessions and questionnaire and would not be included in the analysis. Of the 39 remaining participants, there were 20 participants in the plain condition (without gamified elements) and 19 participants in the gamification condition.

RQ2, "Do gamification elements relate to indicators of higher engagement compared to a non-gamified version?"

Upon initial analysis, the two groups differed in their elapsed time using the software, the number of attempts made, and the number of problems completed. An Independent samples t-test shows that the gamified condition spent almost 1 standard deviation longer (roughly 8 minutes) in the software than the non-gamified condition on average (Table 6.1). Though it may not seem like a significant difference, 8 minutes is a long time considering that the total session was only an hour including the surveys and gesture tutorial. Independent samples t-tests also indicate that the number of problem attempts (the initial attempt plus any additional resets) was significantly higher for the gamification group as well as the number of additional attempts of a problem after the problem had been already completed once ("go-backs"). We believe that this is due to the clover reward feedback of the gamified condition. Though the rewards are meaningless in terms of gameplay or ability and were never mentioned in any instruction, this gamified element seems to be a contributing factor into resetting and re-attempting the same problem over and over for a better score after completion. This can be highlighted even further when we look at the number of problems completed between the two conditions. The average number of distinct problems for the gamified condition was 6 problems less than the plain condition and the number of total problems completed was 3 less in the gamified condition than that of the plain condition.

Table 6.2

Independent samples t-test of subgoal states vs. no subgoal states in FH2T

Measure	Group	*N*	*M*	*SD*	*t*	*p*
Elapsed Time	Non-Gamified	20	1809.75	444.730	-2.533	0.016
	Gamified	19	2281.37	696.921		
Distinct Completed	Non-Gamified	20	36.850	6.499	2.155	0.038
	Gamified	19	30.895	10.413		
Total Completed	Non-Gamified	20	36.850	6.499	1.001	0.323
	Gamified	19	33.526	13.268		
Go Backs	Non-Gamified	20	0.000	0.000	-2.074	0.053
	Gamified	19	2.632	5.529		
Highest Completed	Non-Gamified	20	37.050	6.452	1.147	0.259
	Gamified	19	33.684	11.334		
Total Attempts	Non-Gamified	20	48.750	10.269	-1.886	0.074
	Gamified	19	65.789	38.084		

Math Courses	Non-Gamified	20	7.65	1.785	1.611	0.116
	Gamified	19	6.53	2.525		

We originally thought this may be due to an imbalance in the conditions or that perhaps the plain condition was just faster on average than the gamified condition (Table 6.3). However, the number of attempts of the gamified condition exceeded the plain condition by 17 attempts on average, which is quite high. If we break it down to the amount of time per attempt, we see that the differences between the two conditions are negligible and not statistically significant (Table 6.4). This highlights that though the two conditions are moving at about the same pace, the gamified condition is re-attempting problems more often and engaging with the system more than the plain condition.

Table 6.3

Descriptives comparing the rate of speed between two conditions

Measure	Group	N	Mean	Std. Dev.	Std. Error
Time Per Problem	Non-Gamified	20	49.866	12.126	2.711
	Gamified	19	83.059	36.061	8.273
Attempts Per Problem	Non-Gamified	20	1.329	0.188	0.042
	Gamified	19	2.162	0.933	0.214
Time Per Total Problem	Non-Gamified	20	49.866	12.126	2.711
	Gamified	19	76.841	30.347	6.962
Time Per Attempt	Non-Gamified	20	37.470	6.777	1.515
	Gamified	19	40.941	14.930	3.425

Table 6.4

Independent samples t-test comparing the rate of speed between two conditions

Measure	Group	*N*	*M*	*SD*	*t*	*p*
Time Per Problem	Non-Gamified	20	49.866	12.126	-3.813	0.001
	Gamified	19	83.059	36.061		
Attempts Per Problem	Non-Gamified	20	1.329	0.188	-3.818	0.001
	Gamified	19	2.162	0.933		
Time Per Total Problem	Non-Gamified	20	49.866	12.126	-3.610	0.001
	Gamified	19	76.841	30.347		
Time Per Attempt	Non-Gamified	20	37.470	6.777	-0.927	0.363
	Gamified	19	40.941	14.930		

We believe that this is due to gamified participants spending more time perfecting their steps in order to gain a better score instead of just going through the task as they did in the plain condition. This suggests that gamified elements do in fact engage participants and motivate them to spend more time in the system rather than just work through the problems (RQ1).

RQ3: Do students' levels of interest in math puzzles interact with conditions to predict engagement?

We then explored how interest factored into the data using the prior interest survey that participants completed at the beginning of the study. Each of the questions was on a scale of 1 to 7 with 1 meaning low and 7 meaning high and some questions were reverse-scored (1 meaning high and 7 meaning low). Each participant's scores were averaged (the reverse-scored questions were re-encoded) and a median split was performed to separate participants into either a high or low interest group (Table 6.5). A t-test was then performed comparing participants between the high and low interest groups regardless of whether or not they were in the gamified condition (Table 6.6). We found that there were very few differences between the two groups and only the average time spent per completed problem and average time per attempt were statistically significant. This indicates that the high interest group was able to solve problems more quickly than the low interest group, but doesn't indicate any type of influence on the overall elapsed time or other engagement measures (RQ2).

Table 6.5

Descriptives table comparing low and high interest groups

Measure	Group	N	Mean	Std. Dev.	Std. Error
Elapsed Time	Low Interest	21	2073.76	637.304	139.071
	High Interest	18	1999.56	617.929	145.647
Distinct Completed	Low Interest	21	31.857	10.7345	2.3425
	High Interest	18	36.389	5.922	1.396
Total Completed	Low Interest	21	32.714	11.279	2.461
	High Interest	18	38.167	8.563	2.018
Go-Backs	Low Interest	21	0.857	2.308	0.504
	High Interest	18	1.777	5.440	1.282
Highest Completed	Low Interest	21	33.048	11.342	2.475
	High Interest	18	38.167	4.768	1.123
Total Attempts	Low Interest	21	53.619	24.177	5.276
	High Interest	18	61.056	33.178	7.820
Math Courses	Low Interest	21	7.05	2.224	0.485
	High Interest	18	7.17	2.282	0.538

Table 6.6

Independent samples t-test of low and high interest groups

Measure	Group	*N*	*M*	*SD*	*t*	*p*
Elapsed Time	Low Interest	21	2073.76	637.304	0.368	0.715
	High Interest	18	1999.56	617.929		
Distinct Completed	Low Interest	21	31.857	10.7345	-1.593	0.120
	High Interest	18	36.389	5.922		
Total Completed	Low Interest	21	32.714	11.279	-1.677	0.102
	High Interest	18	38.167	8.563		
Go-Backs	Low Interest	21	0.857	2.308	-0.706	0.485
	High Interest	18	1.777	5.440		
Highest Completed	Low Interest	21	33.048	11.342	-1.883	0.070
	High Interest	18	38.167	4.768		
Total Attempts	Low Interest	21	53.619	24.177	-0.808	0.424
	High Interest	18	61.056	33.178		
Math Courses	Low Interest	21	7.05	2.224	-0.165	0.870
	High Interest	18	7.17	2.282		

We then looked at whether or not the interaction between high and low interest participants and the intervention conditions affected participant engagement (RQ3). We had previously hypothesized that the gamified elements might perhaps catch the attention of low interest participants while possibly turning off high interest participants. In order to answer this question, we ran our analyses with the participants allotted to one of four groups: low interest, plain intervention; low interest, gamified intervention; high interest, plain intervention; and high interest, gamified intervention. We then ran a one-way ANOVA and compared the means of each of the four groups and found that while the amount of time spent wasn't statistically significant between all four of the groups, the amount of attempts per problem, the amount of time per problem, the amount of time per attempt, and the amount of time per distinct problem were all statistically significant (Table 6.7). For three out of the four variables, the two gamified groups (low and high interest) had a higher average than the two plain groups. The amount of time per attempt found that the gamified, high interest group was actually the lowest by four seconds on average while the high interest plain group was the second lowest on average. This indicates that while the high interest groups are working slightly faster on each attempt, the gamified conditions are actually spending more time on each problem and attempting problems more regardless of interest.

Table 6.7

Descriptives of gamification and non-gamification conditions against low and high interest groups

Measure	Group	N	Mean	Std. Dev.	Std. Error
Elapsed Time	Low-Plain	11	1902.45	483.111	145.664
	High-Plain	10	2262.20	753.603	238.310
	Low-Gamified	9	1696.44	389.530	129.843
	High-Gamified	9	2302.67	673.013	224.338
	Total	39	2039.51	621.283	99.485
Distinct Completed	Low-Plain	11	35.636	7.145	2.154
	High-Plain	10	27.700	12.755	4.033
	Low-Gamified	9	38.333	5.657	1.886
	High-Gamified	9	34.444	5.833	1.944
	Total	39	33.949	9.032	1.446
Total Completed	Low-Plain	11	35.636	7.1452	2.154
	High-Plain	10	29.500	14.285	4.517
	Low-Gamified	9	38.333	5.657	1.886
	High-Gamified				

	Total	9	38.000	11.124	3.708
		39	35.231	10.361	1.659
Go-Backs	Low-Plain	11	0.000	0.000	0.000
	High-Plain	10	1.800	3.155	0.998
	Low-Gamified	9	0.000	0.000	0.000
	High-Gamified	9	3.556	7.468	2.489
	Total	39	1.282	4.032	0.646
Highest Completed	Low-Plain	11	35.818	7.291	2.198
	High-Plain	10	30.000	14.391	4.551
	Low-Gamified	9	38.556	5.271	1.757
	High-Gamified	9	37.778	4.494	1.498
	Total	39	35.410	9.196	1.472
Total Attempts	Low-Plain	11	50.909	11.794	3.556
	High-Plain	10	56.600	33.550	10.609
	Low-Gamified	9	46.111	7.897	2.632
	High-Gamified	9	76.000	42.122	14.041
	Total	39	57.051	28.534	4.569

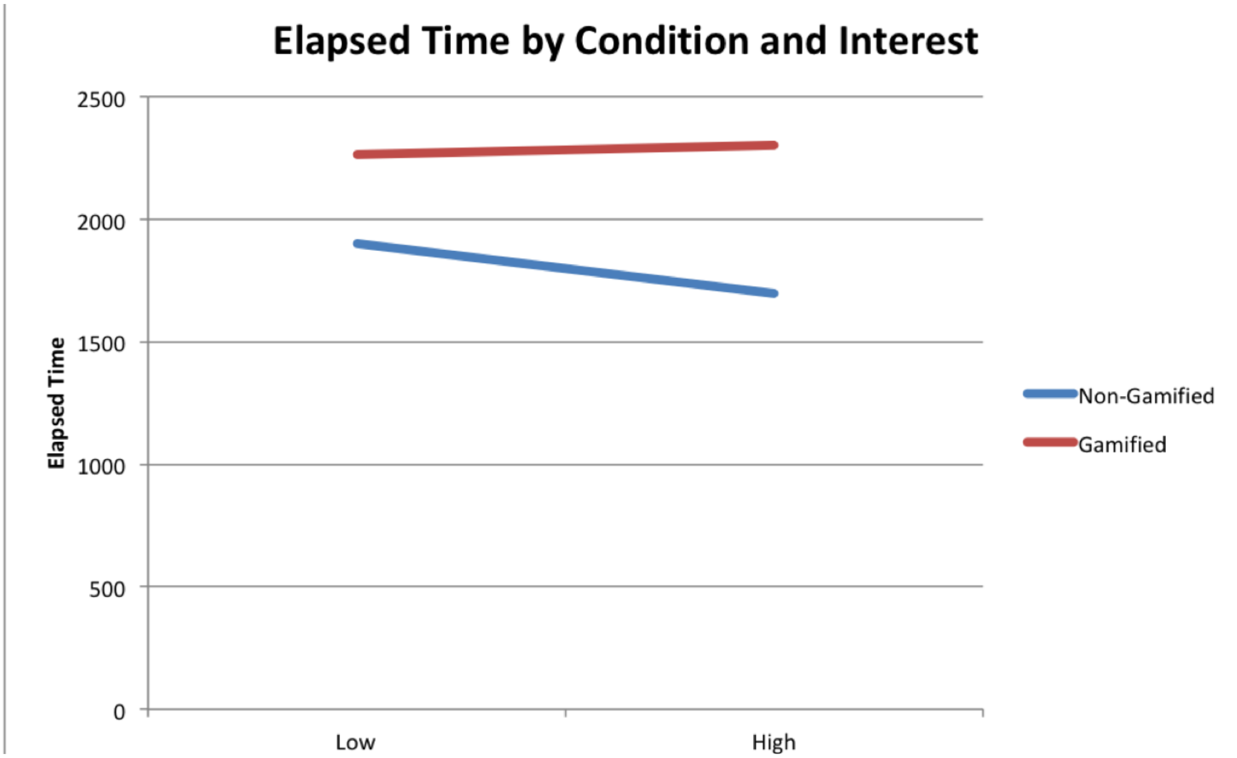
Table 6.8

ANOVA of gamification and non-gamification conditions against low and high interest groups

Measure	Group	*Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*
Elapsed Time	Between Groups	2385043.194	3	795014.398	2.265	0.098
	Within Groups	12282662.549	35	350933.216		
	Total	14667705.744	38			
Distinct Problems Completed	Between Groups	597.030	3	199.010	2.783	0.055
	Within Groups	2502.868	35	71.511		
	Total	3099.897	38			
Total Completed	Between Groups	485.878	3	161.959	1.578	0.212
	Within Groups	3593.045	35	102.658		
	Total	4078.923	38			
Go-Backs	Between Groups	82.075	3	27.358	1.787	0.168
	Within Groups	535.822	35	15.309		
	Total	617.897	38			

Highest Problem Completed	Between Groups	434.022	3	144.674	1.822	0.161
	Within Groups	2779.414	35	79.412		
	Total	3213.436	38			
Total Attempts	Between Groups	4725.699	3	1575.233	2.103	0.118
	Within Groups	26214.198	35	748.977		
	Total	30939.897	38			

We then explored the relationship between prior interest and gamification between high and low interest participants. Though the data suggests that there was no significant relationship between all four groups, we ran t-tests comparing interest between each of the groups separately. There were no significant differences between the low interest gamification group and the low interest plain group. There were also no additional findings between comparing alternate interest groups of one condition to another, such as low interest groups of plain to high interest groups of gamification. Any findings between these groups appear to be consistent with the initial differences of the intervention. However, there was a significant finding in elapsed time between the high interest groups. The elapsed time between high interest, plain condition was roughly 600 seconds or 1 standard deviation less than the high interest, gamified condition. It is also worth noting that the number of problems completed, the rate at which problems were completed, and the rate of number attempts were all found not to be significant between the two groups. This suggests that although they were working at the same pace and ultimately completing the same amount of problems, the high interest, gamified group spent more time engaging in the intervention while the high interest, plain group simply completed the task and stopped. Additionally, the high interest plain group spent the least amount of time using the system on average while the high interest gamified group spent the most amount of time.



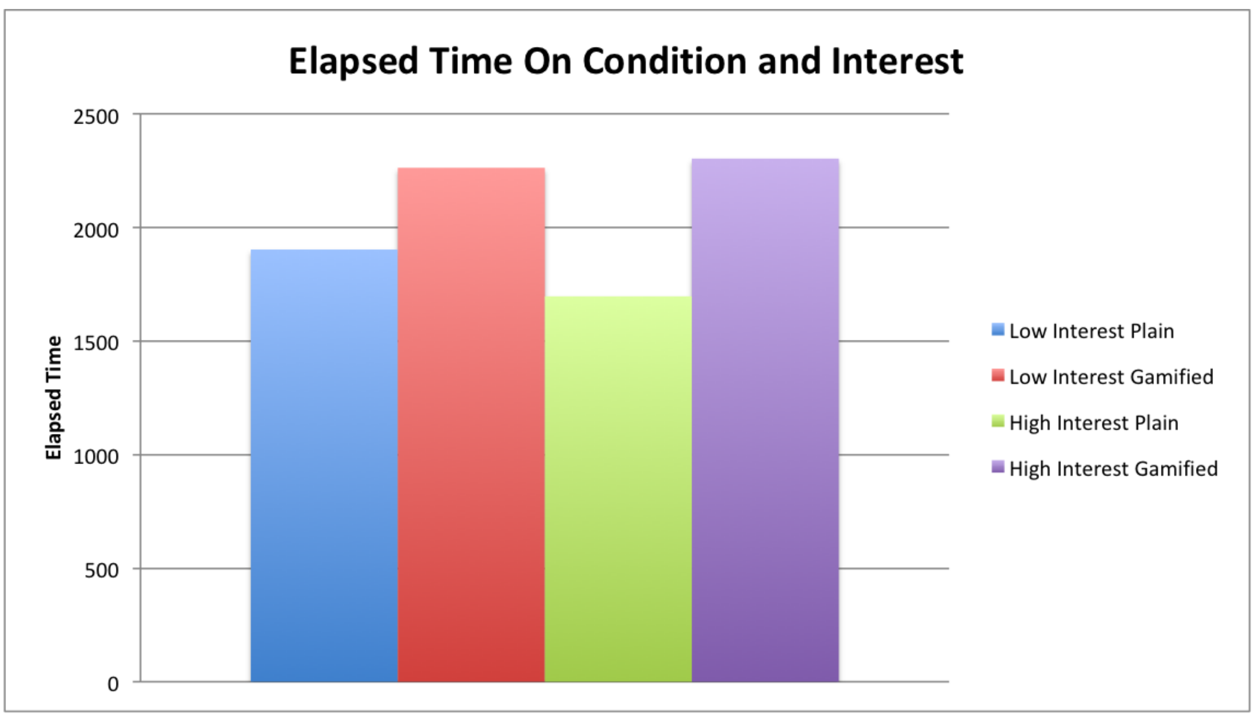


Figure 6.5. A comparison of elapsed time between the four groups

6.4 Discussion

6.4.1. Overview

The difference in results between conditions is highly correlated to the differences between the two styles of intervention. The elapsed time in the gamified condition is significantly higher than that of the plain condition, showing that students in the gamified condition chose to spend a longer time in the intervention than those in the non-gamified (plain) condition. Despite elapsed time being higher for the gamified condition, the number of distinct problems completed was significantly higher for the plain condition. We originally thought this might be due to an imbalance in the prior knowledge between the two groups. However, the number of go-backs and total number of attempts for the gamified condition were significantly higher than those of the plain condition. One possible explanation is due to the clover reward feedback of the gamified condition. Though the rewards are meaningless in terms of gameplay or ability and were never mentioned in any instruction, this gamified element seems to be a contributing factor of resetting and re-attempting completed problems over and over for a better score. This interesting finding may also explain why the number of distinct levels completed was significantly lower for the gamified condition. We believe that this is due to participants in the gamified condition spending an excessive amount of time perfecting their steps in order to gain a better score instead of just going through the task as participants did in the plain condition. This suggests that gamified elements do in fact engage participants and motivate them to spend more time in the system rather than just work through the problems.

When factoring in interest, we found that the amount of time spent remains relatively consistent to our initial findings for the low interest plain group and the low interest gamification group. However, the elapsed time polarizes for our high interest groups with the high-interest, plain condition spending the least amount of time in the intervention across all four groups while the high-interest, gamification condition spending the most amount of time in the intervention across all four groups. The elapsed time in the intervention had p-values nearing significance, but this is considering that the sample size dropped for each group when we split it by interest, leaving us with very small sample sizes. Additionally, we attribute this behavior to goal-objectives that are found in the gamification condition. Because we gave the plain condition just the problems with no realistic other goals other than just to solve the problem, their behavior reflected this by simply completing the problems and stopping. The gamification condition, though there was no explicit instruction for gaining clovers or additional objectives, embraced the gamified objectives of the intervention and spent more time trying to meet these objectives. If this is true, we may be able to invoke longer engagement by adding additional objectives in the intervention such as multiple goal states or competitive play. Additionally, it is worth noting that there were no significant correlations between the number of math courses taken, the self-assessments on prior interest in math puzzles and activities, and the participant feedback on enjoyment and engagement after the intervention.

6.4.2 Conclusion

Our initial findings suggest that there is a significant correlation to higher engagement in the gamification group than that of the control (non-gamified) group. We measured this in the total amount of time spent using the intervention. Our in-app data allowed us to gain insight into what differed between the two groups in terms of user actions and other useful information such as the amount of time spent on each problem and attempt. Our findings suggest that gamification elements do help students engage in math-related tasks longer than they normally would have. This result was also consistent across both low and high interest participants. Additionally, while the rate of time per attempt was consistent across the high interest plain and high interest gamified groups, the high interest plain spent the least amount of time of all four groups while the high interest gamified group spent the most amount of time despite having completed the same number of distinct problems. We believe this is due to the extrinsic rewards of the gamified condition. Though they do not actually give an advantage when earned or serve a purpose other than existing as an achievement, the clovers seem to motivate students to engage more with the system. This supports our initial hypothesis that these gamified elements do encourage engagement and motivation in performing less than desirable tasks. Lastly, it should be noted that there were no major differences in preference based on the post-test survey involving the enjoyment of the interventions themselves though participants were not subjected to both conditions to compare between the two.

6.4.3 Future Work

Other factors that should be considered for future work include the difficulty of the problems themselves. Initially, we planned this study with the intention of removing the math ability factor in order to only focus on enjoyment and engagement. However, because almost all of the participants have completed up to Calculus II, the content may have been not challenging enough to truly distinguish the difference of engagement when considering prior knowledge as a factor. Thus, it may be that difficulty or challenge may be a factor that would bear different results in terms of engagement amongst the different conditions. The proprietary gestures of the technology proved to be difficult for some of the participants as that was a common complaint on the post-survey across all conditions. Additionally, future work in other gamification elements such as leaderboards, pointification, and multi-objective goal states should be explored to see how these affect engagement and motivation.

References

Anderson, J. R. (2007). How can the human mind exist in the physical world? Oxford, England: Oxford University Press.

Braith, L., Daigle, M., Manzo, D., & Ottmar, E. (2017). Even Elementary Students Can Explore Algebra!: Testing the Feasibility of from Here to There!, a Game-Based Perceptual Learning Intervention. American Psychological Society Conference, Boston, MA.

Dempsey, J., Lucassen, B., Gilley, W., & Rasmussen, K. (1993). Since Malone's theory of intrinsically motivating instruction: What's the score in the gaming literature?. Journal of Educational Technology Systems, 22(2), 173-183.

Deterding, S., Sicart, M., Nacke, L., O'Hara, K., & Dixon, D. (2011, May). Gamification. using game-design elements in non-gaming contexts. In *CHI'11 extended abstracts on human factors in computing systems* (pp. 2425-2428). ACM.

Durik, A. M., & Harackiewicz, J. M. (2007). Different strokes for different folks: How individual interest moderates the effects of situational factors on task interest. Journal of Educational Psychology, 99(3), 597.

Falstein, N. (2002). Interactive stealth learning. Retrieved October, 4, 2004.

Hogle, J. G. (1996). Considering games as cognitive tools: In search of effective.

Jacobs, J. W., & Dempsey, J. V. (1993). Simulation and gaming: Fidelity, feedback, and motivation. Interactive instruction and feedback, 197-227.

Lepper, M. R., & Malone, T. W. (1987). Intrinsic motivation and instructional effectiveness in computer-based education. Aptitude, learning, and instruction, 3, 255-286.

Malone, T. W. (1981). What makes things fun to learn? A study of intrinsically motivating computer games. Pipeline, 6(2), 50.

Malone, T. W. (1983). Guidelines for designing educational computer programs. Childhood Education, 59(4), 241-247.

Malone, T. W., & Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. Aptitude, learning, and instruction, 3(1987), 223-253.

Malouf, D. B. (1988). The effect of instructional computer games on continuing student motivation. The journal of special education, 21(4), 27-38.

Mayland, V. (1990). Carmen Sandiego Is in Your Classroom. An Idea Packet.

Mitchell, M. (1993). Situational interest: Its multifaceted structure in the secondary school mathematics classroom. Journal of educational psychology, 85(3), 424.

Pierfy, D. A. (1977). Comparative simulation game research: Stumbling blocks and steppingstones. Simulation & Games.

Ragupathi, K. (2012). Gamification: How to do it Right and Why it is No Good. *Technology in Pedagogy series*, 4-5.

Rieber, L. P. (1996). Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games. Educational technology research and development, 44(2), 43-58.

Rosas, R., Nussbaum, M., Cumsille, P., Marianov, V., Correa, M., Flores, P., ... & Rodriguez, P. (2003). Beyond Nintendo: design and assessment of educational video games for first and second grade students. Computers & Education, 40(1), 71-94.

Tüzün, H., Yılmaz-Soylu, M., Karakuş, T., İnal, Y., & Kızılkaya, G. (2009). The effects of computer games on primary school students' achievement and motivation in geography learning. Computers & Education, 52(1), 68-77.

Warren, A., Littleton, F., & Macleod, H. (2009). Welcome to my World: Induction to Games for Learning. In Proceedings of the 3rd European Conference on Games Based Learning (pp. 169-291).

Wood, L. E., & Stewart, P. W. (1987). Improvement of practical reasoning skills with a computer game. Journal of Computer-Based Instruction.

7 STUDY 2: EFFECTS OF POINTS AND LEADERBOARDS IN EDUCATIONAL TECHNOLOGY

The following chapter explores the effects of social competition using a modified version of the web app From Here to There! This modified version introduces subgoals and leaderboards that are manipulated by condition to reflect an always leading, always trailing, or back and forth standing for the user. A study was conducted to examine the changes in engagement among undergraduate students when using the application featuring gamified elements such as pointification and leaderboards. Though humans are competitive in nature, it is worth exploring whether or not these types of components which promote competition discourages users from engaging when compared to the same intervention that does not include such features in its design. The results suggest that there is no statistically reliable difference in student performance when incorporating leaderboards or subgoals into the learning intervention.

7.1 Introduction

7.1.1 Overview

As discussed in the previous chapter, it is often difficult for students to feel engaged or motivated when it comes to mathematics. As Sarason (1983) expressed, “schools are not very interesting places for most of the people in them”. This results in students feeling bored or disinterested in when it comes to math and math-related activities. The intended outcome for our work is to ultimately help motivate students to want to learn and understand mathematics. While this is not an easy task, we find that adding gamification and gamified elements to our interventions may alleviate some of the more distasteful aspects of math education at least until they fully understand the value of mathematics themselves. In Ottmar et al. (2015), we found that learning gains when using the From Here to There! app were positively linked to exposure, thus the more worlds that a user played through the higher they would perform on the posttest. Additionally, other research also suggests that increased exposure to STEM initiatives and activities positively impacts perceptions and dispositions in mathematics (Bybee, & Fuchs, 2006). Thus among other use cases, we are exploring gamification and gamified elements as a means to increase prolonged exposure to math-related activities to pique interest and motivate students to further explore mathematics.

Though gamification as a whole is a relatively new topic in education research, there has been an overwhelming amount of research on the effects of intrinsic and extrinsic components that are similar to those of gamification. Points, badges, and leaderboards, for example, exemplify both intrinsic and extrinsic rewards among users and are often the main focus of traditional gamification research. In Chapter 6, we focused on the effects of unlocking worlds and receiving clover awards influenced user engagement when using From Here to There! Deci and Ryan (1985) suggests that these incentives may actually have a negative effect by rewarding someone to complete a task that they may already be motivated to do. Our findings in that study suggest that these extrinsic rewards do help students engage in math-related tasks longer than they normally would have regardless of prior interest, though we do not know the effects of gamification as a long-term design strategy.

Though introducing elements such as points and badges to educational activity are often an easy method of appealing to less interested audiences, one of the more natural methods of gamification that occurs both in and out of the classroom is competition. Leaderboards are the typical method of social comparison in gamified environments. Often these leaderboards reflect the in-game performance of users' points in descending order to reflect their performance standings amongst each other. In addition to influencing academic performance (Christy & Fox, 2014), there is also an overwhelming amount of pressure in social comparisons to either maintain a high standing or ascend to an acceptable standing (Wells & Skowronski, 2012). Festinger's Theory of Social Comparison outlines a series of hypotheses in social comparison between groups and individuals (Festinger, 1954). Most notably, the first three hypotheses deal with direct comparison between individuals based off of the need to evaluate themselves. Festinger's first hypothesis states that "There exists, in the human organism, a drive to evaluate his opinions and his abilities". This is followed up by defining the functional ties between opinion and ability, as well as the necessity for knowing the true extent of one's own ability in certain circumstances as a matter of life and death. Festinger (1954) states his second hypothesis "To the extent that objective, non-social means are not available, people evaluate their opinions and abilities by comparison respectively with the opinions and abilities of others." Festinger explains this hypothesis with an example that a person's evaluation of their ability to write poetry depends significantly on the opinions of others on his ability to write poetry. In cases where there is a clear objectiveness in criterion, the evaluation of one's ability does not matter as much on the opinion of others, but actually on the comparison between that person's performance and the performance of others (Festinger, 1954). Festinger's third hypothesis focuses on the differences between others in comparison: "The tendency to compare oneself with some other specific person decreases as the difference between his opinion or ability and one's own increases." This is expressed as the ability to *accurately* determine one's ability by comparing to others who are similar in ability to themselves. When one is comparing their performance to others that are far too above or below their own performance, a self-imposed restriction is enforced invalidating any type of comparison due to the incomparable differences between the two abilities. For this chapter, we are expanding on Festinger's Theory of Social Comparison using gamification in an educational environment. When implementing a leaderboard in From Here to There! to reflect social competition in the intervention, we hypothesize that users will behave differently depending on their determined ability among their peers indicated by their position on the leaderboard. For this study, however, the leaderboard will be controlled based on condition to naturally reflect a superior, inferior, or fluctuating assessment of their performance in contrast to their matched opponents. We anticipate that in concordance with Festinger's work, the fluctuations of position will engage them more due to their capability of assessing their own performance whereas the other conditions will not be able to accurately assess their performance due to outperforming or underperforming their peers.

7.1.2 From Here to There!

From Here to There! (FH2T) is a web-based application for the exploration of arithmetic, symbolic algebra, and logic. This learning technology allows students to physically and dynamically interact with algebraic expressions and equations rather than having to write and rewrite processes as they would using a pen and paper. FH2T uses the *Graspable Math* library, a dynamic mathematical notation tool developed at Indiana University Bloomington. The web-based application was developed at Worcester Polytechnic Institute and is available online in both elementary and middle school versions.

Figure 7.1. Screenshot of FH2T welcome screen

In FH2T, students solve math problems by transforming an equation or expression from the original state to a different end state through spatial transformations. These “goal states” often require the user to perform various mathematical procedures such as commutation and decomposition in order to solve the problem in the fewest number of steps. There are over a dozen worlds to complete in FH2T with around 20 problems per world. With each world, the user is introduced to more difficult problems and more advanced mathematical procedures ranging from commuting terms to factoring and distribution. Prior research has shown that FH2T is related to higher learning gains and has success in introducing early algebra concepts to students on an elementary level (Braith et al., 2017).

In addition to being a successful learning technology, FH2T can also be used as a tool for teachers and researchers to gain a better understanding of the student perspective. As students solve problems in FH2T, interaction data is being logged on the backend that allows researchers to recreate the problem-solving process for each student's attempt as well as identify strategies in how they play the game. The system records the start and end time of each attempt, the number of resets a player makes, the number of steps in each attempt, the number of stars gained with each completion, and the mouse trajectories used when interacting with the software for playback.

7.1.3 Research Goals

1) Do "subgoal states" and pointification lead to higher engagement of the intervention?
2) Does the social/competitive feature present in the leaderboard component of the intervention further promote engagement than not having competitive virtual peers present?
3) Do certain standings on the leaderboard promote or discourage engagement more than others?

7.2 Methodology

7.2.1 Participants

A study was conducted that explored the effects of social competition engagement factors of gamification in interactive learning environments. 200 undergraduates from a private engineering university in the Northeastern United States participated in a study where they solved math problems using a modified version of the math game, *From Here to There!*. Each participant was given credit for an unrelated course as compensation for taking part in the study. College students were selected because of their prior knowledge of mathematics required to solve the problems. Because the study outcome is focused on engagement and motivation not learning gains, they were not given a pretest to assess prior knowledge.

7.2.2 Procedures

Participants signed up to the study through an online portal through their university to fulfill a requirement for a course. Upon starting the study, each participant was to complete an online questionnaire that focused on their prior math education as well as their interest in math-related puzzles and activities, a short tutorial page that teaches them the gestures used in From Here to There!, and finally given two main tasks. They were allowed to spend as much time or as little time on the first task (FH2T intervention) as they would like up to 1 hour, but were required to spend the remainder of the hour completing the second task that was not revealed to them (a separate gamified learning environment). The study was designed this way so the students felt the freedom to explore the first intervention (From Here to There!) as long as they were engaged, but also dedicate the full hour toward the study so that they were not tempted to end early if they knew there was no second task. Each participant was given credit for an unrelated course as compensation for taking part in the study. College students were selected because of their prior knowledge of mathematics required to solve the problems. Because the study outcome is focused on engagement and motivation rather than learning gains, students were not given a pretest to assess prior knowledge.

7.2.3 Survey

Participants were then asked to complete an online questionnaire assessing their interest in mathematics. The first question asked how many math courses they have taken ranging from Algebra I to Calculus IV and higher with an option to add additional courses. This targeted their initial exposure and indicates whether or not they are in a math-related program. They were also asked to assess their general standing in the classroom as worst in class, below average, average, above average, or best in class. Lastly, the questionnaire gave 11 statements regarding math and math activities that they were to agree or disagree with on a 7-degree scale from strongly disagree (1) to strongly agree (7). This targeted a participant's interest in math using positively inflective statements such as “I find math enjoyable”, “Puzzles like Sudoku are a fun way to pass the time” and negatively inflective statements such as “I rarely try math problems not required by the teacher”. The feedback of this questionnaire was used to assess student interest in mathematics to see if high-interest students performed differently than low-interest students depending on conditions. This questionnaire was based on the one featured in Durik and

Harackiewicz (2007) and is the same one featured in the Gamification v. Non-Gamification study in Chapter 6.

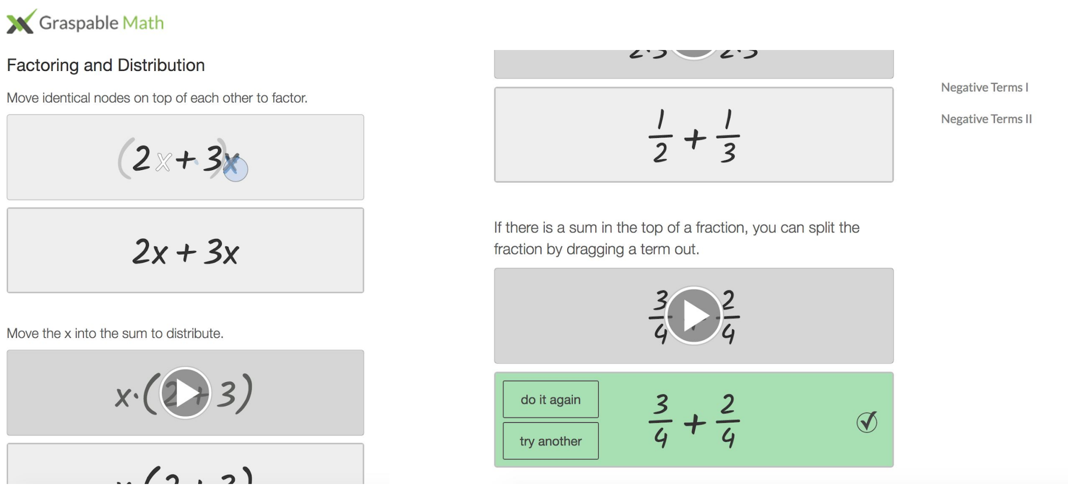


Figure 7.2. Screenshots of the gesture tutorial pages

After completing the questionnaire, each participant was directed to a tutorial web-page that showed how to use the proprietary gestures and movements of the application that will be needed to play From Here to There!. There are 6 pages of tutorials with roughly 4 examples each page demonstrating how to manipulate terms and perform mathematical operations to solve math problems. The tutorial goes through basic interactions of the software, but does not promote any particular strategy in solving problems nor does it provide any additional practice in solving problems.

7.2.4 Conditions

Upon completing the tutorial, the participant is redirected to the modified version of FH2T! and randomly assigned to one of four conditions by an internal random number generator in the project's code. All groups are then automatically logged into the web version of From Here to There and are asked to solve the same problems in the same order. A timestamp is logged from the moment they start the application. Each of the four groups are given a modified version of From Here to There! that features three additional subgoals that players may reach for additional points before solving the main goal state in lieu of the traditional three clover reward system found in the original game. A leaderboard that features the players' total points was also introduced in the study.

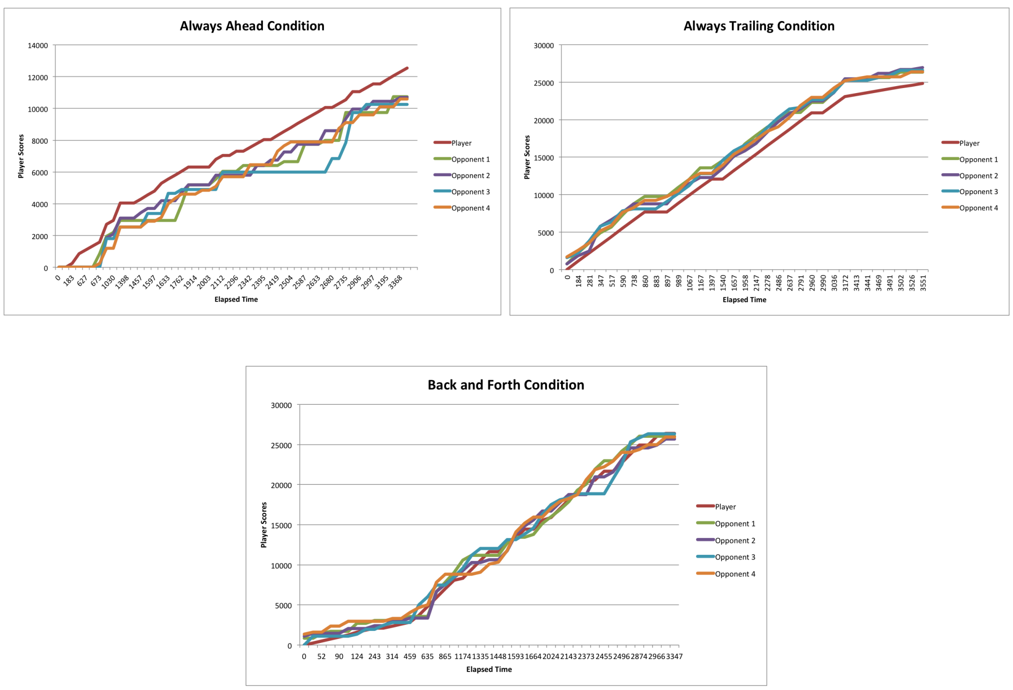


Figure 7.3. Visualizations of the participant's score versus AI scores in the three competitor conditions with the player's score represented in red

In three out of the four conditions, the leaderboard shows the player's score amongst four other "players" whose scores are intelligently generated based on the player's current score and point in time during the intervention. In the "always ahead" condition, players quickly form a lead early on and maintain their lead throughout the remainder of the intervention. In the “always behind” condition, players may find themselves equal or even ahead of their fictitious peers, but they will quickly lose their lead and will always trail in last place on the leaderboards for the remainder of the intervention. In the “back and forth” condition, players’ positions fluctuate based on time and their score. For example, during the first 200 seconds, AI players are programmed to randomly trail the player’s score by a specific amount, but are then programmed to randomly lead the player’s score by another amount in the following 200 seconds. This back and forth pattern continues throughout the remainder of the intervention so that AI players intermittently change positions with the player as long as they continue solving problems and gaining points. Lastly, there is a control condition that explores the same intervention, but features the player on the leaderboard without any additional players (no competitors condition). All versions of the program had a link on the top of each screen that read “I don’t want to continue” that participants were instructed to click when they were finished. When the link is clicked, the user is logged out, redirected to a non-related task, and a timestamped event is logged indicating that the user is finished. This timestamp corresponds to the timestamped event that is created when a user is initially logged in.

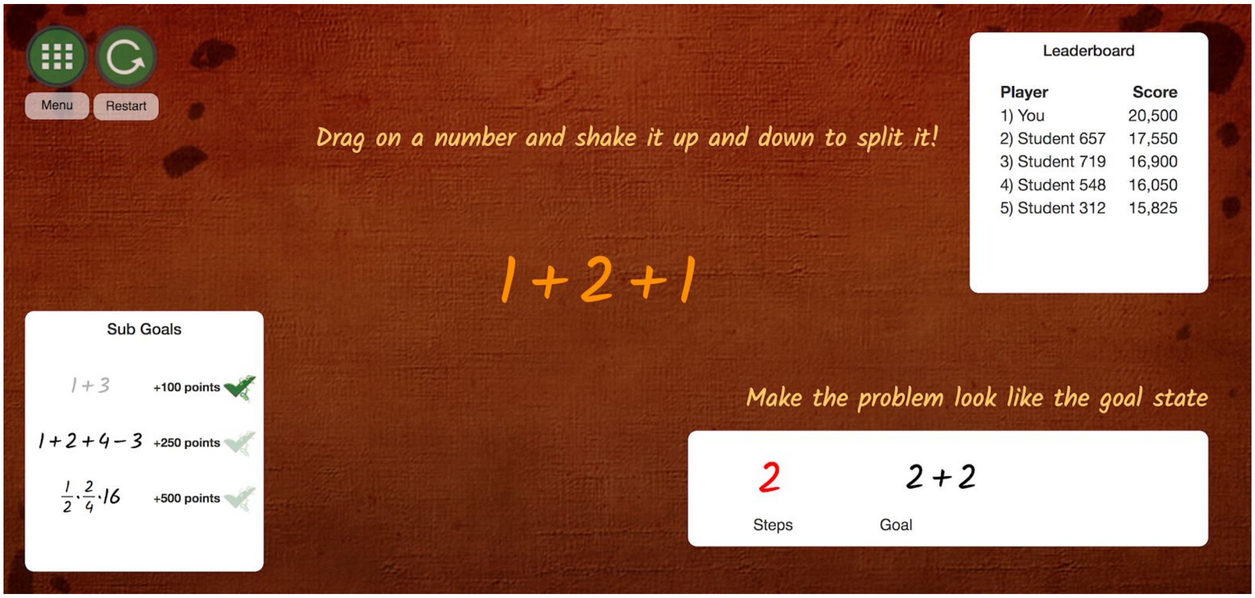


Figure 7.4. A screenshot of the subgoal state goals on the bottom left of the screen as well as the leaderboard with the participant currently in the top rank on the top of the right screen

All conditions were given the same 42 problems to solve that overall increase in difficulty. Each problem featured an initial state, three subgoals, and a main goal state. Participants were instructed to manipulate the problem to make it look like the main goal state to solve the problem using the gestures that they learned in the training session. The three subgoals varied in difficulty and were worth 100, 250, and 500 points, respectively. Simply solving the problem awarded 250 additional points. Players must reach the subgoal and then solve the main goal state without resetting the problem. Resetting the problem would reset any previously reached subgoals. Upon completing the problem, the acquired points would be awarded to the player's score, however, if they went back and reattempted the same problem, it would only change their score if their total awarded points was greater than any previous attempt for that problem. All conditions sorted each of the 42 problems or "levels" into one of four worlds -

Easy, Hard, Insane, and Impossible. Each level was unlocked when you completed the previous one. After completing the 8th level of a world, the next world was unlocked and a reward text was displayed to the player. After completing the intervention, participants are then logged out and redirected to an unrelated gamified learning environment for the remainder of their hour. This was designed so that participants would not rush through the first task with the intention of ending early, but instead allowing them to freely choose when they no longer wanted to complete this task given their allotted timeslot.

Table 7.1. A description of each of the four conditions of the intervention

Condition	**Description**
No Competition	FH2T! intervention featuring subgoal states and points. The participant's score is listed by itself on the leaderboard
Always Leading	FH2T! Intervention featuring subgoal states and points. The participant's score is always in the top ranking (1st) on the leaderboard regardless of performance.
Always Trailing	FH2T! Intervention featuring subgoal states and points. The participant's score is always in the bottom ranking (5th) on the leaderboard regardless of performance.
Back and Forth/Fluctuating	FH2T! Intervention featuring subgoal states and points. The participant's score fluctuates between rankings on the leaderboard regardless of performance.

7.2.5 Measures

1) **Interest in math puzzle questionnaire.** Participants completed a 13-question questionnaire that helped to measure their likely interest and engagement in math puzzles and activity. The first two questions asked the number of math courses they have taken as well as their self-assessment of their ranking in class (i.e. below average, best in class, etc.). This remainder of the questionnaire was modeled after the questionnaire used in Durik & Harackiewicz (2007) and consisted of 11 statements participants would agree with on a scale of 1-7. These statements were designed to gain an understanding of the participants' valuation of mathematics and included statements such as "I find math enjoyable" or "Math just doesn't really appeal to me". For measurement, these statements were summed with a lowest possible score of 11 and highest possible score of 77. For the negatively inflective statements such as "Math just doesn't really appeal to me", these were reverse scored in the total sum. The participants' scores were then averaged for data analysis as our factor of interest in math puzzles.

2) **Engagement.** Engagement is our main focus of the study and was measured using a combination of data recorded in the From Here to There web application.

   a) **Time spent in game.** Once the participant finishes the training tutorial, they are automatically logged in to the FH2T intervention and a timestamp is recorded for that individual participant. When the participant finishes the intervention by clicking the "I don't want to continue" link at the top of the screen, the system records a second timestamp indicating their logout time. Once the participant logouts, they are not able to solve any more problems and are redirected to a post-survey. We used these two timestamps to measure the amount of time spent in the app. Because the study was designed so that a student was free to stop the application at any point, we feel that the amount of time spent in the intervention is a good indication of engagement. The participants that willingly chose to continue playing the intervention would have a higher amount of time spent in the intervention thus demonstrating more engagement than a participant that stopped after only a few minutes.

b) **Number of distinct problems completed.** Similar to time spent, the number of problems that a participant completed in the intervention is useful as it refines the perspective of the "time spent" measure. The intervention had 42 problems increasing in difficulty. While the amount of time spent in the intervention can show us one facet of engagement, the number of problems can be used as an indicator of engagement to show how far a participant progressed in the intervention.

c) **Number of attempts.** When solving a problem, a participant can reset the problem as many times as they like before completing it. When the participant starts a problem, they are on attempt #1 and with each reset, the number of attempts increases. We can use the number of attempts a participant makes in the intervention as a form of engagement.

d) **Number of "go-backs".** The intervention awards points based on the user's performance. A participant is awarded points when they match any of the three subgoal as well as when they complete a problem. The three subgoals vary in difficulty and are worth 100, 250, and 500 points, respectively. Simply solving the problem awarded 250 additional points. Players must reach the subgoal and then solve the main goal state without resetting the problem to earn points. Resetting the problem would reset any previously reached subgoals. Upon completing the problem, the acquired points would be awarded to the player's score, however, if they went back and reattempted the same problem, it would only change their score if their current total awarded points were greater than any previous attempt for that problem. However, this intrinsic motivation to go back and complete a problem for a total possible score of 1,100 points (250 for the main goal and 850 for all three subgoals) can be used as a measure of engagement. We can calculate the number of "go-back attempts" by calculating the total number of problems completed and then subtracting it by the number of distinct problems completed.

e) **Number of subgoals completed.** Subgoals are optional achievements that players can meet while solving a problem. The number of subgoals that a participant met in the intervention is useful as it refines the perspective of the "time spent" measure. The intervention had 42 problems with 3 subgoals per problem.

While these variables will distinguish specific in-game activity, we will focus our analysis and how we measure in engagement using the amount of time spent using the technology (elapsed time). We will also be analyzing at what point in the intervention did participants log out as well as their final position on the leaderboard upon quitting, if applicable. Our initial hypothesis is that participants in the always trailing condition will spend the least amount of time in the intervention due to the frustration of not being able to ascend to a self-imposed acceptable standing, the always leading condition will spend more time in the intervention ascending to the top of the leaderboard where they can easily maintain their lead before becoming bored, and the fluctuating condition will spend the most time due to the constant challenges of leading and trailing to maintain their competitive standing.

7.2.6 Analysis

Research Question 1. *Do "subgoal states" and pointification lead to higher engagement of the intervention?*

As stated earlier, subgoal states and points for achieving these states are a new component of From Here to There! that has not been implemented or tested in any of our other studies using this application. We are interested in exploring whether or not these components will have any effect on user engagement. Just as the extrinsic rewards for clovers in Part II, Chapter 2 led to more time spent in the intervention than the condition that did not have any rewards or clovers, we are curious to see if providing additional goals and rewarding points will work in the same manner. We have kept the same problems, interface, and application that was used in Chapter 2 to examine gamification vs non-gamification versions of FH2T!. Thus, we ran the same study using FH2T! featuring leaderboards to see if this has any impact or effect on engagement. To do so, we will measure engagement by the amount of time spent in the intervention before quitting using our condition that has no competitors on the leaderboard. This will provide us with an isolated perspective on the effects of gaining points and solving subgoal states. We ran basic descriptives as well as an independent samples t-test against the data from Chapter 2, as it is from the same population pool using the same intervention with the same problems. When compared to the gamification condition of the previous chapter, the no competitors condition allows us to isolate the effects of the subgoal states and pointification as there are no additional effects involving social comparison or competition that are present in the other three conditions. We expect that there will be a slightly higher measure of engagement for the subgoal states condition than we previously found in Part II, Chapter 2.

Research Question 2. *Does the social/competitive feature present in the leaderboard component of the intervention further promote engagement than not having competitive virtual peers present?*

As a whole, we are interested to see the effects of presenting a leaderboard to the participant. Much like the comparison for RQ1, we are interested in seeing if this component engages or disengages users from the intervention. Though humans are competitive in nature, it is worth exploring whether or not these types of components which promote competition, though unbeknownst to the user is fictitious in this project, discourages users from engaging when compared to the same intervention that does not include such features in its design. We find this interesting because many games both in entertainment and in education often include scores and allow for friendly competition among players or students. However, this competitive aspect may not be embraced in the way that it was designed to be. We will examine the total amount of time that participants interacted with the intervention to determine engagement. We will analyze this data similarly to RQ1 in that we will run basic descriptives to compare means as well as run independent samples t-tests against the data from Part II, Chapter 2 to see how it compares. We expect that social competition as a whole will lead to longer interactions with the system than recorded in the gamified version featured in Part II, Chapter 2. However, depending on the results of RQ3, there may be certain conditions that might fall short of the interaction times of the previous chapter.

Research Question 3. *Do certain standings on the leaderboard promote or discourage engagement more than others?*

As mentioned earlier, participants will be randomly assigned to a condition where their scores will either be always ahead of their fictitious competition, always behind, or fluctuating back and forth on the leaderboards. We designed this intervention with the intention of seeing whether or not this social competition component of gamification discourages users from interacting with the application. It is our suspicion that participants in the always leading condition will not feel challenged enough because they will maintain a significant lead over their peers regardless of their true performance and ability. We believe that this will make them grow bored of the application and ultimately stop the task prematurely. Likewise, we feel that the always trailing condition will become frustrated and overwhelmed that they cannot compete with their peers and will want to stop even earlier than the always leading condition due to discouragement. On the contrary, we suspect that the fluctuating condition will allow the participant to feel the struggle of being behind in the leaderboards, but also achieve the feeling of success while leading their peers even for a short time, ultimately leading to longer interaction times than the other two leaderboard conditions. Like the other two research questions, we will be comparing the descriptives and performing an ANOVA across the conditions to further examine the effects of each intervention. We hope that this will provide a more conclusive answer into the effects of social competition in our educational technology. Furthermore, we plan to examine the positions on the leaderboards when participants quit to see if there are any additional patterns regardless of condition. Because we monitor and record each attempt that a participant occurs, we can see if these positions themselves have any effect on *when* a user decides to stop.

7.2.7 Procedures

At the start of the study, each participant was to complete a questionnaire, a short tutorial page (Figure 7.3), and two main tasks. They were allowed to spend as much time or as little time on the first task (FH2T intervention) as they would like up to one hour, but were required to spend the remainder of the hour completing a second unrelated task. The study was designed this way so the students felt the freedom to explore the intervention as long as they were engaged, but also dedicate the full hour toward the study so that they were not tempted to leave early. Before using the system, participants were asked to complete an online questionnaire assessing their interest in mathematics. The first question asked how many math courses they have taken ranging from Algebra I to Calculus IV and higher with an option to add additional courses. These questions targeted their overall exposure to mathematics and indicates whether or not they are in a math-related program in their current studies. They were also asked to assess their general standing in their math classes as worst in class, below average, average, above average, or best in class. Lastly, the questionnaire gave 11 statements regarding math and math activities that participants were to agree or disagree with on a 7-degree scale from strongly disagree (1) to strongly agree (7). This targeted participant interest in math using positively inflective statements such as “I find math enjoyable”, “Puzzles like Sudoku are a fun way to pass the time” and negatively inflective statements such as “I rarely try math problems not required by the teacher”. This was the same questionnaire that we included in the study on gamification discussed in the previous chapter. This questionnaire was included in this study to keep the study procedures similar between the two studies. Though the questionnaire may be used in future analysis, the emphasis of this research will focus on analyzing the in-game data and identifying student behavior based on condition.

After completing the questionnaire, each participant was directed to a tutorial web-page that showed how to use the proprietary gestures and movements of the application that will be needed to play *From Here to There!*. There are 7 pages of tutorials with roughly 4 examples on each page demonstrating how to manipulate terms and perform mathematical operations to solve math problems. The tutorial goes through basic interactions of the software, but does not promote any particular strategy in solving problems nor does it provide any additional practice in solving problems. On completing the tutorial, the participant was redirected to a specific version of the intervention based on the participant's condition. All four groups were then automatically logged into the web version of *From Here to There!*. Though the leaderboard component is different for each of the four conditions as described above, all four conditions were asked to solve the same problems in the same order. A timestamp was logged from the moment they started the application. All versions of the program had a link on the top of each screen that read "I don't want to continue" that participants were instructed to click when they were finished. When the link was clicked, the user was logged out and a timestamped event was logged indicating that the user finished. This corresponds to the timestamped event that was created when a user initially logged in. If they did not click the link, they were automatically logged out after the hour was complete.

7.3 Results

RQ1, "Do "subgoal states" and pointification lead to higher engagement of the intervention?"

Two hundred (200) undergraduate students from a private engineering university in the Northeastern United States participated in the hour-long study focusing on leaderboards in the gamified version of FH2T. Of these 200 students, 32 students experienced technical issues that may have caused them to restart the intervention and thus were removed from the analysis. Of the 168 participants that were included in the analysis, 39 participants were in the no leaderboard condition; 39 participants were in the always leading condition; 45 participants were in the always trailing condition; and 45 participants were in the fluctuating condition. We began our analysis in an effort to answer RQ1. As mentioned earlier, the intervention is similar to the study discussed in the previous chapter with the addition of subgoal states and leaderboards. However, in order to accurately measure the difference in engagement between the original version and the newly modified version of FH2T, we only compared the data from the gamified condition of the previous study and the no leaderboards condition of the current study. Both versions of the intervention used the same problems, worlds, and gameplay as well as the same intervention design. This was intentionally kept the same so the two interventions could be directly compared. The data from each study were aggregated and were compared against the duration of the time spent in the intervention, the number of distinct problems completed, the total number of problems completed, and the number of problem attempts overall. An independent samples t-test was used to compare these measures between the two groups and the results are found in Table 7.3.

Table 7.2

Descriptives of subgoal states (control) vs. no subgoal states (experimental) in FH2T

Measure	Group	N	Mean	Std. Dev.	Std. Error
Elapsed Time	No Subgoals	22	2047.409	887.5190	189.2197
	Subgoals	39	2205.037	898.7617	172.9668
Distinct Completed	No Subgoals	22	28	12.2513	2.612
	Subgoals	39	23.519	10.8606	2.0901
Total Completed	No Subgoals	22	30.273	14.9226	3.1815
	Subgoals	39	26.037	12.2741	2.3621
Highest Completed	No Subgoals	22	30.500	13.412	2.8594
	Subgoals	39	26.222	12.6349	2.4316
Total Attempts	No Subgoals	22	59.773	38.6707	8.2446
	Subgoals	39	33.074	14.7568	2.84

Table 7.3

Independent samples t-test of subgoal states vs. no subgoal states in FH2T

Measure	Group	*N*	*M*	*SD*	*t*	*p*
Elapsed Time	No Subgoals	22	2047.409	887.5190	-0.614	.542
	Subgoals	39	2205.037	898.7617		
Distinct Completed	No Subgoals	22	28	12.2513	1.356	.181
	Subgoals	39	23.519	10.8606		
Total Completed	No Subgoals	22	30.273	14.9226	1.091	.281
	Subgoals	39	26.037	12.2741		
Highest Completed	No Subgoals	22	30.500	13.412	1.147	.257
	Subgoals	39	26.222	12.6349		
Total Attempts	No Subgoals	22	59.773	38.6707	3.310	.002
	Subgoals	39	33.074	14.7568		

RQ2, "Does the social/competitive feature present in the leaderboard component of the intervention further promote engagement than not having competitive virtual peers present?"

We also explored the effects of the leaderboard component of FH2T that we developed for this study. As discussed earlier, the leaderboard component fictitiously places participants into predetermined rankings during the study. All participants were grouped into one of four possible conditions: no leaderboard, leaderboard where they are always leading, leaderboard where they are always trailing, or leaderboard where their ranking fluctuates. For this analysis, we will be analyzing the differences between the no leaderboard condition and the other three conditions using FH2T's in-app measures to answer Research Question #2 (RQ2). We aggregated the data points for all participants and calculated the number of distinct problems completed, the number of problems completed in total, the highest problem completed, and the total number of attempts made during the intervention. Lastly, we ran an independent samples t-test comparing the no leaderboard condition to the remaining three conditions against these in-app measures and presented the results in Table 7.5.

Table 7.4

Descriptives of leaderboard (experimental) vs. no leaderboard (control) in FH2T

Measure	Group	*N*	*M*	*Std. Dev.*	*Std. Error*
Elapsed Time	No Leaderboard	39	2078.73	1573.968	258.759
	Leaderboard	129	1830.32	1404.723	125.642
Distinct Completed	No Leaderboard	39	19.65	11.509	1.892
	Leaderboard	129	17.30	10.713	0.958
Total Completed	No Leaderboard	39	21.73	12.989	2.135
	Leaderboard	129	19.97	13.057	1.168
Highest Completed	No Leaderboard	39	21.73	13.391	2.201
	Leaderboard	129	18.63	11.623	1.040
Total Attempts	No Leaderboard	39	27.35	16.158	2.656
	Leaderboard	129	26.39	17.996	1.610

Table 7.5

Independent samples t-test of leaderboard conditions vs. no leaderboard condition in FH2T

Measure	Group	*N*	*M*	*SD*	*t*	*p*
Elapsed Time	No Leaderboard	39	2078.73	1573.968	0.919	0.360
	Leaderboard	129	1830.32	1404.723		

Distinct Completed	No Leaderboard	39	19.65	11.509	1.150	0.252
	Leaderboard	129	17.30	10.713		
Total Completed	No Leaderboard	39	21.73	12.989	0.722	0.471
	Leaderboard	129	19.97	13.057		
Highest Completed	No Leaderboard	39	21.73	13.391	1.374	0.171
	Leaderboard	129	18.63	11.623		
Total Attempts	No Leaderboard	39	27.35	16.158	0.291	0.771
	Leaderboard	129	26.39	17.996		

RQ3), "Do certain standings on the leaderboard promote or discourage engagement more than others?"

Finally, we wanted to identify any behavioral changes that may be present when examining the effects of the fictitious leaderboard. As a reminder, the leaderboard is present in three of the four conditions and is designed to keep the participant always ahead, always in last, or fluctuating based on their assigned conditions. Ultimately, we attempted to use this analysis to answer Research Question 3 (RQ3). For this exploratory analysis, we examined the data for mainly the fluctuating condition as that is the only condition where players are constantly shifting back and forth between ranks where we can observe behavioral changes in game strategy and other performance measures. We also looked at the amount of subgoals and overall scores earned during problem attempts and examined interactions with their leaderboard rankings at the time to see if certain rankings or positions were linked with specific patterns. Lastly, we observed similar measures across the three leaderboard conditions to see if the differences were causal due to the shifting upward and downward between ranks or simply being in specific ranks themselves. The descriptive and ANOVA results are featured in Table 7.7.

Table 7.6

Descriptives of fluctuating leaderboard attempts vs. no leaderboard attempts in FH2T

Measure	Group	*N*	*M*	*Std. Dev.*	*Std. Err.*
Step Count	No Leaderboard	39	7.965	4.141	0.663
	Fluc. Leaderboard	45	9.899	6.733	1.004
Total Steps	No Leaderboard	39	9.585	4.679	0.749
	Fluc. Leaderboard	45	11.853	7.730	1.152
Total Attempts	No Leaderboard	39	1.246	0.232	0.037
	Fluc. Leaderboard	45	1.305	0.325	0.048
Overall Score	No Leaderboard	39	386.339	210.335	33.681
	Fluc. Leaderboard	45	448.585	229.795	34.256
Subgoals	No Leaderboard	39	0.636	0.750	0.120
	Fluc. Leaderboard	45	0.876	0.836	0.125

Table 7.7

Independent samples t-test of fluctuating leaderboard vs. no leaderboard in FH2T

Measure	Group	*N*	*M*	*SD*	*t*	*p*
Step Count	No Leaderboard	39	7.965	4.141	-1.556	0.124
	Fluc. Leaderboard	45	9.899	6.733		
Total Steps	No Leaderboard	39	9.585	4.679	-1.596	0.114
	Fluc. Leaderboard	45	11.853	7.730		
Total Attempts	No Leaderboard	39	1.246	0.232	-0.956	0.342
	Fluc. Leaderboard	45	1.305	0.325		
Overall Score	No Leaderboard	39	386.339	210.335	-1.287	0.202
	Fluc. Leaderboard	45	448.585	229.795		
Subgoals	No Leaderboard	39	0.636	0.750	-1.376	0.173
	Fluc. Leaderboard	45	0.876	0.836		

Table 7.8

Descriptives of attempts based on current ranking

Measure	Ranking	*N*	*M*	*Std. Dev.*	*Std. Error*
Overall Score	1st	8	477.083	394.726	139.557
	2nd	18	558.565	375.480	88.502
	3rd	28	508.384	345.986	65.385
	4th	40	440.069	280.183	44.301
	5th	45	454.461	254.769	37.979
	Total	139	475.965	305.287	25.894
Subgoals	1st	8	1.135	1.119	0.396
	2nd	18	1.291	1.247	0.294
	3rd	28	0.969	1.184	0.224
	4th	40	0.823	0.964	0.152
	5th	45	0.939	0.884	0.132
	Total	139	0.969	1.031	0.087
Elapsed Time	1st	8	38.510	29.926	10.580
	2nd	18	51.903	44.831	10.567
	3rd	28	44.872	24.275	4.586
	4th	40	63.981	38.594	6.102
	5th	45	115.213	87.984	13.116
	Total	139	73.688	64.724	5.489
Total Steps	1st	8	7.156	5.132	1.814
	2nd	18	9.463	7.907	1.864
	3rd	28	8.562	4.802	0.908
	4th	40	10.239	5.343	0.845
	5th	45	13.019	8.557	1.276
	Total	139	10.523	6.972	0.591
Total Attempts	1st	8	1.167	0.236	0.083
	2nd	18	1.088	0.247	0.058
	3rd	28	1.159	0.424	0.080
	4th	40	1.124	0.390	0.062
	5th	45	1.369	0.464	0.069
	Total	139	1.242	0.409	0.035

Table 7.9

Analysis of Variance (ANOVA) of all ranks in the fluctuating condition in FH2T

Measure	Group	*Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*
Overall Score	Between Groups	224597.195	4	56149.299	0.595	0.667
	Within Groups	12636993.3	134	94305.920		
	Total	12861590.4	138			
Subgoals	Between Groups	2.992	4	0.748	0.698	0.595
	Within Groups	143.674	134	1.072		
	Total	146.666	138			
Steps	Between Groups	246.400	4	61.600	1.566	0.187
	Within Groups	5271.294	134	39.388		
	Total	5517.694	138			
Total Steps	Between Groups	502.104	4	125.526	2.711	0.33
	Within Groups	6205.370	134	46.309		
	Total	6707.474	138			
Total Attempts	Between Groups	1.394	4	0.349	2.155	0.077
	Within Groups	21.676	134	0.162		
	Total	23.070	138			

Table 7.10

Descriptives across all four conditions for attempts in FH2T

Measure	Ranking	*N*	*M*	*Std. Dev.*	*Std. Error*
Overall Score	None	39	386.339	210.335	33.681
	Leading	39	364.183	197.464	31.619
	Trailing	45	389.711	227.589	33.927
	Fluctuating	45	448.585	229.795	34.256
	Total	168	398.772	217.958	16.816
Subgoals	None	39	0.636	0.750	0.120
	Leading	39	0.600	0.684	0.109
	Trailing	45	0.664	0.763	0.114

	Fluctuating	45	0.876	0.837	0.125
	Total	168	0.699	0.764	0.059
Steps	None	39	7.965	4.141	0.663
	Leading	39	8.119	5.198	0.832
	Trailing	45	8.157	4.648	0.692
	Fluctuating	45	9.899	6.733	1.004
	Total	168	8.570	5.324	0.411
Total Steps	None	39	9.585	4.679	0.749
	Leading	39	10.195	7.577	1.213
	Trailing	45	10.121	5.312	0.792
	Fluctuating	45	11.853	7.730	1.152
	Total	168	10.477	6.479	0.499
Total Attempts	None	39	1.245	0.232	4.141
	Leading	39	1.306	0.429	5.198
	Trailing	45	1.309	0.371	4.648
	Fluctuating	45	1.305	0.325	6.733
	Total	168	1.292	0.345	5.324

Table 7.11

Analysis of variance across all four conditions for attempts in FH2T

Measure	Group	*Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*
Overall Score	Between Groups	168041.721	3	56013.907	1.183	0.318
	Within Groups	7765374.72	164	47349.846		
	Total	7933416.44	167			
Subgoals	Between Groups	2.001	3	0.667	1.144	0.333
	Within Groups	95.590	164	0.583		
	Total	97.591	167			
Steps	Between Groups	109.359	3	36.453	1.293	0.279
	Within Groups	4623.645	164	28.193		
	Total	4733.005	167			
Total Steps	Between Groups	125.045	3	41.682	0.993	0.398
	Within Groups	6884.248	164	41.977		
	Total		167			

		7009.293				
Total Attempts	Between Groups	0.114	3	0.038	0.317	0.813
	Within Groups	19.721	164	0.120		
	Total	19.835	167			

7.4 Discussion

7.4.1 Outcome

For Research Question 1 (RQ1), "Do "subgoal states" and pointification lead to higher engagement of the intervention?", we compared the data from the gamification condition of the previous chapter's study to that of the non-leaderboard condition of this study to isolate the effects of having subgoal states without the leaderboard component. The independent samples t-test that was used to compare these two groups examined the in-app measures of the application. Although the subgoals condition spent an average of 3 minutes longer in the intervention, the data suggests that the subgoals condition completed far fewer distinct problems, overall problems, and made far fewer attempts than the condition without subgoals. Thus, the subgoals led to more engagement on the problem, but less overall progress in the app despite playing for a longer amount of time. However, the statistical data was not statistically significant for any of these in-app measures except for attempts. We interpret these findings to conclude that although they may not be engaged any further than the gamified condition, the subgoals themselves are engaging and result in the participant to proceed slower through the application, thus resulting in fewer attempts, but with increased effort.

Similarly, we examined Research Question 2 (RQ2), “Does the social/competitive feature present in the leaderboard component of the intervention further promote engagement than not having competitive virtual peers present?”. To do so, we compared our condition without leaderboards to the remaining three conditions with leaderboards. We hypothesized that the leaderboard aspect of the application might engage players to try harder to rank up or maintain a leader while also solving problems. To test this, we analyzed the in-app data to explore any changes in behavior from the leaderboard conditions to the no leaderboard condition. The data in Table 7.3 suggests that similarly to the results of RQ1, the no leaderboard condition was able to complete more problems and progress through the application more than the conditions with leaderboards. However, these findings were also not statistically significant like the findings of RQ1. We interpret these findings to suggest that the leaderboard component of the application does affect behavior and ultimately slows the participant’s progress down even further like subgoals do. The engagement that does occur from the leaderboard component seems to slow down the pace of completion, however, the elapsed time and the rate of attempts are pretty consistent between the conditions.

Finally, we took a closer look at the in-game data and behaviors exerted in the application when answering Research Question #3 (RQ3), “Do certain standings on the leaderboard promote or discourage engagement more than others?”. For this analysis, we examined the data belonging to the leaderboard conditions that artificially fluctuates the leaderboard condition. First, we wanted to see how each of the conditions compared using the in-app measures. Our ANOVA results suggest that the always fluctuating condition completed the most subgoals than any of the other three conditions followed by the always trailing condition, the no leaderboard condition, and finally the always leading condition. These findings are also present in the same order when examining the amount of points earned per problem as well. This suggests that the fluctuating condition promoted a more competitive environment that motivated participants to complete more subgoal tasks and earn more points overall. While in the always leading condition, however, players did not feel motivated to earn points or complete subgoal tasks. We believe that this is because the players did not feel challenged by the artificial players because they were always in the lead regardless of their effort. Furthermore, the lack of competition seemed to discourage players from attempting subgoals or earning points when compared to the no leaderboard condition who averaged a much higher amount of each. We believe that these findings suggest that competition needs to be carefully balanced to ensure maximum motivation and engagement. Too much competition will lead to some motivation, but not enough competition will disengage participants to the point where they were better off having no competition. Thus, a fluctuating amount of competition seems to motivate participants the most and keep them engaged. In addition to the previous in-game measures that we have used up to this point, we explored some of the decision-based measures that are associated with the leaderboard. More specifically, we looked at the number of subgoals attempted when players are at various rankings on the leaderboard as well as the amount of points that each of the subgoals that they were attempting while in those ranks. As a reminder, the leaderboard rankings of this condition artificially fluctuate throughout the intervention from being highest (1st place) to lowest (5th place) to give the participant the illusion of having equally skilled competitors. The data in Table 7.5 suggests that the amount of completed subgoals are directly tied with their respective rankings. When players are in the 2nd place position, they complete more subgoals than any other position while the players in the lowest ranking complete the fewest amount of subgoals. However, while the number of subgoals increases with rank, the number of overall points gained on a problem peaks when players are in second and third place, drops slightly when players are leading in first place, and continues to drop significantly as they rank lower to fourth and fifth place. We believe players are attempting more difficult subgoals to earn higher point values per attempt while trying to rank up, but once they are in first, they are trying to maintain and coast off of their lead by going for the easier subgoals which yield less points, but require less time to complete. This trend continues in elapsed time per problem and the number of steps per problem where first place players spend the least amount of time and use the fewest number of steps which also gradually increases as players rank lower on the leaderboard. On the other end, players seem to not want to apply themselves as much when they are ranked in fourth and fifth place; they strive for fewer subgoals and earn fewer points resulting in their perpetually low rankings. However, this behavior does change in the fluctuating condition because the rankings are manipulated.

7.4.2 Limitations and Future Work

Our initial testing of social competition in a gamified learning environment proved to yield some interesting results. While there were no statistically significant changes in engagement from simply adding subgoals, it appeared that in-game progress slows down slightly when introducing subgoals. There were also no statistically significant changes simply by adding a leaderboard to the application, however, anecdotal analysis of the data suggests that this also slows participant progression through the app. Furthermore, when exploring the in-app behaviors of the leaderboard component itself, the fluctuating condition outperformed the other three conditions in the number of subgoals attempted and points earned. Meanwhile, the always leading condition attempted the fewest subgoals and earned the least amount of points of the four conditions including the condition that did not have a leaderboard. Thus, we believe that the lack of equal competition was actually demotivating to participants than it would have been if they had no competition. Finally, the fluctuating condition exerted a few interesting behaviors when players shuffle between rankings. Players attempt more subgoals when they are in higher rankings and decrease as they rank down the leaderboard. However, the points associated with these subgoals are not the same. Players in the middle of the leaderboard strive for harder subgoals that yield higher points, while first place players go for easier subgoals that they will be able to solve quickly. Overall, the subgoal component does seem to be an easy extension of the application that can further promote number exploration as an extension of the original application. While it may not significantly lead to any additional engagement, it does seem to engage as well and can be used as additional content to the application. The leaderboard, however, can have negative effects on a player's engagement and motivation and should only be implemented in a controlled competitive environment where it would lead to a positive experience.

References

Braith, L., Daigle, M., Manzo, D., & Ottmar, E. (2017). Even Elementary Students Can Explore Algebra!: Testing the Feasibility of from Here to There!, a Game-Based Perceptual Learning Intervention. American Psychological Society Conference, Boston, MA.

Christy, K. R., & Fox, J. (2014). Leaderboards in a virtual classroom: A test of stereotype threat and social comparison explanations for women's math performance. *Computers & Education*, *78*, 66-77.

Deci, E., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Springer Science & Business Media.

Durik, A. M., & Harackiewicz, J. M. (2007). Different strokes for different folks: How individual interest moderates the effects of situational factors on task interest. Journal of Educational Psychology, 99(3), 597.

Festinger, L. (1954). A theory of social comparison processes. *Human relations*, *7*(2), 117-140.

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*.

Sarason, S. B. (1983). *Schooling in America: Scapegoat and salvation*. New York: Free Press; London: Collier Macmillan.

Wells, B. M., & Skowronski, J. J. (2012). Evidence of choking under pressure on the PGA tour. *Basic and Applied Social Psychology*, *34*(2), 175-182.

8 STUDY 3: THE EFFECTS OF GAMIFICATION ON LEARNING IN THE ELEMENTARY CLASSROOM

The following chapter includes two published efforts from the same study: Braith, Daigle, Manzo, and Ottmar (2017) and Hulse et al. (2019). Originally submitted as a senior project by a member of the educational psychology lab, this work explores the feasibility of gamified and non-gamified versions of From Here to There! Web that were modified to be suitable for elementary school students. In addition to its feasibility, the findings suggest that student learning gains improved just as much using the learning applications as the controlled worksheet condition. While the original study found no significant learning differences between the gamified and non-gamified versions of the intervention, further analysis using in-app measures suggests that the gamified condition elicits strategies and behaviors associated with increased learning gains.

8.1 Introduction

8.1.1 Overview

In early elementary school, students begin to understand mathematics as numbers by creating concrete representations, and often do not recognize the flexible potential or function of the digit (Carr et al., 2011). In a child's early years of education, math is focused primarily on the recognition of patterns in numerical expressions, and their ability to apply these patterns to other math expressions (Lins & Kaput, 2004). While these patterns are taught early on in school, students are not readily introduced to the relation or flexibility within the patterns. By second grade, children often gain the capacity to cognitively represent these numbers do so through abstract reasoning of numbers and their relations to one another (Carr et al., 2011). More broadly, when students are provided with ample opportunity to practice new strategies and understand the effectiveness of these new strategies then they are able to acquire and use these strategies independently (Bay-Williams, 2001; Carr et al., 2011). These cognitive representations of number patterns allow students to think flexibly about numbers. This decomposition then allows them to manipulate expressions into something that they can more easily reason through. However, the educational topics taught in schools are often introduced as standalone topics and do not facilitate much understanding on flexibility and reasoning. While efforts such as Project LEAP are primarily focused around the fundamental belief that the practice of early algebra education is critical to success in mathematics later in life (Lins & Kaput, 2004; Carraher & Schliemann, 2008), it is generally recognized that early introduction to algebra is relatively untested. Thus, if decomposition of numbers was introduced at an earlier age, as a precursor to formal algebra learning, children could potentially increase their conceptual understanding of number relations and flourish in their flexibility and number reasoning.

8.1.2 From Here to There! Elementary

As discussed in earlier chapters, From Here to There! Web was developed through an iterative design process that allowed it to be accessible from any computer while still maintaining its data integrity and customizations for research. In an effort to explore the efficacy of From Here to There! as a learning intervention, our research team developed an elementary school version of the app, From Here to There! Elementary (FH2T-E). While most of the studies that we have discussed up until this point have largely ignored learning gains and focused primarily on the effects and behaviors of our gamified interventions, we decided that we needed to implement a version of an intervention for a demographic that was relatively new to the content in an effort to gauge our app's efficacy as a learning technology. While the structure of the application would remain largely intact, we shifted the content focus from basic algebra to basic number reasoning and number decomposition. While the classic version of FH2T! would progress through operations such as factoring, distribution, and Euclidean's algorithm, From Here to There! Elementary introduces commuting and decomposing using basic operators such as addition, subtraction, multiplication and addition.

8.1.3 Research Goals

In this chapter, we examine the effectiveness of early introduction of number sense and algebraic principles using From Here to There! Elementary (FH2T-E). In addition to the pre-test and post-test assessments that were given to participants, we leverage student log data created during mathematical problem solving in FH2T-E to reveal constructs of proficiency and answer the following research questions:

1) Can we establish feasibility of FH2T-E for introducing students to early mathematics concepts?

2) Were there differences in math posttest performance between students who received the gamified and non-gamified conditions?
3) Do in-app measures of student problem solving process predict learning gains?

8.2 Methodology

8.2.1 Participants

One hundred eighty five (N=185) second grade students from ten classrooms in three different elementary schools in Massachusetts (116 female, 78 male) participated in this study. Students were randomized into one of two experimental conditions: the regular gamified version of FH2T-E or a non-gamified version of FH2T-E. In the gamified condition, students played through the version of the game that possessed game-like features. The non-gamified version of FH2T-E featured the exact same problems and content of the gamified version and was stripped down to display only the 18 math problems within each level. As participants progressed through the non-gamified version, there was no acknowledgement of completion or rewarded points for accuracy and efficiency. This design choice was intended to assess the degree to which the learning gains stemmed from the gamification features or the goal-state dynamic approach that the FH2T-E application provided.

8.2.2 Procedures

At the beginning of the study, students were given a 15-question pre-test that assessed prior knowledge. These questions reflected second grade math standards and tested baseline understanding of decomposition, operational strategies, and basic notation. The problems and expressions on the posttest were similar to those found on the pretest. To ensure baseline equivalence, an independent-samples t-test was conducted to compare pretest scores for gamified and non-gamified conditions. There were no significant differences in pretest scores for the gamified ($M = 9.85$, $SD = 3.89$) and non-gamified ($M = 9.95$, $SD = 3.60$) conditions; $t(183) = 0.17$, $p = 0.865$. All students interacted with the app during their regularly scheduled math period during the school day for 20 minutes each day over four separate days for a combined total of 80 minutes. We feel that this amount of time is both reasonable and practical for elementary students to practice and learn these concepts using typical classroom instruction as well as 80 minutes of play using the intervention. After the four sessions of the intervention, participants were given an additional session to complete the post-test assessment which mirrored the pre-test assessment.

8.2.3 Conditions

As previously discussed, participants were randomly assigned to one of two versions of FH2T-E: the gamified version and the non-gamified version. The gamification condition's version of FH2T included the colorful backgrounds, achievements, reward messages, and levels while the plain condition's version of FH2T consisted of a directory of the problems themselves without any gamified elements, designs, rewards, or levels. Both groups were given the same problems to solve that increased in overall difficulty. Each problem featured an initial state and a goal state. Participants were instructed to manipulate the problem to make it look like the goal state using the gestures that they learned in the tutorial. Each level was unlocked upon completion of the previous one. After completing the 14th level of a world, the next world was unlocked and a reward text was displayed to the player. The participants could still continue to finish the remaining levels of that world, if they chose to. The plain condition presented each problem in a directory on the main page that increased when participants solved the previous problem. To keep the exposure to available problems consistent, the problem that would have been unlocked when solving the 14th level of a world in the gamification condition's also became unlocked when solving that same problem in the non-gamified condition.

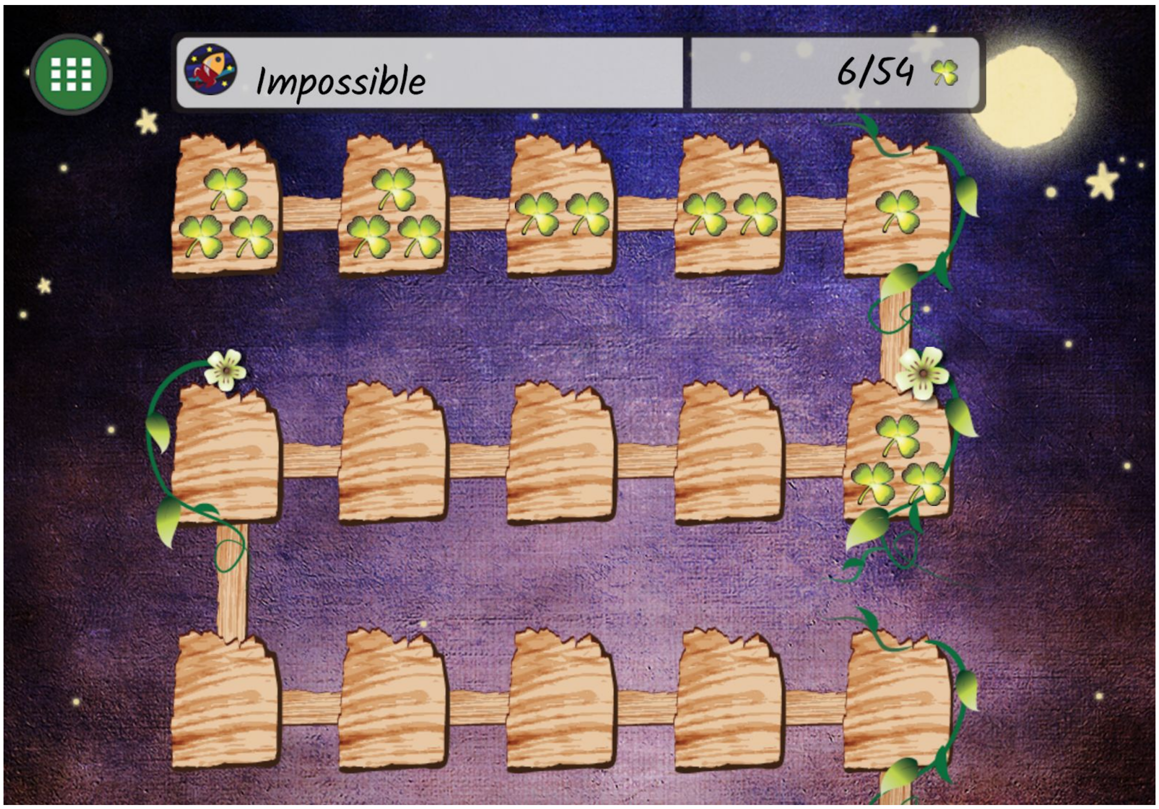


Figure 8.1. An example of the gamification condition's version of FH2T!

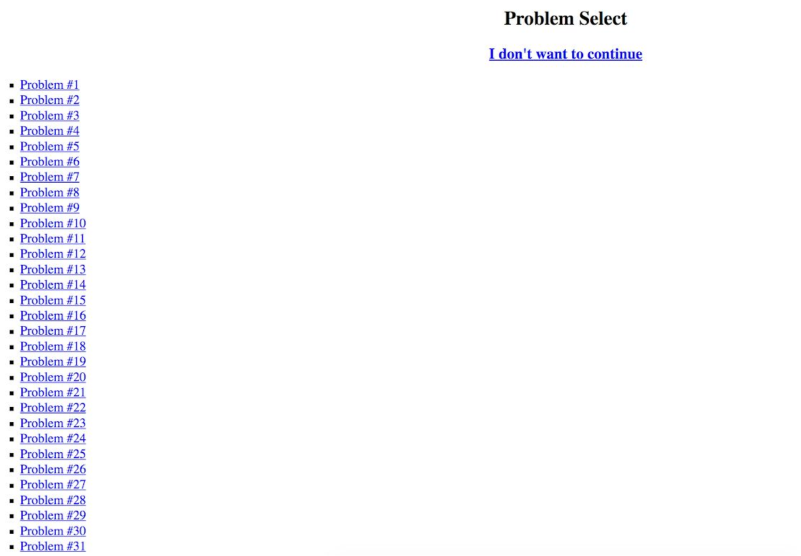


Figure 8.2. An example of the plain version of FH2T! without gamification

8.2.4 Measures

While there are hundreds of variables that we account for when participants progress through From Here to There!, the analysis of this study primarily focuses on the following measures:

1) **Pre-test/Post-test assessment.** Participants completed a 15-question pretest and posttest that helped to measure their existing knowledge of arithmetic as well as decomposition tasks. This assessment is our main focus when exploring our research questions and is administered to participants before and after they are given their condition-specific intervention.

2) **Number of distinct problems completed.** Similar to time spent, the number of problems that a participant completed in the intervention is useful as it refines the perspective of the "time spent" measure. While the amount of time spent in the intervention can show us

one facet of engagement, the number of problems can be used as an indicator of engagement to show how far a participant progressed in the intervention.

3) **Number of additional problems completed.** The intervention had over a dozen worlds of problems increasing in difficulty. When a participant completed 14 of the 20 problems available in a particular world, the next world was unlocked and able to be played. Although the additional problems of a world are completely optional, we consider the completion of the additional problems as additional practice that measures both engagement and progression in the intervention.

4) **Number of attempts.** When solving a problem, a participant could reset the problem as many times as they liked before completing it. When the participant started a problem, they were on attempt #1 and with each reset, the number of attempts increased. We can use the overall number of attempts a participant makes in the intervention as a form of engagement because it shows effort or what?.

5) **Number of "go-backs".** The gamification version of the intervention included a feature that awards clovers based on the user's performance. A participant was awarded clovers when they completed a problem. Completing a problem in the fewest number of moves awarded 3 clovers, completing a problem in two moves more than the fewest number awarded 2 clovers, and completing a problem in any additional moves awarded 1 clover. The number of clovers that a participant achieved in the game does not affect the study in

any way and the goal of maximizing the number of clovers earned serves only as an intrinsic reward. However, this intrinsic motivation to go back and complete a problem for 3 clovers can be used as a measure of engagement. We can calculate the number of "go-back attempts" by calculating the total number of problems completed and then subtracting it by the number of distinct problems completed.

8.2.5 Analysis

Upon completion of the study, our original analysis consisted of comparing pre-test and post-test assessment scores to determine the level of learning gains attributed to each intervention: control, non-gamified and gamified. Using an independent samples t-test, we explored whether there were any significant differences in learning gains between the two versions of FH2T-E and the control condition. Additionally, we ran a second independent samples t-test to determine whether there were any changes in learning gains between the gamified and non-gamified versions of the application. Lastly, an exploratory factor analysis was conducted to identify the number and structure of the factors underlying the in-app measures that were recorded within FH2T-E as students solved problems and to discover any links to learning gains between in-app measures and condition. Of the groups of factors that we analyzed, we have primarily focused on two significant factors that represent engagement and progression of our intervention.

8.3 Results

RQ1, Can we establish feasibility of FH2T-E for introducing students to early mathematics concepts?

In the initial pilot run of the study that we discussed in Chapter 4 of this dissertation, the FH2T-E program was given to 23 first grade students, all within a single classroom with a single teacher. Of the 23 first graders, 11 students used a mouse on a computer and 12 students used an iPad tablet. We examined whether there was improvement in math performance using a two-tailed t-test on the pre and post assessment mean scores. Test scores, vindicated by the total percentage of items correct, improved significantly from 63.6% to 71.8%, comparing pre- ($M= 0.636$, $SD=.26$) to post- worksheets ($M= 0.718$, $SD= .24$). There is an approximate difference of 8.1% ($SD=.33$, $\eta p^2= 0.33$, $t(1) = 1.717$, $p = 0.05$). A linear regression was also used to examine the contributions of pretest performance, gender, and input device on mathematics performance. Pretest scores ($M=0.636$, $SD=0.256$, $\beta=-0.189$) for individuals (using either input device) significantly predicted posttest scores ($M =0.718$, $SD=0.243$), accounting for 0.6 of the posttest difference, ($\beta=0.607$, $t(22) =14.21$, partial $eta^2 = 0.33$, $R^2= 0.036$, $p=0.05$, two-tailed test). No gender differences were observed; girls and boys performed similarly on pre/post-assessments. Analyses failed to reveal significant differences between computer and iPad groups, indicating the utility of both input devices. Overall, the intervention was feasible for implementation by the students and the teacher, both of whom offered positive comments at the end of the intervention. In addition to the quantitative in-app data collected from using FH2T, the teacher and in-class researchers made observations across the five days of using the intervention. The following observations recorded in regards to the students' experience using each input device. Some of the more interesting observations included: "Students were not used to the mouse movements and had difficulty clicking and selecting", "Computers pose the biggest challenge/Children struggle with using the mouse", "Out of frustration, kids will try to physically touch the computer screen and move numbers manually with their hands", and "Students who achieved higher levels were on iPads".

RQ2, Were there differences in math posttest performance between students who received the gamified and non-gamified conditions?

In addition to exploring the behavioral effects of gamification, we wanted to explore how gamified learning environments and educational games influences learning gains. In Hulse et al. (2019), we modified our app *From Here to There!* to be used by elementary school students to measure learning gains. We conducted a randomized controlled trial involving 185 elementary school students from 9 elementary schools in the northeast where students played with the elementary school version of FH2T (FH2T-E) for four (4) twenty-minute sessions across 3 weeks. Students were randomly assigned into one of three conditions: (1) a paper-based worksheet (control); (2) a non-gamified version of FH2T-Elementary where rewards, background images, and unlockables are removed; and (3) a gamified version of FH2T-Elementary. In addition to our goals involving the feasibility of deployment of the modified version of the app, we wanted to explore whether or not the exploratory nature of FH2T-E led to learning gains in both a gamified and non-gamified setting. In addition to the four sessions of the intervention, students were also given a pre-test and post-test assessment to measure learning gains. Based on the pre-test and post-test assessment data, we found that while there were significant learning gains in all three conditions, the gains of the two FH2T-E conditions were slightly higher than the gains of the control condition. However, the assessment data found that there were no statistically significant differences in learning gains between the gamified and non-gamified condition.

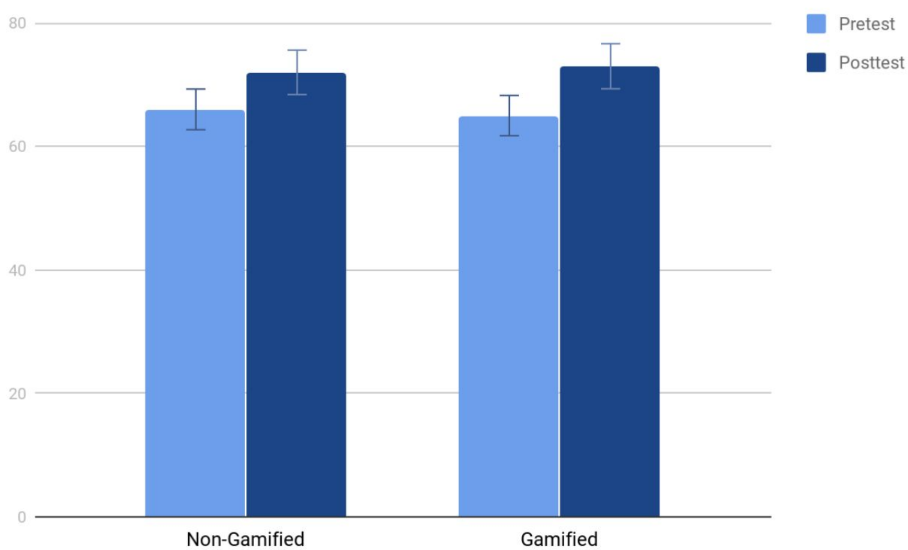
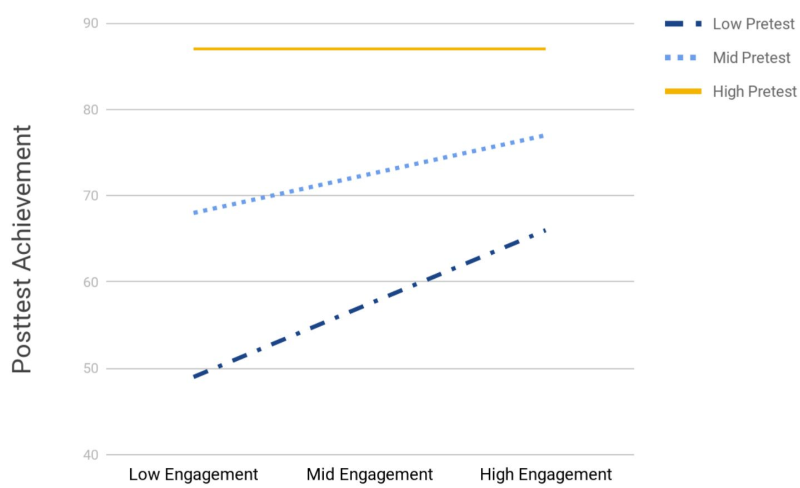


Figure 8.3. Pre-test and post-test scores comparing the non-gamified and gamified versions of From Here to There! Elementary

RQ3, Do certain student behaviors within FH2T:E differentially predict learning for high or low-knowledge students?

While RQ2 suggests that there were no difference in learning gains between the gamified and non-gamified conditions, these findings were based solely on the assessment data and did not utilize any of the data from the in-app measures that are recorded when using FH2T-E. To account for this, we ran an exploratory factor analysis (EFA) to identify the number and structure of the factors underlying the overall data variables that were recorded within FH2T-E as students solved problems. Two significant factors found in the analysis were associated with the students' levels of engagement as well as their progression in the application throughout the study. Using these factors, we found that students in the gamified condition averaged 6.58 points higher on the posttest than students in the non-gamified condition ($p < 0.05$) and students who progressed faster and completed more unique problems in the app may demonstrate higher posttest scores.



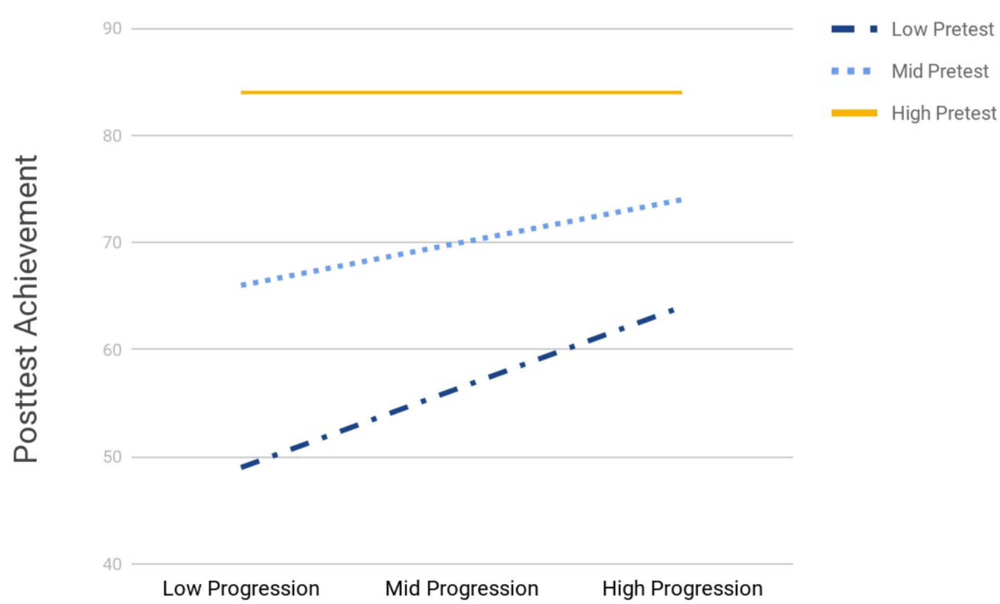


Figure 8.4. Differences in learning gains based on factor analysis of in-app measures of FH2T-E for both engagement (top) and progression (bottom)

8.4 Discussion

8.4.1. Overview of Findings

The overall goal of this effort was to explore the interaction between gamification and education in dynamic learning technologies. We achieved this by conducting a randomized controlled trial involving 185 elementary school students from 9 elementary schools in the northeast where students played with the elementary school version of FH2T (FH2T-E) for four (4) twenty-minute sessions across 3 weeks. Students were randomly assigned into one of three conditions: (1) a paper-based worksheet (control); (2) a non-gamified version of FH2T-Elementary where rewards, background images, and unlockables are removed; and (3) a gamified version of FH2T-Elementary. Our findings suggest that when comparing solely assessment learning gains, our intervention, FH2T-E, worked just as well as the control condition. Additionally, there were no differences in learning gains when comparing the gamified version of FH2T-E to the non-gamified version. However, when you incorporate the data from FH2T-E's in-app measures, an exploratory factor analysis found statistically significant factors that linked behaviors exerted by the gamified version of FH2T-E to significant learning gains. This suggests that gamified components such as rewards, unlockable levels, and worlds promote engagement and progression and may also lead to greater learning gains.

8.4.2. Conclusion

At the beginning of this effort, we set out to accomplish a series of tasks including testing the feasibility of our elementary version of From Here to There!, evaluating the learning gains associated with our intervention, and comparing the differences in behavior and learning gains between the gamified and non-gamified versions of our application. After conducting a randomized controlled trial, we found that our iterative design process proved to be feasible in deployment and accessibility for average classrooms. Additionally, our assessment data suggests that while significant learning gains were associated with the control condition and both versions of our intervention, there were no significant differences in learning gains between the three conditions. While this finding suggests that using our dynamic technology leads to learning gains, it also indicates that there are no benefits to the gamification aspects and components that have been featured in other chapters of this dissertation. However, when taking the in-app measures into account and performing an exploratory factor analysis, our findings suggest that the gamified components featured in the gamified condition were strongly linked to learning through the engagement and progression factors of our model. As participants became more engaged with the gamification components of the intervention and progressed further through each world, their posttest scores increased significantly. This finding is consistent with the original findings in Ottmar et al. (2015) that found that higher amounts of exposure to these dynamic learning technologies lead to increased posttest scores. We also believe that this result compliments the behaviors associated with gamification that we have highlighted in the previous chapters. Additionally, we believe that the playful nature of the gamified learning environment provides an opportunity to practice new strategies that further promote conceptual understanding as outlined in Bay-Williams (2001) and Carr et al. (2011). Furthermore, it also emphasizes the need for additional research in this intersection of gamification and learning. Our last chapter is highly influenced by this study as well as our prior work and applies it to a newly developed technology, Treasure Hunter. This final study explores mathematical concepts such as number decomposition and number reasoning in a gamified learning environment.

References

Bay-Williams, J. M. (2001). What is algebra in elementary school? *Teaching Children Mathematics, 8*(4), 196-200. http://www.jstor.org.ezproxy.wpi.edu/stable/41197754

Carr, M., Taasoobshirazi, G., Stroud, R., & Royer, J. M. (2011). Combined fluency and cognitive strategies instruction improves mathematics achievement in early elementary school. *Contemporary Educational Psychology, 36*(4), 323-333.

Carraher, D. W., Martinez, M. V., & Schliemann, A. D. (2008). Early algebra and mathematical generalization. *ZDM*, *40*(1), 3-22.

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*(2), 423-441.

Lins, R., & Kaput, J. (2004). The early development of algebraic reasoning: The current state of the field. In *The Future of the Teaching and Learning of Algebra The 12thICMI Study* (pp. 45-70). Springer, Dordrecht.

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*.

9 STUDY 4: GAMIFIED LEARNING ENVIRONMENTS FOR NUMBER DECOMPOSITION AND REASONING

The following chapter introduces an arithmetic-based video game, *Treasure Hunter*, as well as the discussion of cognitive load in early algebraic activities. A study was conducted which explored student behavior when solving the in-game arithmetic puzzles and decomposition tasks. In addition to an accuracy assessment featuring difficult concepts such as order of operations, participants were also scored on the complexity of their answers before and after interacting with the application. The study also featured three conditions that affect the presentation of number options that students are shown when formulating their solutions for these early algebraic tasks.

9.1 Introduction

9.1.1 Overview

Dynamic technology has the potential to record information on the student problem solving process and support teachers in identifying student understanding and misconceptions in innovative and efficient ways. In our previous work, we have applied this potential toward algebraic concepts to further understand the student perspective using projects such as *From Here to There!*. We are exploring arithmetic problems involving number decomposition and discussing how dynamic technology can further aid in number reasoning and order of operations. A crucial conceptual progression for students is to transition from the foundational arithmetic problems to a more abstract world of algebra problems with variables (Heffernan and Koedinger, 1998). The crucial grounding of arithmetic requires students to not just recall mathematical facts, but also understand the relationship between numbers. Prior research suggests that early childhood math proficiency is a good indicator for future academic achievement (VanDerHeyden & Burns, 2009). However, the deficit in mathematical understanding begins to occur in this transition due to a lack of understanding of number sense and flexibility of expressions through mathematical operations (Kalchman, 2011). In elementary school, students begin to learn math as numbers by creating concrete representations, and often do not recognize the flexible potential or function of the digit (Carr et al., 2011). In a child's first years of schooling, math instruction is focused predominantly on the recognition of patterns in numerical expressions, and their ability to apply these patterns to other math expressions (Lins & Kaput, 2004). While these patterns are taught early on in school, students are not readily introduced to the relation or flexibility within the patterns. By second grade, children gain the capacity to cognitively represent such numbers, and begin to cognitively represent numbers through abstract reasoning of numbers and their relations to one another (Carr et al., 2011). At this stage, children first acquire the ability to mentally represent numbers and operations. More broadly, when students are provided with ample opportunity to practice new strategies and understand the effectiveness of these new strategies then they are able to acquire and use these strategies independently (Bay-Williams, 2001; Carr et al., 2011). These cognitive representations of number patterns allow students to think flexibly about numbers. This decomposition then allows them to manipulate expressions into something that they can more easily reason through. However, the educational topics taught in schools are often introduced as standalone topics and do not facilitate much understanding on flexibility and reasoning. Thus, if decomposition of numbers was introduced at an earlier age, as a precursor to formal algebra learning, children could potentially increase their conceptual understanding of number relations and flourish in their flexibility and number reasoning. One of the more difficult conceptual tasks in early Algebra learning is order of operations. Order of operations has been noted as a major area of confusion for students learning algebra (Welder, 2012). To combat this, students are often taught clever mnemonics such as PEMDAS to help remember the order of operations. However, Glidden (2008) suggests that tactics such as this might lead students to interpret the order literally and disregard the underlying conceptual understanding. Students may incorrectly believe that addition has a priority over subtraction and disregard the position and ordering in the expression or equation (Glidden, 2008). Furthermore, research has shown that the equal sign is a common point of misconception even in mathematically advanced college students (Clement, Narode & Rosnick, 1981). Instead of identifying the equal sign as a symbol of equivalence, students often describe the equal sign as an operator with an operation on the left and a result on the right (Kieran, 1981). This distinction is crucial in algebraic reasoning as students must be able to correctly understand the equal sign and identify its relation and equivalence (Knuth et al., 2006; Welder, 2012). Cognitive tasks such as the card game *24*, where players must use various numbers to ultimately total up to a given number, may help improve mathematical reasoning tasks such as number decomposition (van der Maas & Nyamsuren, 2016) while also increasing cognitive functions and working memory. Mathematical reasoning and decomposing numbers are identified as critical learning sectors relative to Common Core Standards. By understanding equality in mathematical numbers and expressions, students can transition from concrete representations and progress to a more abstract way of thinking. Therefore, current research suggests that introducing number reasoning interventions may result in an increased understanding of decomposition (Aleven et al., 2010; Brendefur et al., 2013). The current work, *Treasure Hunter*, focuses on number decomposition and order of operations in a fun web-based video game. In the game, players are given a set of numbers and a limited set of operators to comprise a specific number. This activity introduces topics in both arithmetic and early algebra as players must both carry and calculate numbers while also strategically filling in each space (much like solving for a variable) to attain the specific number. Furthermore, this activity also includes procedures such as the order of operations in the final levels of the game where addition, subtraction, multiplication, and division are all included as potential operators.

9.1.2 Games in Education

The technical design of *Treasure Hunter* is the product of this prior research. Our earlier research in engagement through gamification and educational games has ultimately led us to the design choices featured in the current work. As we have discussed, video games have become a culturally relevant medium in the current generation (Seel, 2001; Aarsand, 2007; Gee, 2007). The immediate feedback and immersive, adaptive environments found in video games and other technologies are a crucial part of their culture. In Beck and Wade's book *The Kids Are Alright: How the Gamer Generation is Changing the Workplace* (2006), they argue that because the millennial generation has grown up with technology that it has become a central part of their lives (Beck & Wade, 2006). Educators took advantage of the fun and entertaining exterior of video games to further push pedagogical content in the early 1980's. This push gave birth to the edutainment (educational entertainment) genre of video games with popular titles such as *Math Blasters*, *Where in the World is Carmen Sandiego?* and *the Oregon Trail*. Research suggests that video games can motivate and interest learners (Dempsey et al., 1994; Jacobs & Dempsey, 1993; Lepper & Malone, 1987; Malone, 1981, 1983; Malone & Lepper, 1987; Malouf, 1988), increase retention rates (Dempsey et al., 1994; Jacobs & Dempsey, 1993; Pierfy, 1977), and improve reasoning skills and higher-order thinking (Mayland, 1990; Rieber, 1996; Wood & Stewart, 1987; Hogle, 1996). In view of these findings, we find that video games are a culturally relevant medium that can administer pedagogical content while still maintaining a fun "sugarcoated" experience, thus we feel that games are a suitable shell for our number decomposition tasks.

Figure 9.1. Screenshot of the Treasure Hunter welcome screen

9.1.3 Treasure Hunter

Treasure Hunter is a web-based application for the exploration of number decomposition, arithmetic, symbolic algebra, and logic. This learning technology allows students to dynamically interact with numbers and operators to solve number sense and decomposition problems.

Figure 9.2. Example of a successful problem completion

In Treasure Hunter, students solve number decomposition puzzles by using a given set of numbers and operators. Each problem or treasure chest has an “unlock code”. These “unlock codes” are the number that you must amount to using the given numbers and operators for each problem. When a player correctly solves a problem, the treasure chest opens and the player is awarded gold before being presented with their next problem. As players complete more and more problems, they will eventually earn enough gold to unlock the next world where they may earn even more gold. Starting with only addition, players work through 6 worlds of problems with each problem getting progressively harder and varying in operations such as addition, subtraction, multiplication, and division. For example, players are given a set of numbers such as 2, 4, 8, 11, and 16 and are told to make up the number 7 using only two of the numbers and the subtraction operator. Similar to other number logic games such as 24, Treasure Hunter targets number logic and reasoning in a gamified environment. However, a strong advantage of Treasure

Hunter over card-based games such as 24 is its foundation in technology which allows certain affordances such as data logging of attempted combinations as well as event-level timestamps when players are solving problems. Thus, teachers and researchers are able to see key information such as the amount of time taken when solving a problem, the number of attempts that were made or reset, and every iteration of the solution that a player may make when solving.

Figure 9.3. A screenshot of one of the levels in the world select menu

Some treasure chests require players to use only one operator while others require a combination of operators. Players may use an operator as many times as they want, but may only use a given number once in the sequence. Furthermore, the number of number spaces (the amount of numbers in the answer) required in the answer sequence varies from problem to problem. For example, players may be asked to amount to 8 using two number spaces, the addition operator, and a series of numbers, but are asked in the following problem to amount to the same number with the same operator and number series using three number spaces instead. By framing the problem as a fill in the blank-style algebraic equation, this activity can help teachers and researchers introduce topics such as algebraic thinking without having to involve variables explicitly.

We have designed this game with the intention of exploring number reasoning and decomposition in middle school students. While there are not many video games on topics such as number decomposition, we feel that we have stumbled upon an excellent cognitive task that involves number reasoning, working memory, and perception. Among other interesting areas to explore with this intervention, we are focusing on creativity and complexity in number reasoning, difficulties around order of operations and early algebra reasoning, and cognitive load based on perception. We feel that this is important as it not only affects the design choices that we wish to explore in games like Treasure Hunter, but also in understanding some of the cognitive struggles that occur when students are completing similar activities involving number reasoning and decomposition. The intention of this project is to promote flexible thinking when students are working with numbers to complete a task. The goal of each level is to use the given number patterns to find their relation to the specified number. These cognitive representations of number patterns allow students to think flexibly about numbers. Thus, if they understand the patterns between numbers, they should be able to further identify the patterns in the activity to attain their goal. The game's introduction tutorial informs the participant of the game's objectives as well as how to play the game. With each correctly solved puzzle, participants are awarded points that they use to unlock new worlds and mathematical operations. As the participant progresses through each world, the math puzzles become more difficult and introduce more operations starting with just addition being available and ending with addition, subtraction, multiplication, and division being available.

Figure 9.4. After earning enough points, players unlock new worlds featuring new operators and number combinations

In addition to being a fun and engaging technology where students can refine their number reasoning skills, Treasure Hunter can also be used as a tool for teachers and researchers to gain a better understanding of the student perspective. As students solve problems in Treasure Hunter, interaction data is being logged on the backend that allows researchers to recreate the problem-solving process for each student's attempt as well as identify strategies in how they play the game. The system records the start and end time of each attempt as well as the number of resets a player makes and a snapshot of the individual attempt itself. We can then use this in-app data to answer our research questions and further explore these concepts.

9.1.4 Research Goals

1) Do students' assessment scores increase more when interacting with Treasure Hunter than compared to a similar worksheet activity?

2) Do in-game behaviors of participants explain learning above and beyond prior knowledge and condition?

3) Do certain aspects of the problem structure such as the number of terms or operators impact student performance and behavior?

9.2 Methodology

9.2.1 Participants

154 students from 5 middle schools in the Northeastern United States participated in an hour long study consisting of a pretest and posttest assessment as well as a 30 minute intervention of the game Treasure Hunter. Upon the start of the study, students were randomly assigned conditions and given a 9-question assessment that explored prior knowledge of decomposition and number reasoning. This assessment was the same for all three conditions and was developed for this study and based on the number reasoning assessments of Blanton, et al. (2015) as well as Geary, et al. (2009). The assessment takes approximately 10 - 15 minutes to complete and is administered via an online platform built specifically for this study that allows students to enter only valid answers using an on-screen keyboard.

Once the assessment was completed, participants were automatically redirected to their version of the online intervention. Based on condition, participants were given one of three versions of Treasure Hunter – the normal gamified version of Treasure Hunter; a black and white, non-gamified version of Treasure Hunter with the points, graphics, and levels removed; or an online worksheet with the same problems as Treasure Hunter, but without any feedback or correctness displayed to the user. Participants were given 30 minutes to interact with their intervention before being notified that their 30 minutes was up and automatically redirected to the posttest assessment.

After the intervention, students were redirected to the posttest where they were given the same assessment as the pretest. Once they completed the posttest, they were thanked for their participation, automatically logged out of the study, and told that they may safely close their browsers.

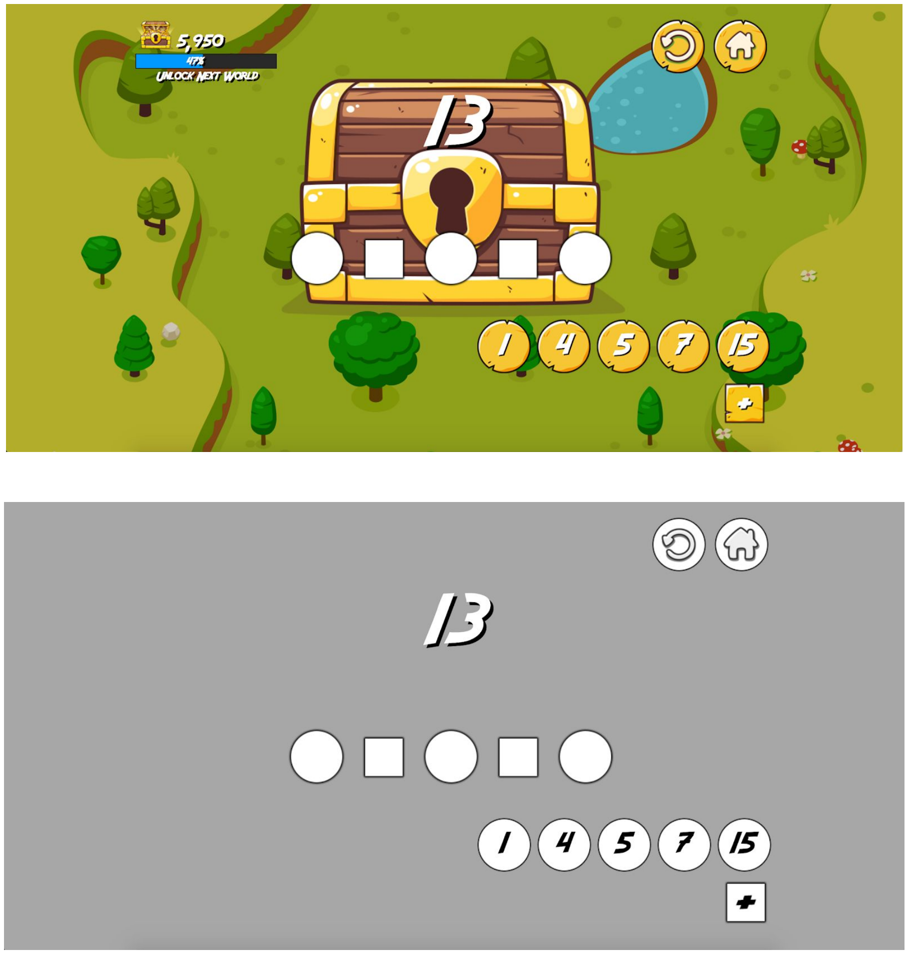


Figure 9.5. Screenshots of the gamified (top) and non-gamified (bottom) versions of Treasure Hunter and the online worksheet.

9.2.2 Measures

1) **Pretest/posttest Assessment.** Participants completed a 9-question pretest and posttest that helped to measure their existing knowledge of arithmetic as well as decomposition tasks. These tasks are broken into two main sections that examine correctness on skills such as order of operations. This assessment is our main focus when exploring our research questions and is administered to participants before and after they are given

Treasure Hunter.

2) **Resets per problem.** When trying to solve a problem, a reset occurs when a player clicks the restart button, thus clearing the slots and leading to a new attempt of the problem. A higher number of attempts can indicate a low understanding of number relation mastery as well as potential struggle due to low knowledge.

3) **Incorrect answers per problem.** When trying to solve a problem, an incorrect answer occurs when a player fills all of the number and operator slots, but does not unlock the treasure chest. The treasure chest then gets reset, thus clearing the slots and leading to a new attempt of the problem. A higher number of incorrect answers can indicate a low understanding of number relation mastery as well as potential gaming of the system due to disengagement and/or low knowledge.

4) **Time spent per problem.** We can measure the amount of time that a participant takes when solving each problem. The elapsed time spent in a problem is measured in seconds and indicates difficulty of a problem for a particular participant. Additionally, we can use the amount of time spent per problem as an indication of conceptual understanding. As more complex mathematical operators, numbers, and spaces are introduced in the game, the problems become more difficult to answer, thus resulting in a longer amount of time spent. We can also compare the time spent per problem across conditions and problem structures to identify any existing behavioral patterns.

9.2.3 Analysis

In Treasure Hunter, participants are asked to solve math puzzles by dragging and dropping numbers and operators into place to total up to a specific number, the unlock code. As the participant progresses through each level, the puzzles become more difficult and complex. These complexities are found in the list of given numbers, spaces, and operators available per problem. For example, players may be asked to amount to 8 using two number spaces, the addition operator, and a series of numbers, but are asked in the following problem to amount to the same number with the same operator and number series using three number spaces instead. Depending on the list of given numbers and operators, it may be more or less difficult to total up to 8 using three number spaces and two operators than it is to total up to 8 using two number spaces and one operator. As researchers, we are interested in seeing the different behaviors that can be observed based on the different types of problems in the application. We can measure this by the amount of time taken per problem as well as the number of resets or attempts that a participant uses when trying to solve the problem. We are able to look at this data using timestamps measured in seconds and compare it amongst participants to identify which types of problems required the most amount of time.

9.3 Results

RQ1, Do students' assessment scores increase more when interacting with Treasure Hunter than compared to a similar worksheet activity?

Of the 154 students that participated in the study, 53 participants did not complete the posttest and were dropped from the analysis regarding learning gains. Of the 101 students that completed both the pretest and posttest assessment, 35 participants were in the no feedback, worksheet activity; 35 participants were in the non-gamified version of Treasure Hunter; and 31 participants were in the gamified version of Treasure Hunter. Table 9.1 features the descriptives of both pretest and posttest by condition.

Table 9.1

Descriptives across all three conditions for Treasure Hunter assessments

Measure	Condition	*N*	*M*	*Std. Dev.*	*Std. Error*
Pretest	Worksheet	35	5.991	2.219	0.375
	Non-Gamified	35	5.273	2.191	0.370
	Gamified	31	5.814	2.392	0.430
	Total	101	5.688	2.263	0.225
Posttest	Worksheet	35	5.930	2.412	0.408
	Non-Gamified	35	5.187	2.209	0.373
	Gamified	31	5.670	2.677	0.481
	Total	101	5.593	2.426	0.241
Gains	Worksheet	35	-0.060	0.889	0.150
	Non-Gamified	35	-0.086	1.809	0.306
	Gamified	31	-0.143	0.968	0.174
	Total	101	-0.095	1.290	0.128
Sec. 1 Change	Worksheet	35	0.057	0.539	0.091
	Non-Gamified	35	0.029	0.891	0.151
	Gamified	31	-0.097	0.539	0.097
	Total	101	0.000	0.678	0.067
Sec. 2 Change	Worksheet	35	-0.032	0.113	0.019
	Non-Gamified	35	0.000	0.129	0.022
	Gamified	31	-0.014	0.098	0.018
	Total	101	-0.015	0.114	0.011

Sec. 3 Change	Worksheet	35	-0.086	0.612	0.103
	Non-Gamified	35	-0.114	1.078	0.182
	Gamified	31	-0.032	0.706	0.127
	Total	101	-0.079	0.821	0.082

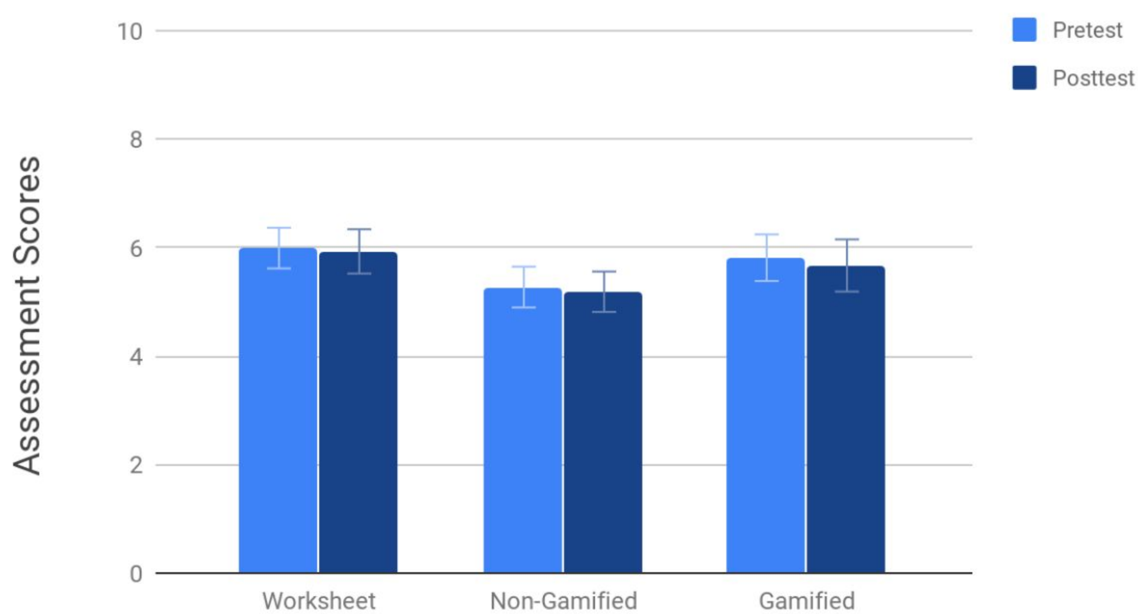


Figure 9.6. Pre-test and post-test score comparison by condition

All three conditions average around 5 out of 9 problems correct for the pretest and posttest. We ran an ANOVA to determine whether or not there was any statistical significance among the three conditions assessment performance. There were no significant differences in pretest to posttest gains for the control ($M = -0.060$, $SD = 0.889$), non-gamified ($M = -0.086$, $SD = 1.809$) and gamified ($M = -0.097$, $SD = 0.539$) conditions; $t(98) = 0.035$, $p = 0.966$. This indicates that exposure to the 30 minute intervention does not affect the posttest scores much at all.

Table 9.2

Analysis of variance across all three conditions for Treasure Hunter assessment

Measure	Group	*Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*
Pretest	Between Groups	9.720	2	4.860	0.948	0.391
	Within Groups	502.223	98	5.125		
	Total	511.942	100			
Posttest	Between Groups	9.924	2	4.962	0.840	0.435
	Within Groups	578.816	98	5.906		
	Total	588.740	100			
Gains	Between Groups	0.118	2	0.059	0.035	0.966
	Within Groups	166.306	98	1.697		
	Total	166.424	100			
Sect. 1 Change	Between Groups	0.433	2	0.217	0.466	0.629
	Within Groups	45.567	98	0.465		
	Total	46.000	100			
Sect. 2 Change	Between Groups	16.420	2	0.009	0.672	0.513
	Within Groups	1278.040	98	0.013		
	Total	1294.460	100			
Sect. 3 Change	Between Groups	16.420	2	0.056	0.082	0.921
	Within Groups	1278.040	98	0.676		
	Total	1294.460	100			

RQ2, Do in-game behaviors of participants explain learning above and beyond prior knowledge and condition?

Though there appears to be no significant learning gains in any of the three conditions, a median split was performed to bin all of the participants into either a high or low group based off of their pretest scores for the two Treasure Hunter conditions (gamified and non-gamified). The two groups were then compared using an independent samples t-test to look at the in-app measures and observe their performance in Treasure Hunter. This was done to observe the validity of the pre and post assessment for this study.

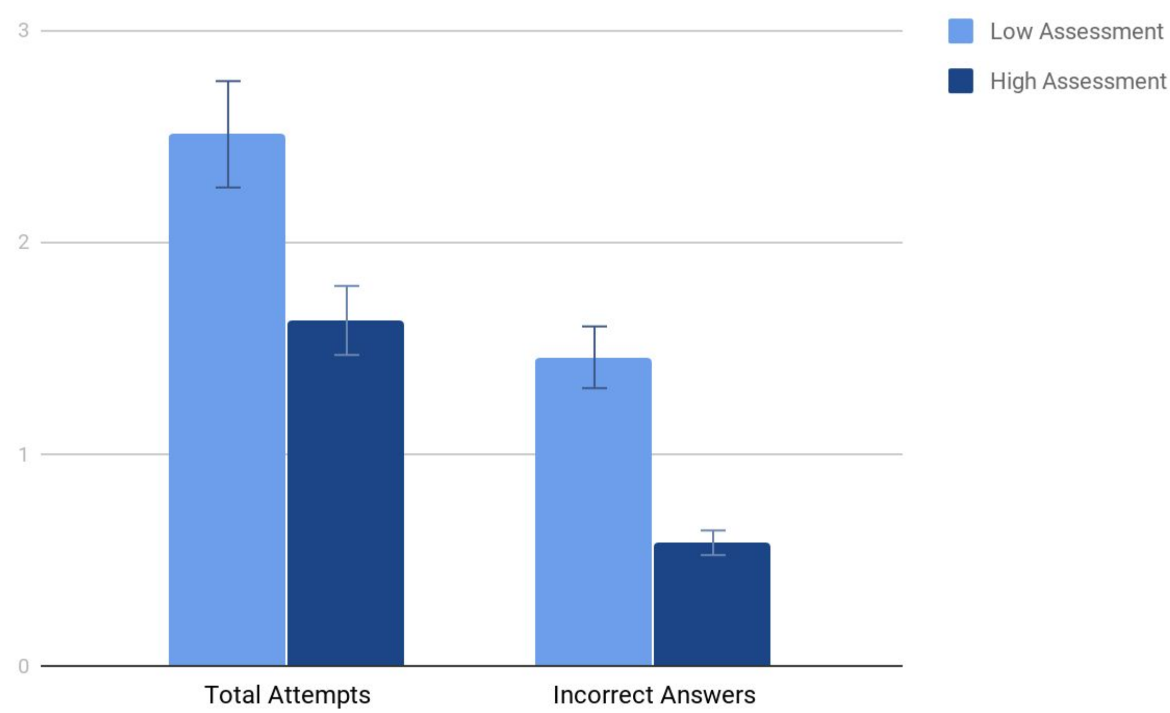


Figure 9.7. A median split of the pretest scores reflects data of in-app performance measures in Treasure Hunter

Table 9.3

Descriptives of high and low assessment performance groups

Measure	Group	*N*	*M*	*Std. Dev.*	*Std. Err.*
Total Attempts	Low	33	1.508	1.046	0.182
	High	33	0.630	0.490	0.085
Completed	Low	33	0.895	0.188	0.032
	High	33	0.906	0.237	0.041

Incorrect Answers	Low	33	1.456	1.035	0.180
	High	33	0.581	0.502	0.087
Resets	Low	33	0.055	0.128	0.022
	High	33	0.048	0.105	0.018
Returns	Low	33	1.785	1.102	0.191
	High	33	1.613	0.892	0.155
Steps	Low	33	13.117	5.505	0.958
	High	33	9.581	3.797	0.661

Table 9.4

Independent samples t-test of high and low assessment performance groups

Measure	Group	*N*	*M*	*SD*	*t*	*p*
Total Attempts	Low	33	1.508	1.046	4.367	0.000
	High	33	0.630	0.490		
Completed	Low	33	0.895	0.188	-0.203	0.839
	High	33	0.906	0.237		
Incorrect Answers	Low	33	1.456	1.035	4.366	0.000
	High	33	0.581	0.502		
Resets	Low	33	0.055	0.128	0.238	0.812
	High	33	0.048	0.105		
Returns	Low	33	1.785	1.102	0.694	0.490
	High	33	1.613	0.892		
Steps	Low	33	13.117	5.505	3.038	0.003
	High	33	9.581	3.797		

As the data suggests, the lower performing group ($M = 1.508$, $SD = 1.046$) took more steps and made more attempts due to incorrect answers and resets than the higher performing group ($M = 0.630$, $SD = 0.490$) when accounting for in-app measures ($t(31) = 4.367$, $p = 0.000$).

While the findings on learning gains were not significant, this data suggests at least some validity to the assessment as a proper measure for learning, if it would have been present. Nevertheless, in-game actions and behaviors are consistent with assessment performance. While this does not necessarily mean that playing Treasure Hunter leads to any increase in learning gains, it can be argued that Treasure Hunter's design can properly reflect learning knowledge based on in-app performance measures such as incorrect answers, step counts, and resets.

RQ3, Do certain aspects of the problem structure such as the number of terms or operators impact student performance and behavior?

In-app measures were then analyzed across the gamified and non-gamified versions of Treasure Hunter to observe behaviors on particular problem attempts. In addition to the data analysis on the pretest and posttest assessment gains, the study was structured to assess and analyze behaviors of performance against variations in problem structure. These problem structure variations include number magnitude, available operators on a given problem, and the number of spaces (terms) required when solving a problem. For this analysis, we used attempt level data from over 2300 problem attempts. Problem attempts occur when participants load any problem. In the event that a participant resets the problem or guesses an incorrect answer, the problem is reset on-screen, but the logged attempt record is still maintained until a participant either solves the problem correctly or exits the problem.

Our first behavioral analysis explores the effect of the number of spaces or terms required for a problem. As discussed earlier, Treasure Hunter supplies a list of given numbers, spaces, and operators available per problem. For example, players may be asked to amount to 8 using two number spaces, the addition operator, and a series of numbers, but are asked in the following problem to amount to the same number with the same operator and number series using three number spaces instead. In an effort to explore whether or not having additional spaces impacted student performance, we created a pair of problems in each world that asked for the same answer using the same number of operators and the same set of given numbers, but required participants to solve it using 2 terms for one problem and then solve it using 3 terms for another. We then ran an independent samples t-test for the 2 term problems against the 3 term problems and found that the 2 term problems (M = 32.988, SD = 26.131) required less time, attempts, and steps than the problems with 3 or more terms/spaces (M = 94.658, SD = 96.108) across the in-app measures($t(103) = -6.344$, $p = 0.000$).

Table 9.5

Descriptives of high and low space problems

Measure	Group	*N*	*M*	*Std. Dev.*	*Std. Err.*
Elapsed Time	Low Spaces	105	32.988	26.131	2.550
	High Spaces	105	94.648	96.108	9.379
Total Attempts	Low Spaces	105	0.565	0.899	0.088
	High Spaces	105	1.138	1.447	0.141
Completed	Low Spaces	105	0.972	0.085	0.008
	High Spaces	105	0.927	0.143	0.014
Incorrect Answers	Low Spaces	105	0.552	0.878	0.086
	High Spaces	105	1.053	1.416	0.138

Resets	Low Spaces	105	0.015	0.115	0.011
	High Spaces	105	0.084	0.273	0.027
Returns	Low Spaces	105	0.652	0.944	0.092
	High Spaces	105	3.309	2.693	0.263
Steps	Low Spaces	105	6.223	4.547	0.444
	High Spaces	105	17.389	11.186	1.092

Table 9.6

Independent samples t-test of high and low space problems

Measure	Group	*N*	*M*	*SD*	*t*	*p*
Elapsed Time	Low Spaces	105	32.988	26.131	-6.344	0.000
	High Spaces	105	94.648	96.108		
Total Attempts	Low Spaces	105	0.565	0.899	-3.446	0.001
	High Spaces	105	1.138	1.447		
Completed	Low Spaces	105	0.972	0.085	2.842	0.005
	High Spaces	105	0.927	0.143		
Incorrect Answers	Low Spaces	105	0.552	0.878	-3.087	0.002
	High Spaces	105	1.053	1.416		
Resets	Low Spaces	105	0.015	0.115	-2.417	0.017
	High Spaces	105	0.084	0.273		
Returns	Low Spaces	105	0.652	0.944	-9.539	0.000
	High Spaces	105	3.309	2.693		
Steps	Low Spaces	105	6.223	4.547	-9.476	0.000
	High Spaces	105	17.389	11.186		

When compared to a problem requiring 2 terms, this data suggests that requiring participants to solve the same problem using 3 terms, their performance decreases significantly in every measure. For these three term problems, participants show a significant increase in elapsed time, number of steps taken, the number of attempts due to both an incorrect answer and problem resets, slight decrease in completion rates, and an increase in the amount of numbers and operators returned to the bin.

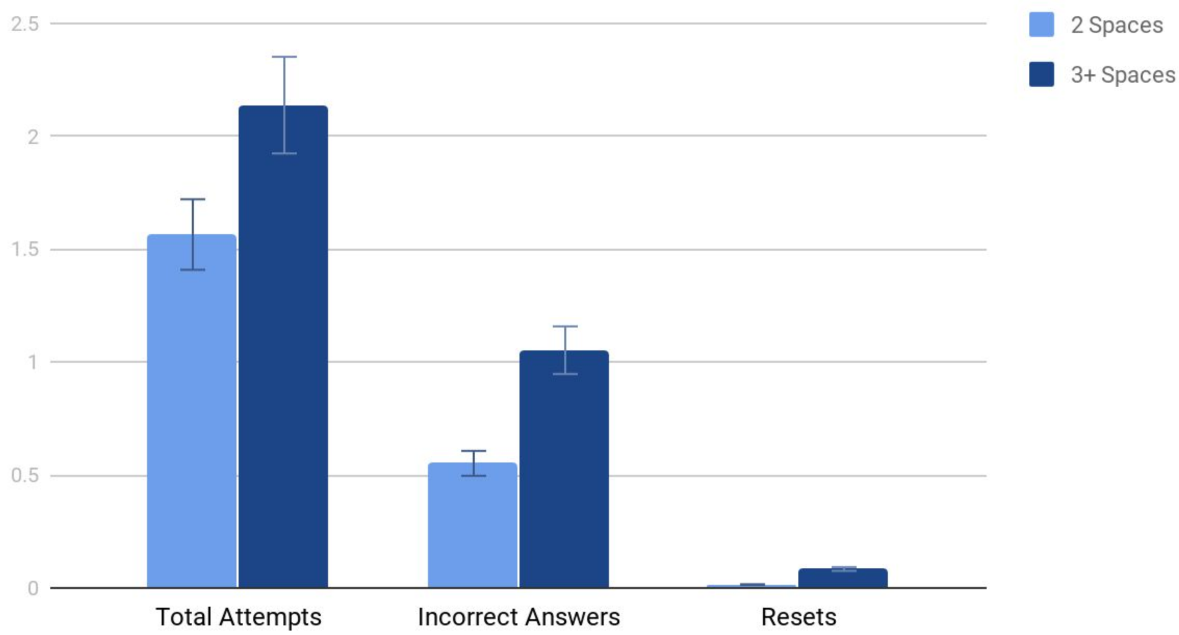


Figure 9.8. Comparison of in-app measures on study problems requiring answers with 2 terms against problems requiring answers with 3 or more terms

We also explored the behavioral effect of number magnitude when participants are solving Treasure Hunter problems. Similarly to how we observed the effects of the number of terms per problem, we created pairs of problems with the same unlock code, available operators, and number of terms required, but used small numbers for the first problem and then used larger numbers for the problem immediately following it. We followed this structure for every world with the exception of the solely addition and solely multiplication worlds due to not being able to achieve the same small-number answer without a subtraction or division operator. We then ran an independent samples t-test for the low magnitude problems against the high magnitude problems and found that the low magnitude problems (M = 57.968, SD = 67.391) required less time, attempts, and steps than the high magnitude problems (M = 94.218, SD = 104.318) across the in-app measures(t(100) = -2.952, p = 0.004).

Table 9.7

Descriptives of high and low magnitude problems

Measure	Group	*N*	*M*	*Std. Dev.*	*Std. Err.*
Elapsed Time	Low Magnitude	102	57.968	67.391	6.673
	High Magnitude	103	94.218	104.318	10.279
Total Attempts	Low Magnitude	102	0.752	1.455	0.144
	High Magnitude	103	2.905	3.041	0.299
Completed	Low Magnitude	102	0.935	0.121	0.012
	High Magnitude	103	0.890	0.231	0.023
Incorrect Answers	Low Magnitude	102	0.730	1.444	0.143
	High Magnitude	103	2.832	3.043	0.300
Resets	Low Magnitude	102	0.022	0.129	0.013
	High Magnitude	103	0.077	0.424	0.042
Returns	Low Magnitude	102	2.024	4.224	0.418
	High Magnitude	103	1.362	1.524	0.150
Steps	Low Magnitude	102	10.546	12.850	1.272
	High Magnitude	103	14.706	10.272	1.012

Table 9.8

Independent samples t-test of high and low magnitude problems

Measure	Group	*N*	*M*	*SD*	*t*	*p*

Elapsed Time	Low Magnitude	102	57.968	67.391	-2.952	0.004
	High Magnitude	103	94.218	104.318		
Total Attempts	Low Magnitude	102	0.752	1.455	-6.454	0.000
	High Magnitude	103	2.905	3.041		
Completed	Low Magnitude	102	0.935	0.121	1.736	0.084
	High Magnitude	103	0.890	0.231		
Incorrect Answers	Low Magnitude	102	0.730	1.444	-6.308	0.000
	High Magnitude	103	2.832	3.043		
Resets	Low Magnitude	102	0.022	0.129	-1.264	0.208
	High Magnitude	103	0.077	0.424		
Returns	Low Magnitude	102	2.024	4.224	1.495	0.136
	High Magnitude	103	1.362	1.524		
Steps	Low Magnitude	102	10.546	12.850	-2.561	0.011
	High Magnitude	103	14.706	10.272		

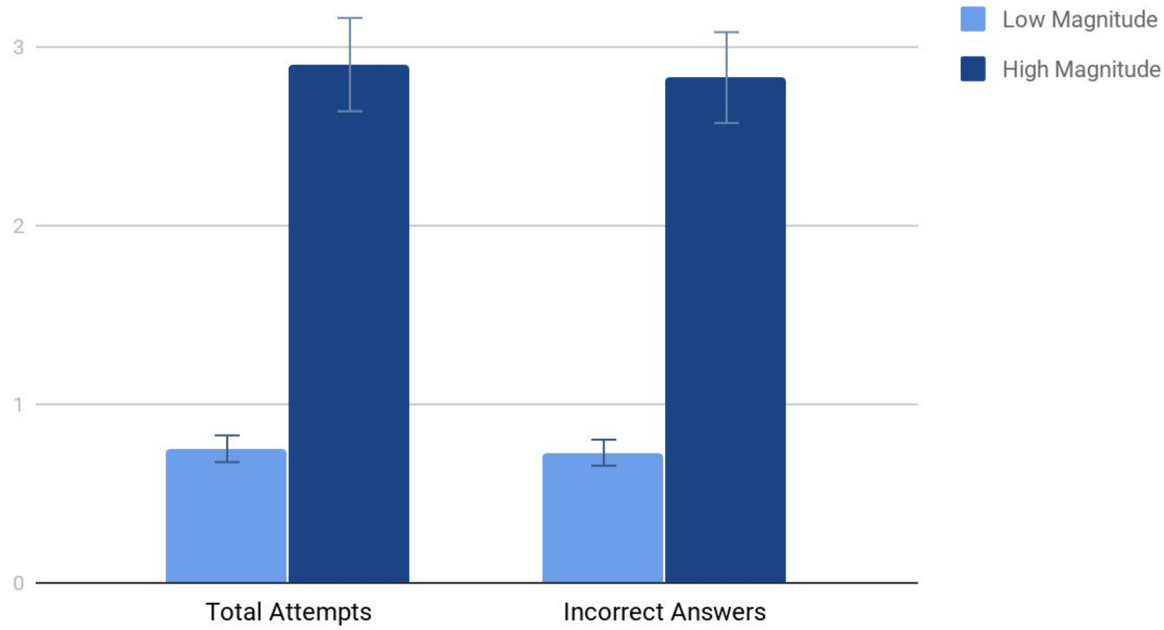


Figure 9.9. Comparison of in-app measures on study problems with high magnitude numbers against problems with low magnitude numbers

Lastly, we wanted to explore the effects of available operators while progressing through Treasure Hunter. As discussed earlier, Treasure Hunter begins with solely addition operators and gradually introduces worlds where there are solely subtraction operators, a combination of both addition and subtraction operators, solely multiplication and solely division operators, and combination of multiplication and division operators, and ends with a series of problems allowing all operators. We ran an analysis across our in-app measures to explore the effects that the availability of the operators themselves may have on student behavior and performance and found that performance measured by our in-app measures decreases significantly in any of the worlds where there are more than one operator to choose from ($t(554) = 43.208$, $p = 0.000$). Even in the early worlds, participants actively perform worse in the addition and subtraction world (M

= 92.466, SD = 98.670) where they are given the choice of two operators than they do in just addition (M = 39.235, SD = 30.986) and just subtraction (M = 50.495, SD = 34.804) where they are only given one. Not surprisingly, participants' performance drops the most in the final world where they are given all four operators to choose from (M = 238.267, SD = 199.279).

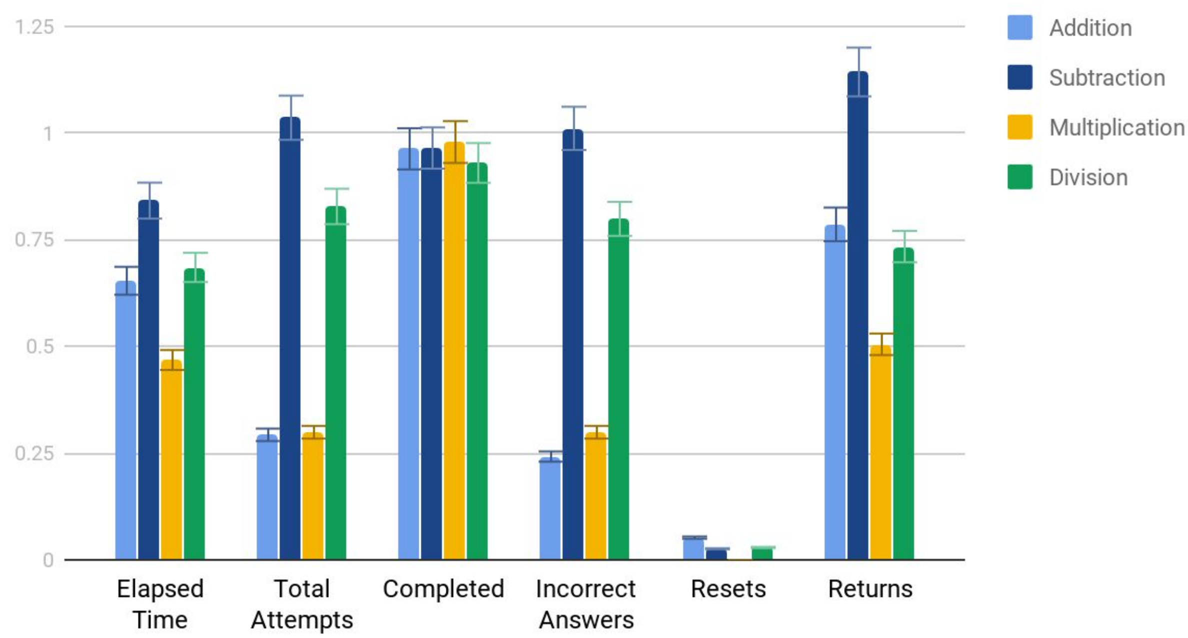


Figure 9.10. Comparison of in-app performance measures related to the operators that are available in problems

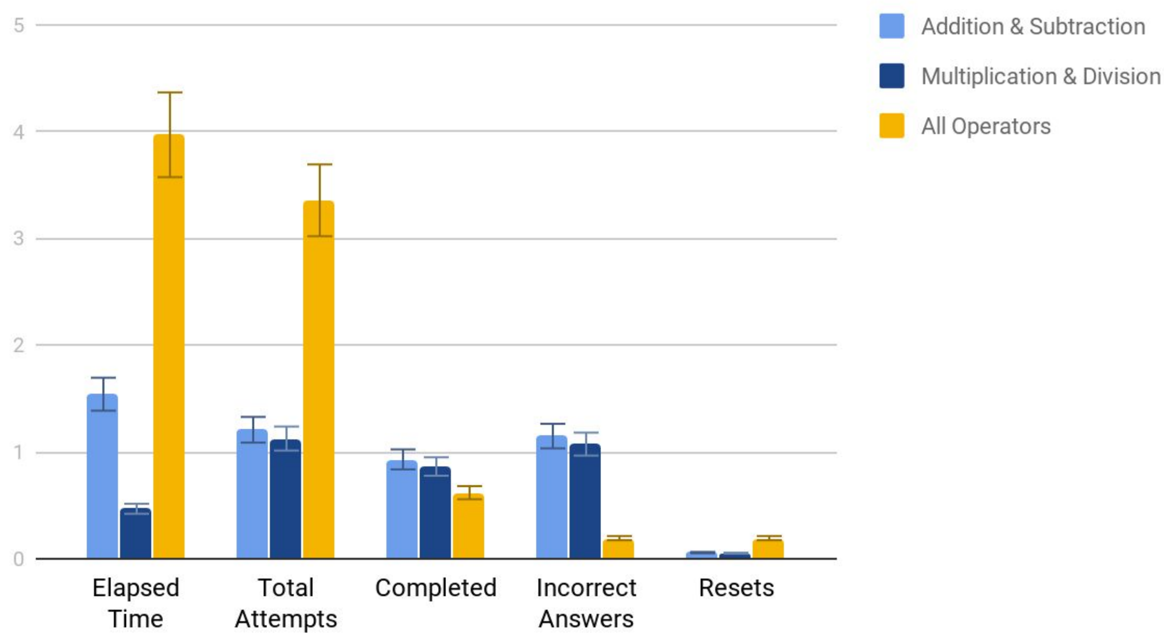


Figure 9.11. Comparison of in-app performance measures related to the number of operators that are available in problems

Table 9.9

Descriptives of measures across each world in Treasure Hunter

Measure	Group	*N*	*M*	*Std. Dev.*	*Std. Err.*
Elapsed Time	Addition	106	39.235	30.986	3.009
	Subtraction	105	50.495	34.804	3.396
	Add. & Sub.	97	92.466	98.670	10.018
	Multiplication	80	28.105	29.992	3.353
	Division	72	41.109	43.835	5.166
	Mult. & Div.	62	74.859	46.930	5.960
	All Operators	33	238.267	199.279	34.690
	Total	555	65.122	85.884	3.646
Total Attempts	Addition	106	0.293	0.512	0.049
	Subtraction	105	1.036	1.199	0.117
	Add. & Sub.	97	1.208	1.499	0.152
	Multiplication	80	0.299	1.126	0.126
	Division	72	0.828	1.249	0.147
	Mult. & Div.	62	1.126	1.791	0.227

	All Operators	33	3.357	3.284	0.571
	Total	555	0.939	1.604	0.068
Completed	Addition	106	0.963	0.116	0.011
	Subtraction	105	0.965	0.125	0.012
	Add. & Sub.	97	0.931	0.142	0.014
	Multiplication	80	0.979	0.072	0.008
	Division	72	0.930	0.193	0.023
	Mult. & Div.	62	0.864	0.239	0.030
	All Operators	33	0.619	0.383	0.067
	Total	555	0.924	0.190	0.008
Incorrect Answers	Addition	106	0.242	0.429	0.042
	Subtraction	105	1.011	1.189	0.116
	Add. & Sub.	97	1.149	1.435	0.146
	Multiplication	80	0.299	1.126	0.126
	Division	72	0.799	1.217	0.143
	Mult. & Div.	62	1.074	1.777	0.226
	All Operators	33	3.194	3.228	0.562
	Total	555	0.895	1.563	0.066
Resets	Addition	106	0.052	0.220	0.021
	Subtraction	105	0.026	0.114	0.011
	Add. & Sub.	97	0.058	0.242	0.025
	Multiplication	80	0.000	0.000	0.000
	Division	72	0.029	0.161	0.019
	Mult. & Div.	62	0.052	0.340	0.043
	All Operators	33	0.193	0.607	0.106
	Total	555	0.046	0.247	0.010
Returns	Addition	106	0.786	0.948	0.092
	Subtraction	105	1.143	1.020	0.099
	Add. & Sub.	97	2.669	3.369	0.342
	Multiplication	80	0.505	1.016	0.114
	Division	72	0.734	0.711	0.084
	Mult. & Div.	62	2.715	2.328	0.296
	All Operators	33	15.279	15.874	2.763
	Total	555	2.212	5.407	0.230
Steps	Addition	106	6.876	3.434	0.334
	Subtraction	105	9.434	5.289	0.516
	Add. & Sub.	97	13.990	11.042	1.121
	Multiplication	80	5.828	5.269	0.589
	Division	72	7.759	4.568	0.538
	Mult. & Div.	62	14.168	8.962	1.138

All Operators	33	53.368	49.257	8.575
Total	555	12.146	17.363	0.737

Table 9.10

Analysis of variance across all seven worlds in Treasure Hunter

Measure	Group	*Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*
Elapsed Time	Between Groups	1312355.12	6	218725.853	43.208	0.000
	Within Groups	2774042.87	548	5062.122		
	Total	4086397.99	554			
Total Attempts	Between Groups	281.068	6	46.845	22.420	0.000
	Within Groups	1145.019	548	2.089		
	Total	1426.087	554			
Completed	Between Groups	3.878	6	0.646	21.886	0.000
	Within Groups	16.185	548	0.030		
	Total	20.063	554			
Incorrect Answers	Between Groups	258.503	6	43.084	21.554	0.000
	Within Groups	1095.366	548	1.999		
	Total	1353.869	554			
Resets	Between Groups	0.963	6	0.160	2.684	0.014
	Within Groups	32.765	548	0.060		
	Total	33.728	554			
Elapsed Time	Between Groups	6396.191	6	1066.032	59.591	0.000
	Within Groups	9803.237	548	17.889		
	Total	16199.428	554			
Total Attempts	Between Groups	64953.484	6	10825.581	58.123	0.000
	Within Groups	102066.460	548	186.253		
	Total	167019.945	554			

9.4 Discussion

9.4.1 Outcome

Research Question 1 (RQ1), “Do students' assessment scores increase more when interacting with Treasure Hunter than compared to a similar worksheet activity?”, we used our pre and posttest assessment data to answer this question. As mentioned in the results section, this assessment was created for this study and was based off of similar assessments found in Blanton, et al. (2015) and Geary, et al. (2009). The assessment took approximately 10 - 15 minutes to complete and was administered via an online platform built specifically for this study that allows students to enter only valid answers using an on-screen keyboard. As discussed in the previous section, all three conditions averaged around 5 out of 9 problems correct for both the pretest and posttest. While there was virtually no significant change in performance from pretest to posttest for any of the three conditions, however, the descriptives and t-test performed during this analysis indicates that the groups were balanced in prior knowledge. In regard to RQ1, however, the data suggests that exposure to the 30 minute intervention does not affect posttest scores for any of the three conditions. While this suggests that students gained no knowledge after interacting with our technology, it is worth noting that Treasure Hunter was not designed as a teaching tool, but rather a means of skill assessment. While there are aspects of its design that might give students practice when solving math problems, there is no point during the intervention that we are actually teaching or introducing new mathematical topics or content other than how to perform actions and play Treasure Hunter. Thus, we do not find it to be surprising or discouraging that no significant learning gains were detected during the 30 minute intervention.

Though there appears to be no significant learning gains in any of the three conditions, we continued our analysis of the assessment to identify whether or not the assessment itself could have been a factor for no significant changes in learning gains. In an effort to gauge the validity of the assessment, a median split was performed to bin all of the participants into either a high or low group based off of their pretest scores for the two Treasure Hunter conditions (gamified and non-gamified). The two groups were then compared using an independent samples t-test to look at the in-app measures and observe their performance in Treasure Hunter. As the data suggests, the lower performing group took more steps, returned more tiles, and made more attempts due to incorrect answers and resets than the higher performing group. While the findings on learning gains were not significant, this data suggests at least some validity to the assessment as a proper measure for learning for the skills associated with this intervention.

In addition to the assessment analysis, we ran several analyses on the in-app measures of Treasure Hunter to determine student behavior when given various problem structures. This exploration was structured to assess and analyze behaviors of performance against variations in problem structure. These problem structure variations include number magnitude, available operators on a given problem, and the number of spaces (terms) required when solving a problem. We used attempt level data from over 2300 problem attempts from both the gamified and non-gamified versions of Treasure Hunter. Problem attempts occur when participants load any problem. In the event that a participant resets the problem or guesses an incorrect answer, the problem is reset on-screen, but the logged attempt record is still maintained until a participant either solves the problem correctly or exits the problem.

The first behavioral analysis explored the effect of the number of spaces or terms required for a problem. For example, players may be asked to amount to 8 using two number spaces, the addition operator, and a series of numbers, but are asked in the following problem to amount to the same number with the same operator and number series using three number spaces instead. In an effort to explore whether or not having additional spaces impacted student performance, we created a pair of problems in each world that asked for the same answer using the same number of operators and the same set of given numbers, but required participants to solve it using 2 terms for one problem and then solve it using 3 terms for another. The data in Table 9.4 suggests that increasing the number of required terms from 2 to 3 while keeping every other component the same shows a significant increase in elapsed time, number of steps required, the number of attempts due to both incorrect answers and manual resets, and the number of tiles returned. Though it may seem obvious or trivial that requiring an extra term would decrease student performance for these tasks, consider that the task itself does not necessarily increase in complexity just because it requires more terms. The complexity of a Treasure Hunter problem lies in the student’s perception and ultimately their decomposition and number reasoning ability in conjunction with the given parameters of the problem structure. Additionally, consider that they were already exposed to the same answer, available number options, and given operators in the problem immediately preceding each of these problems. It is surprising that even despite being familiar with the operators and number relations from the previous problem, participants required approximately double the amount of time to complete the second problem of the pair. It is possible that a type of functional fixedness would be a factor where students are trying to force the solution of the two term problem into the same solution for the three term problem and cannot see any possible alternatives than what they have previously used. Unfortunately, this study only featured pairs of problems that started with two terms and ended with three terms. If a follow-up study were conducted that would alternate between the two-term/three-term format and the three-term/two-term format, it may find that there is a significant difference when increasing the number of terms that a problem has or perhaps that it depends on which format is introduced first. Testing specifics such as this would be an excellent follow-up to the current work.

Similarly to the term pairs, the data suggests that when we introduce a pair of problems that differ in number magnitude we get a similar effect. Like we did with the number of terms, we introduced pairs of problems in specific worlds where the first problem had relatively small number options while the second problem had much larger numbers to see how the magnitude of the numbers affected the in-app measures of Treasure Hunter. The paired problems were only administered in worlds that included a subtraction or division operator because students needed to be able to work down to the same answer in the high magnitude problem as the low magnitude problem. Aside from the magnitude difference in the number options, the answer, available operators, and required number of terms/spaces were consistent throughout both problems. On average, participant behavior increased significantly in the elapsed time, number of steps, and the number of attempts due to incorrect answers. This finding is not surprising as we expected larger numbers would be less familiar when decomposing or breaking down into smaller numbers. This was consistent across all of the worlds where the magnitude pairs were introduced. However, it is not clear as to whether or not there would be any difference in reversing the order of the pairs so the high magnitude problems were introduced first. Though the difficulty of high magnitude numbers would not change much, it is unsure whether or not that would improve the efficiency of the lower magnitude problems.

Finally, participant behavior was explored around the number of available operators in a problem as well as the behaviors exerted across all 7 worlds. As discussed earlier, Treasure Hunter progresses through 7 worlds each varying in the types of operators available. The worlds progress in the following order: solely addition, solely subtraction, addition and subtraction, solely multiplication, solely division, multiplication and division, and finally all four operators. Because of this structure, we can clearly compare how the number of operators available and the operators themselves affect student behavior and performance using the in-app measures. Our data suggests that performance measured by our in-app measures decreases significantly in any of the worlds where there are more than one operator to choose from. Even in the early worlds, participants actively perform worse in the addition and subtraction world where they are given the choice of two operators than they do in just addition and just subtraction where they are only given one. Not surprisingly, participants' performance drops the most in the final world where they are given all four operators to choose from. We believe this has to do with cognitive overload of the task at hand. As students are trying to actively calculate number options when they solve a problem, the possible combinations of numbers and operators exponentially grows with each additional operator. We believe that these potential combinations are thus approximated and evaluated by the students when solving a problem. This could explain why even though a problem does not necessarily require all of the numbers and operators to solve it, simply presenting them with an increased number of options may further delay their response time. Diving in further, we also examined student performance across all seven worlds to compare the effects of the operators themselves against the in-app measures of Treasure Hunter. As mentioned earlier, the progression of the worlds is as follows: solely addition, solely subtraction, addition and subtraction, solely multiplication, solely division, multiplication and division, and finally all operators. In addition to affirming our previous finding on multiple operators versus single operators, the data suggests that student performance is lower in the two worlds with solely addition and solely subtraction compared to solely multiplication or solely division. Our initial hypothesis was that these later worlds would prove to be more difficult as students will generally practice addition and subtraction more in school and struggle more with multiplication and division overall. However, we believe these findings are not focusing on basic fact retrieval, but instead are highlighting a unique aspect of the tasks at hand in Treasure Hunter. We believe that as students play Treasure Hunter there is a level of approximation that is taking place when students are trying to solve problems. This “ballpark” calculation occurs as students evaluate the different number options and operators that are available and compare it to the desired number goal. When solving a problem that involves addition or subtraction, the range of approximation is much broader. For example, when solely adding numbers such as 1, 3, 5, 7, and

12 to reach the answer 15 using 3 terms, the combination of numbers and operators broadly covers most numbers between 9 through 24. This obscures the correct combination further because the ballpark estimate of any number option is very close in proximity to the rest of the available numbers. Meanwhile, if the operator for the same problem were solely multiplication, ballpark estimates for the available number combinations of these problems are much farther apart in proximity. Students can quickly evaluate and remove 7 and 12 as number options knowing that involving these numbers at all will greatly exceed the number goal of 15. This ability to reduce the number of options and find an optimal path forward is an excellent example of the type of number sense and reasoning that we hoped to see when designing Treasure Hunter. Additionally, the cognitive functioning associated with Treasure Hunter is conceptually aligned with the mathematical reasoning tasks featured in van der Maas & Nyamsuren (2016) and provides a beneficial environment for number exploration that Kalchman (2011) and Carr et al. (2011) suggests could provide meaningful cognitive understanding. Furthermore, we believe this finding to be a good example of our technology's ability to reflect a student's number sense and number reasoning through our collected data.

9.4.2 Limitations and Future Work

After running an hour long study exploring the new technology, Treasure Hunter, our findings suggest that there were no significant differences in learning gains between the control, non-gamified, and gamified versions of the intervention. Furthermore, there were no significant differences in pretest and posttest assessments for any of the conditions. While there may have been some small scale learning gains in number sense or decomposition skills, the technology itself was never designed to be a learning intervention, but rather an assessment tool for number decomposition and reasoning. An analysis of the in-app data of Treasure Hunter suggests that student performance on the assessment was significantly linked to the in-app measures of Treasure Hunter, thus the pretest/posttest is likely a valid assessment despite there being no significant gains. Additionally, in-app measures were examined to highlight student behaviors and performance when faced with various problem structures in Treasure Hunter. Some of the more notable findings suggest that student performance decreases significantly as the number of terms required increases, though this should be re-examined with a follow-up study that alternates the ordering of how problem structures were introduced. The magnitude of the number options provided are also indicative of student performance. When students are given larger numbers, they tend to perform worse than when given smaller numbers. One of the more significant findings of the study is associated with student behavior when given certain operators types. Student performance significantly decreases when a problem allows for more than one operator due to cognitive overload. Additionally, student performance is highest on solely multiplication and solely division problems. We suspect that this is due to the approximation that is taking place when students are trying to solve problems. As students begin to evaluate numbers, operators such as addition and subtraction will often lead students to potential number combinations that are in close proximity to other potential combinations. However, using operators such as multiplication and division when evaluating number combinations may lead to potential combinations that result in answers that are farther apart in proximity to other potential number combinations. Thus, students are able to evaluate and eliminate number options more efficiently when using solely multiplication or division. Lastly, though the study proved to be successful in its implementation, data retrieval, and overall enjoyment by the participants, Treasure Hunter is still a work in progress that will continue to be improved, updated, and expanded for future use in research.

References

Aarsand, P. A. (2007). Computer and video games in family life: The digital divide as a resource in intergenerational interactions. *Childhood*, *14*(2), 235-256.

Aleven, V., Myers, E., Easterday, M., & Ogan, A. (2010, April). Toward a framework for the analysis and design of educational games. In *Digital Game and Intelligent Toy Enhanced Learning (DIGITEL), 2010 Third IEEE International Conference on* (pp. 69-76). IEEE.

Bay-Williams, J. M. (2001). What is algebra in elementary school? *Teaching Children Mathematics, 8*(4), 196-200. http://www.jstor.org.ezproxy.wpi.edu/stable/41197754

Beck, J. C., & Wade, M. (2006). *The kids are alright: How the gamer generation is changing the workplace*. Harvard Business Press.

Brendefur, J., Strother, S., Thiede, K., Lane, C., & Surges-Prokop, M. J. (2013). A professional development program to improve math skills among preschool children in head start. *Early Childhood Education Journal*, *41*(3), 187-195.

Carr, M., Taasoobshirazi, G., Stroud, R., & Royer, J. M. (2011). Combined fluency and cognitive strategies instruction improves mathematics achievement in early elementary school. *Contemporary Educational Psychology, 36*(4), 323-333.

Clement, J., Narode, R., & Rosnick, P. (1981). Intuitive misconceptions in algebra as a source of math anxiety. *Focus on Learning Problems in Mathematics*, *3*(4), 36-45.

De Rammelaere, S., Stuyven, E., & Vandierendonck, A. (2001). Verifying simple arithmetic sums and products: Are the phonological loop and the central executive involved?. *Memory & Cognition*, *29*(2), 267-273.

Dempsey, J., Lucassen, B., Gilley, W., & Rasmussen, K. (1993). Since Malone's theory of intrinsically motivating instruction: What's the score in the gaming literature? Journal of Educational Technology Systems, 22(2), 173-183.

Dempsey, J. V., Rasmussen, K., & Lucassen, B. (1994, February). Instructional gaming: Implications for instructional technology. Paper presented at the Annual Meeting of Association for Educational Communications and Technology, Nashville.

Gee, J. P. (2007). *Good video games+ good learning: Collected essays on video games, learning, and literacy* (Vol. 27). Peter Lang.

Glidden, P. L. (2008). Prospective elementary teachers' understanding of order of operations. *School Science and Mathematics*, *108*(4), 130-136.

Heffernan, N. T., & Koedinger, K. R. (1998). A developmental model for algebra symbolization: The results of a difficulty factors assessment. In *Proceedings of the twentieth annual conference of the cognitive science society* (pp. 484-489).

Hogle, Jan G. "Considering Games as Cognitive Tools: In Search of Effective" Edutainment."." (1996).

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*(2), 423-441.

Jacobs, J. W., & Dempsey, J. V. (1993). Simulation and gaming: Fidelity, feedback, and motivation. In J. V. Dempsey & G. C. Sales (Eds.), Interactive instruction and feedback (pp. 197-227). Englewood Hills, NJ: Educational Technology Publications.

Kalchman, M. (2011). Using the math in everyday life to improve student learning: The Math in Everyday Life homework assignment builds student confidence and competence in mathematics. *Middle School Journal*, *43*(1), 24-31.

Kieran, C. (1981). Concepts associated with the equality symbol. *Educational studies in Mathematics*, *12*(3), 317-326.

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*, 297-312.

Lins, R., & Kaput, J. (2004). The early development of algebraic reasoning: The current state of the field. In *The Future of the Teaching and Learning of Algebra The 12thICMI Study* (pp. 45-70). Springer, Dordrecht.

Logie, R. H., Gilhooly, K. J., & Wynn, V. (1994). Counting on working memory in arithmetic problem solving. *Memory & cognition*, *22*(4), 395-410.

Malone, T. (1980) . What makes things fun to learn? A study of intrinsically motivating computer games. Unpublished dissertation, Stanford University.

Malone, T. (1983). Guidelines for designing educational computer programs. Childhood Education, 59(4), 241-47.

Malone, T., & Lepper, M. (1987) . Making learning fun: A taxonomy of intrinsic motivations for learning. In R. Snow & M. Farr (Eds.), Aptitude, learning, and instruction (Vol. 3, pp. 223-253) . Hillsdale, NJ: Earlbaum.

Malouf, D. B. (1988) . The effect of instructional computer games on continuing student motivation. Journal of Special Education, 21(4), 27- 38.

Mayland, V. (1990) . Carmen Sandiego is in your classroom. Miami, Florida: Dade Public Education Fund.

Pierfy, D. A. (1977) . Comparative simulation game research: Stumbling blocks and stepping stones. Simulation and games, 8(2), 255-268.

Rieber, L. P. (1996). Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games. *Educational technology research and development*, *44*(2), 43-58.

Seel, J. (2001) Plugged in, spaced out, and turned on: Electronic entertainment and moral mindfields. Journal of Education, 179 (3), 17–32.

Van Der Maas, H. L., & Nyamsuren, E. (2017). Cognitive analysis of educational games: The number game. Topics in cognitive science, 9(2), 395-412.

VanDerHeyden, A. M., & Burns, M. K. (2009). Performance indicators in math: Implications for brief experimental analysis of academic performance. *Journal of Behavioral Education*, *18*(1), 71-91.

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

Wood, L. E., & Stewart, P. W. (1987). Improvement of practical reasoning skills with a computer game. Journal of Computer-Based Instruction, 14(2), 49-53.

10 SYNTHESIS

This synthesis is intended to summarize the previous chapters and to reiterate the impact of this work to the educational psychology and mathematics research lab, the Graspable Math research team, and the underlying scientific research community. Chapters are broadly summarized with the main takeaways, findings, and contributions. The final sections introduce the future of my research projects and the online research platform that has stemmed from my efforts.

10.1 Chapter Summaries

Starting with Chapter 2, we explored the benefits and drawbacks of using dynamic technology in the classroom as an instructional system, support structure, and assessment tool. Graspable Math is introduced as an assessment tool as it requires students to externalize their problem solving strategies and procedures on screen that they would normally internalize or "do in their head" with write-in or type-in your answer problems. A small observational study was conducted to explore the pros and cons of using static text compared to dynamic technology. The summarizing data of the dynamic technology provided more perspective in determining when students are struggling and may allow teachers to plan an intervention earlier than if they only saw that the student got the problem right or wrong. In Chapter 3, a study was conducted that demonstrates the feasibility of using dynamic technology and highlighted the variation about student problem solving strategies and mathematical flexibility. The variability of a single problem was explored to further demonstrate the potential for informing teacher instruction, identifying gaps, and understanding the student perspective through in-app measures instead of more traditional methods. In Chapter 4, From Here to There! was introduced. A review of the prior work and research surrounding the iPad app was discussed. Inaccessibility of the iPad app was determined to be a major obstruction for adoption in schools, thus a web version of the app was developed. The new web version was evaluated in a randomized controlled pilot study and found to be as effective as the iPad version. More extensive technical developments were added to support most modern browsers, low-speed Internet connections, and dynamically loaded scripts and style sheets. Chapter 5 gives a general summary of how videogames became intertwined with education and paved the road for edutainment and games for learning. In Chapter 6, a study was conducted that explored the concept of gamification in educational environments. The findings suggested that there is a significant correlation to higher engagement in the gamification group than that of the non-gamified group. The in-app data of this modified FH2T! web application allowed us to gain insight into what differed between the two groups in terms of user actions and other useful information such as the amount of time spent on each problem and attempt. This result was also consistent across both low and high interest participants. Additionally, while the rate of time per attempt was consistent across the high interest plain and high interest gamified groups, the high interest plain spent the least amount of time of all four groups while the high interest gamified group spent the most amount of time despite having completed the same number of distinct problems. Chapter 7 continued with gamified elements in learning environments and explored the effects of subgoal states and leaderboard components in From Here to There!. While there were no statistically significant changes in engagement from simply adding subgoals, it appeared that in-game progress slows down slightly when introducing subgoals. There were also no statistically significant changes simply by adding a leaderboard to the application, however, the anecdotal analysis of the data suggests that this also slows participant progression through the app. Furthermore, when exploring the in-app behaviors of the leaderboard component itself, the fluctuating condition outperformed the other three conditions in the number of subgoals attempted and points earned. Overall, the subgoal component does seem to be an easy extension of the application that can further promote number exploration as an extension of the original application. While it may not significantly lead to any additional engagement, it does seem to engage as well and can be used as additional content to the application. The leaderboard, however, can have negative effects on a player's engagement and motivation and should only be implemented in a controlled competitive environment where it would lead to a positive experience. Chapter 8 explored the interaction between gamification and education in dynamic learning technologies. This was achieved by conducting a randomized controlled trial involving 185 elementary school students from 9 elementary schools in the northeast where students played with the elementary school version of FH2T (FH2T-E) for four (4) twenty-minute sessions across 3 weeks. Students were randomly assigned into one of three conditions: (1) a paper-based worksheet (control); (2) a non-gamified version of FH2T-Elementary where rewards, background images, and unlockables are removed; and (3) a gamified version of FH2T-Elementary. Our findings suggest that when comparing solely assessment learning gains, our intervention, FH2T-E, worked just as well as the control condition. Additionally, there were no differences in learning gains when comparing the gamified version of FH2T-E to the non-gamified version. However, these findings were based solely on the assessment data and did not utilize any of the data from the in-app measures that are recorded when using FH2T-E. After we ran an exploratory factor analysis, two significant factors found in the analysis were associated with the students' levels of engagement as well as their progression in the application throughout the study. Using these factors, we found that students in the gamified condition averaged 6.58 points higher on the posttest than students in the non-gamified condition and students who progressed faster and completed more unique problems in the app may demonstrate higher posttest scores. Thus, we believe this indicates that gamification can also lead to increased learning gains in addition to having beneficial behavioral effects on engagement. Lastly, Chapter 9 introduces an original work, Treasure Hunter, as a culminating effort of the previous six chapters. An initial pilot test found that there were no significant differences in learning gains in pretest and posttest assessment for the control, non-gamified, and gamified versions of the intervention. While there may have been some small scale learning gains in number sense or decomposition skills, the technology itself was never designed to be a learning intervention, but rather an assessment tool for number decomposition and reasoning. An analysis of the in-app data of Treasure Hunter suggests that student performance decreases significantly as the number of terms required increases. The magnitude of the number options provided are also indicative of student performance. When students are given larger numbers, they tend to perform worse than when given smaller numbers. One of the more significant findings of the study is associated with student behavior when given certain operators types. Student performance significantly decreases when a problem allows for more than one operator due to cognitive overload. Additionally, student performance is highest on solely multiplication and solely division problems.

10.2 Conclusions

In this work, I touched on the development of dynamic technologies as well as the infrastructure for supporting data-rich gamified interventions. Based on the findings of the studies conducted in this work and the lessons learned when designing and developing the technologies along the way, I would like to conclude this work with comments about designing gamified learning environments and a brief discussion about the need for better collaboration among educators and computer scientists. In the more empirical part of this work, I discussed the effects of engagement in gamified learning environments. In the four studies that were conducted that involved gamification or gamified components, there were some significant effects and results that were found. For the sake of clarity, I would like to refine and consolidate the findings and offer suggestions for future implementations. Gamification and educational games are inherently different. Educational games are games that feature educational content. Gamification or gamified education consists of educational resources that feature gamified elements. In the four studies of this dissertation, a few components of gamification are explored in terms of behavior and learning gains. Our results suggest that there were some increases in engagement when incorporating gamified elements and, in the case of FH2T-Elementary, the gamified version of the intervention encouraged behaviors that were associated with learning. Statistically, however, it was not found that gamification led to a finding that was so significant that we should upend education as we know it and switch to a gamification-centered alternative. What I have found in my time running studies and developing interventions is that gamification does seem to make these technologies more approachable. Though only anecdotally, we have found that our display stations at expos and conferences that feature gamified technologies such as Treasure Hunter or From Here to There! are busier than when featuring similar non-gamified technology. Younger participants also seemed more willing to try our projects at these fairs when given a gamified version. In our first study, we found that participants pay attention to the rewards and often go back to attain all three clovers of a level before moving on to the next problem. We also found differences in behavior when adding sub-goals that participants could optionally complete for additional points. While these components do not seem to hold weight on their own as major contributing factors in learning, they do seem to add an enjoyable enhancement that doesn't appear to be detrimental to learning. Other gamified elements such as the leaderboard component, however, showed no significant benefit in engagement, however, there were instances of decreased performance (or possibly discouragement) when ranked in lower positions on the leaderboard. As an alternative, I would highly suggest that real-time leaderboards or participant comparisons be dropped and replaced with a more formative assessment of performance. A suitable replacement for a leaderboard would be a high-score board that models some of the other more successful components. In this component, participants would be encouraged to keep moving forward to rank higher and higher on the scoreboard without the possibility of ranking down. With each new ranking achieved, they may be able to unlock a new feature of the application or simply earn a new achievement badge or title. This approach would provide a positive goal without the negative reinforcement or pressure of social standings. Participants would also be able to see clear growth in their ability which is a vital part of self-regulation.

In the earlier chapters of this work, I looked at the benefit of implementing dynamic technology as a way to better understand student progress and strategy. While this could be done without aid of computers or technology such as asking students to show their work or write a description of their thought process, the implementation of technology arguably simplifies this process and allows for the thorough analysis of all recorded data. I believe this is important for two reasons: 1) the data collection and overall process relies less on student efforts and is easily scalable 2) allows for easier statistical analysis and machine learning techniques to be applied. However, this does not seem to be the norm in most educational settings. While some may wonder why or even suggest new ways that this can be attained, I would argue that it is not due to lack of technical ability or concept, rather a lack of communication and collaboration between the developer and the educator that should be addressed. As a developer and educational researcher, I have been able to see and understand both perspectives on user experience and classroom innovation. Simply put, it is difficult for a teacher to communicate classroom needs to a developer without a foundation on how the development process works. Conversely, it is difficult for a developer to assess whether or not their implementations are feasible or not for a teacher or student to use without experiencing the classroom environment firsthand. While technology holds very few limits in possibility, the same cannot be said for education, especially in the average classroom. It is not enough that the technology simply resolves a problem. It should resolve the problem in the least intrusive way possible while minimizing the amount of effort or learning curve that is required to operate it. Generally speaking, there is no shortage of innovative math technology, however, technology that can be easily brought into the classroom and painlessly integrated is hard to find. Educators are very good at educating and are often not good at designing technology. Developers are good at developing technology, but are often not experienced teaching in a classroom. With that said, developers should consider what it means to bring a technology into a classroom and the various factors and burdens that are associated with it. Factors such as device access, Internet access, Internet speed, students remembering passwords, students being distracted, inability to navigate technologies, etc. have very little to do with the development of the technology itself and whether or not it functions properly, but will ultimately outweigh the efficiency of the technology if not accounted for. It is worth expanding on the design and infrastructure of the web-based version of From Here to There! Web. In my first year in the Ph.D. program, we attempted to conduct a study using the iPad version that was developed and has been run in previous in-school studies. However, we found that in our region of the country, iPads were not readily available in most schools or were limited (not 1-to-1) in the few schools that did have them. For that reason alone, it is imperative that we should be abandoning native apps for educational technology if we are creating a general use product and solely deploying for a specific device or platform. Deployment problems such as this occur in most areas of software development and have bolstered some platforms while throttling others. Yielding to native app development also further perpetuates this problem and often establishes the monopolization of specific devices and platforms that eventually become antiquated and obsolete. One example of this is the Windows operating system that dominated the software market for personal and commercial computing for decades. Because most developers solely developed Windows executable files (as opposed to Java or other non-OS specific libraries), users and businesses often could not switch to another operating system due to the lack of non-native support of their essential programs. In classrooms, this effect is even more detrimental because the budgets of schools and districts are limited as is and only allows schools to upgrade and purchase new hardware every few years. Schools that rely on native apps for their devices could theoretically be locked in to an obsolete operating system. Because of this, I developed a browser-based version of our technology that is deployed using a cloud server. While this requires an Internet connection, it is designed to front-load the initial download of the data from the server into the browser's cache storage and then read locally for the duration of the session, thus reducing the need for a high-speed Internet connection. Additionally, it is accessible on nearly every device that can run a modern Internet browser. Because the initial download occurs at the beginning of each session, it is fairly easy to update the remote content and allow updates to disburse to devices without having to go through a third-party app store or install software updates.

10.3 Treasure Hunter Research Platform

Treasure Hunter was developed as a culmination of my research efforts during my graduate school career. After positive reception and response from the test groups that it was demonstrated to, it was important to make sure that this was not just an application that was to be used in controlled environments, but rather something that could be publically available for both research and recreational use. To attain this, online parent and teacher portals were developed that can be used to monitor student progress as they play through Treasure Hunter. Similarly to the tools used when analyzing our randomized controlled trials, parents and teachers are able to track student performance, view in-game measures on each problem attempt, and replay each individual problem attempt in real-time using our playback tool. Each parent account can generate an enrollment code that allows students to link their accounts to their parent's account. This linking can be done retroactively and still view any and all problem attempts that were made before and after the linking occurred.

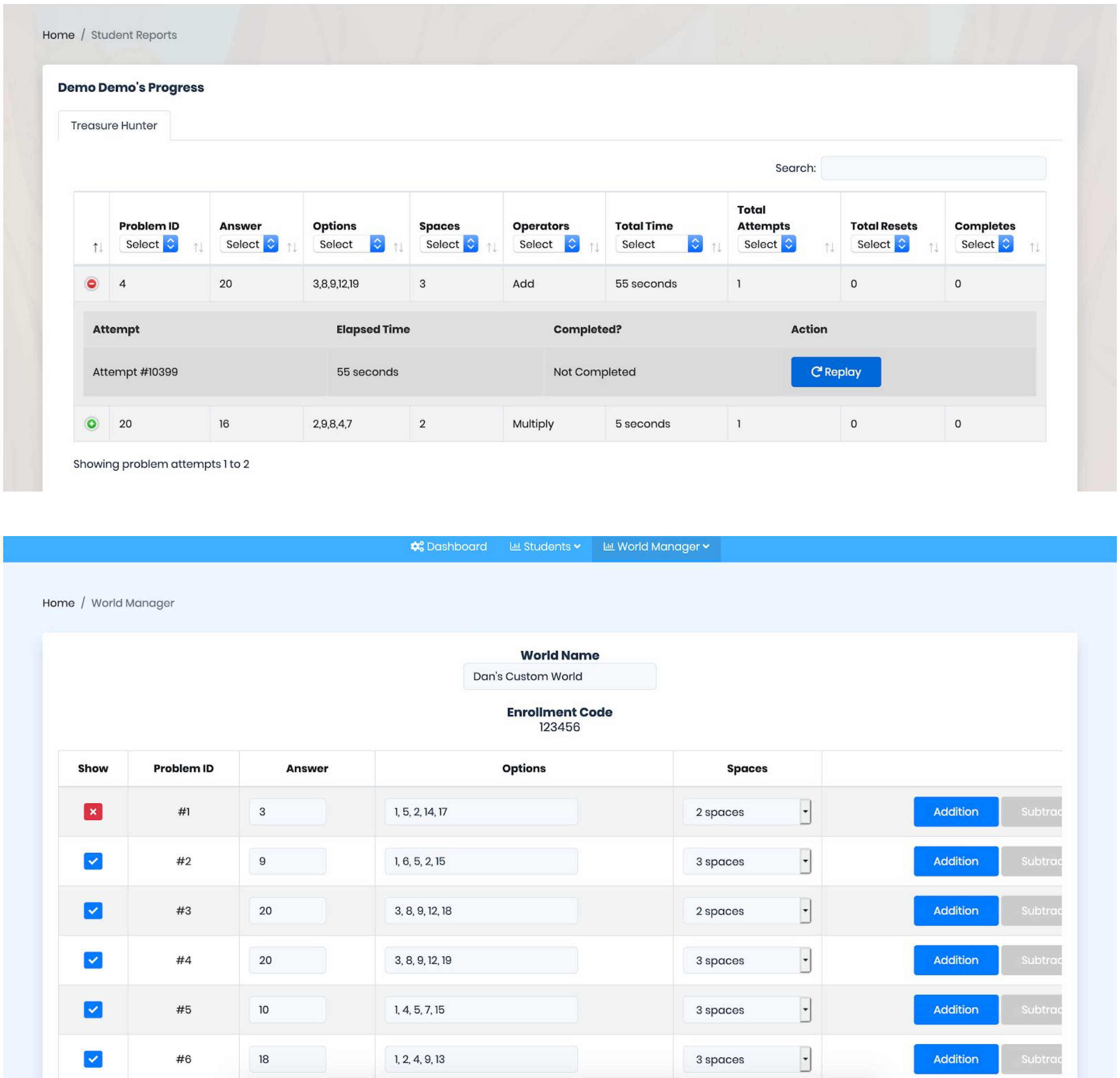


Figure 10.1. Screenshots of the parent and teacher portals of Treasure Hunter where users can access student attempt data and create their own custom worlds in Treasure Hunter

Similarly, teacher accounts provide the same functionality of linking to view problem attempt data. However, unlike parent accounts, teacher accounts can also build their own Treasure Hunter worlds composed of custom problems built with the platform editor. This editor allows teachers to build up to 42 problems across the 7 worlds of Treasure Hunter and select which numbers, spaces, answers, and operators are available in each problem as well as the required amount of gold and gold rewards that each problem has. In addition to viewing the data from the original Treasure Hunter world, parents and teachers will also have access to any data that was created while playing these third-party user-created worlds. We believe that this will allow Treasure Hunter to be embraced more easily in the academic and non-academic worlds by allowing complete transparency in both data retrieval and content generation. While this work will mark the end of my graduate career, I am pleased that my efforts will be available for the benefit of the general public and look forward to developing new learning technologies in the future.

APPENDIX

Appendix A. Gamification vs Non-Gamification Pre-Survey

Math Survey

Please fill out the information below.

* Required

1. **What is your participation code? (ask Research Assistant)** *

   ____

2. **How many math courses have you completed/currently taking?** *
   Check all that apply.
   - [ ] Algebra I
   - [ ] Algebra II
   - [ ] Pre-Calculus
   - [ ] Calculus I
   - [ ] Calculus II
   - [ ] Calculus III
   - [ ] Calculus IV or Higher
   - [ ] Statistics
   - [ ] Trigonometry
   - [ ] Other: ____________________

3. **How good are you in math in general?** *
   Mark only one oval.
   - ( ) Worst in class
   - ( ) Below average
   - ( ) Average
   - ( ) Above average
   - ( ) Best in class

Please select the most appropriate number of each statement which corresponds most closely to your desired response.

4. **"I find math enjoyable"** *
Mark only one oval.

	1	2	3	4	5	6	7	
Strongly Disagree	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Strongly Agree

5. **"Math just doesn't really appeal to me"** *
Mark only one oval.

	1	2	3	4	5	6	7	
Strongly Disagree	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Strongly Agree

6. **"I rarely try math problems not required by the teacher"** *
Mark only one oval.

	1	2	3	4	5	6	7	
Strongly Disagree	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Strongly Agree

7. **"I don't enjoy solving math puzzles"** *
Mark only one oval.

	1	2	3	4	5	6	7	
Strongly Disagree	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Strongly Agree

8. **"I find that math is useful in everyday situations"** *
Mark only one oval.

	1	2	3	4	5	6	7	
Strongly Disagree	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Strongly Agree

9. **"I feel comfortable answering math questions"** *
Mark only one oval.

	1	2	3	4	5	6	7	
Strongly Disagree	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Strongly Agree

10. **"I enjoy playing math-related games"** *
Mark only one oval.

	1	2	3	4	5	6	7	
Strongly Disagree	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Strongly Agree

11. **"Schools should require students to take more math classes"** *
Mark only one oval.

	1	2	3	4	5	6	7	
Strongly Disagree	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Strongly Agree

12. **"Math is boring to me"** *
Mark only one oval.

	1	2	3	4	5	6	7	
Strongly Disagree	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Strongly Agree

13. **"I use math to solve everyday problems"** *
Mark only one oval.

	1	2	3	4	5	6	7	
Strongly Disagree	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Strongly Agree

14. **"Puzzles like Sudoku are a fun way to pass the time"** *
Mark only one oval.

	1	2	3	4	5	6	7	
Strongly Disagree	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Strongly Agree

Appendix B. Gamification vs Non-Gamification Post-Survey

Math Survey

Please fill out the information below.

* Required

1. **What is your participation code? (ask Research Assistant)** *

Please select the most appropriate number of each statement which corresponds most closely to your desired response.

2. **"The application really held my attention"** *

   Mark only one oval.

	1	2	3	4	5	6	7	
Strongly Disagree								Strongly Agree

3. **"I wish I had more time to use the application"** *

   Mark only one oval.

	1	2	3	4	5	6	7	
Strongly Disagree								Strongly Agree

4. **"The problems weren't challenging enough for me"** *

   Mark only one oval.

	1	2	3	4	5	6	7	
Strongly Disagree								Strongly Agree

5. **"I had too much trouble with the gestures and I lost interest"** *
Mark only one oval.

	1	2	3	4	5	6	7	
Strongly Disagree	◯	◯	◯	◯	◯	◯	◯	Strongly Agree

6. **"I would prefer solving math problems on paper instead of using the application"** *
Mark only one oval.

	1	2	3	4	5	6	7	
Strongly Disagree	◯	◯	◯	◯	◯	◯	◯	Strongly Agree

7. **"The application was very boring"** *
Mark only one oval.

	1	2	3	4	5	6	7	
Strongly Disagree	◯	◯	◯	◯	◯	◯	◯	Strongly Agree

8. **"I wish I had this tool when I was learning math"** *
Mark only one oval.

	1	2	3	4	5	6	7	
Strongly Disagree	◯	◯	◯	◯	◯	◯	◯	Strongly Agree

9. **Is there anything else you'd like to add?**

____
____
____
____
____

Appendix C. Leaderboard Study Pre-Assessment

Background info

Welcome to the study. This study contains 1 survey, 1 tutorial, and 2 main tasks. We will be asking you to answer questions and solve math problems. This study will take approximately 1 hour and you will be directed to the next part when you complete this survey. All questions are optional and you may stop at any time if you feel uncomfortable. Thank you!

Age:

Gender:

- ○ Male
- ○ Female
- ○ Other

Year in school at WPI:

- ○ First Year
- ○ Second Year
- ○ Third Year
- ○ Fourth Year
- ○ Fifth Year
- ○ Other

Major(s):

Minor(s):

Please list the college-level math courses that you have taken and are currently enrolled in:

Are you currently enrolled in a math course?

- Yes
- No

Math Courses

How many math courses have you taken?

- ☐ Algebra I
- ☐ Algebra II
- ☐ Pre-Calculus
- ☐ Calculus
- ☐ Calculus II
- ☐ Calculus III
- ☐ Calculus IV or Higher
- ☐ Statistics
- ☐ Trigonometry
- ☐ Other ______

How good are you in math in general?

- ○ Worst in Class
- ○ Below Average
- ○ Average
- ○ Above Average
- ○ Best in Class

How competitive are you in general?

- ○ Least Competitive
- ○ Not Very Competitive
- ○ Average
- ○ A Little Competitive
- ○ Very Competitive

Please rate how much you agree with each statement on a scale of 1-7.

"I find math enjoyable"

1: Strongly Disagree	2	3	4	5	6	7: Strongly Agree
◯	◯	◯	◯	◯	◯	◯

"Math just doesn't really appeal to me"

1: Strongly Disagree	2	3	4	5	6	7: Strongly Agree
◯	◯	◯	◯	◯	◯	◯

"I rarely try math problems not required by the teacher"

1: Strongly Disagree	2	3	4	5	6	7: Strongly Agree
◯	◯	◯	◯	◯	◯	◯

"I don't enjoy solving math puzzles"

1: Strongly Disagree	2	3	4	5	6	7: Strongly Agree
◯	◯	◯	◯	◯	◯	◯

"I find that math is useful in everyday situations"

1: Strongly Disagree	2	3	4	5	6	7: Strongly Agree
◯	◯	◯	◯	◯	◯	◯

"I feel comfortable answering math questions"

1: Strongly Disagree	2	3	4	5	6	7: Strongly Agree
◯	◯	◯	◯	◯	◯	◯

"I enjoy playing math-related games"

1: Strongly Disagree	2	3	4	5	6	7: Strongly Agree
◯	◯	◯	◯	◯	◯	◯

"Schools should require students to take more math classes"

1: Strongly Disagree	2	3	4	5	6	7: Strongly Agree
◯	◯	◯	◯	◯	◯	◯

"Math is boring to me"

1: Strongly Disagree	2	3	4	5	6	7: Strongly Agree
◯	◯	◯	◯	◯	◯	◯

"I use math to solve everyday problems"

1: Strongly Disagree	2	3	4	5	6	7: Strongly Agree
◯	◯	◯	◯	◯	◯	◯

"Puzzles like Sudoku are a fun way to pass the time"

1: Strongly Disagree	2	3	4	5	6	7: Strongly Agree
◯	◯	◯	◯	◯	◯	◯

Appendix D. Leaderboard Study Post-Assessment

Background info

Thank you for participating in the study! Your credit will be awarded within 24 hours.

In FH2T (Task 1), how did your position on the leaderboard (score board) affect your play time?

- ○ The leaderboard motivated me to play longer
- ○ The leaderboard demotivated me from playing longer
- ○ The leaderboard did not impact my playing
- ○ I did not pay attention to the leaderboard at all

Which position on the leaderboard were you in when you stopped Task 1?

- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ I don't remember

How much do you agree with the following statements about Task 1?

I liked doing this activity.

- Strongly Agree
- Agree
- Neutral
- Disagree
- Strongly Disagree

I would like to do more activities like this.

- Strongly Agree
- Agree
- Neutral
- Disagree
- Strongly Disagree

The material in Task 1 was difficult for me.

- Strongly Agree
- Agree
- Neutral
- Disagree
- Strongly Disagree

I worked hard on understanding the material in this activity.

- ○ Strongly Agree
- ○ Agree
- ○ Neutral
- ○ Disagree
- ○ Strongly Disagree

This activity made me feel more like I am good at math.

- ○ Strongly Agree
- ○ Agree
- ○ Neutral
- ○ Disagree
- ○ Strongly Disagree

This activity made me feel that math is fun.

- ○ Strongly Agree
- ○ Agree
- ○ Neutral
- ○ Disagree
- ○ Strongly Disagree

I liked the way the material was presented on the screen.

- ○ Strongly Agree
- ○ Agree
- ○ Neutral
- ○ Disagree
- ○ Strongly Disagree

"I find math enjoyable"

1: Strongly Disagree	2	3	4	5	6	7: Strongly Agree
◯	◯	◯	◯	◯	◯	◯

"Math just doesn't really appeal to me"

1: Strongly Disagree	2	3	4	5	6	7: Strongly Agree
◯	◯	◯	◯	◯	◯	◯

"I rarely try math problems not required by the teacher"

1: Strongly Disagree	2	3	4	5	6	7: Strongly Agree
◯	◯	◯	◯	◯	◯	◯

"I don't enjoy solving math puzzles"

1: Strongly Disagree	2	3	4	5	6	7: Strongly Agree
◯	◯	◯	◯	◯	◯	◯

"I find that math is useful in everyday situations"

1: Strongly Disagree	2	3	4	5	6	7: Strongly Agree
◯	◯	◯	◯	◯	◯	◯

"I feel comfortable answering math questions"

1: Strongly Disagree	2	3	4	5	6	7: Strongly Agree
◯	◯	◯	◯	◯	◯	◯

"I enjoy playing math-related games"

1: Strongly Disagree	2	3	4	5	6	7: Strongly Agree
◯	◯	◯	◯	◯	◯	◯

What did you think of From Here to There (Task 1)? Any comments or suggestions?

What did you think of Treasure Hunter (Task 2)? Any comments or suggestions?

Anything else you'd like to add?

Appendix E. Treasure Hunter Assessment

Question 5. Select all of the sets that add up to 9.

2 7	**6 4**	**3 5**
7 3	**4 5**	**8 1**
6 2	**1 7**	**6 3**

Welcome! Please answer the following problems.

What is your age?

What is your gender?

Questions 1-4. Enter the numbers in the number boxes below to complete the problem.

1) $7 + 3 = \square + 4$

2) $5 + 3 = \square + 3$

3) $2 \times \square + 2 = 12$

4) $\square \div 3 + 2 = 5$

Questions 6-9. Enter the numbers in the number boxes below to complete the problem.

<u>Example</u>
$? \times ? = 8$
$2 \times 4 = 8$

$\square \times \square = 15$

$\square \div \square = 12$

$\square + \square \div \square = 5$

$\square - \square \times \square = 7$

Finished