DESIGNING EDUCATIONAL VIDEO GAMES AND INTELLIGENT TUTORING SYSTEMS FOR DRILL-BASED TRAINING

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Zikai Wen

August 2021

DESIGNING EDUCATIONAL VIDEO GAMES AND INTELLIGENT TUTORING SYSTEMS FOR DRILL-BASED TRAINING

Zikai Wen, Ph.D.
Cornell University 2021

Drill and practice is a well-received approach to repeatedly train learners' skills through a series of exercises and to reward them with corrective feedback. However, drill-based training may not improve learners' performance if its exercises are badly designed (e.g., not fun, not relevant to the learning goal, and becoming too difficult or too simple). To make drill-based training more effective, researchers have been designing two types of e-learning tools: educational video games to motivate learners to practice and intelligent tutoring systems to personalize the exercises. Nonetheless, the existing e-learning tools have two main problems: (1) not providing situational context for training, which limits learners' ability to apply previously learned knowledge and skills to real-life situations; (2) not designed for learners with specific learning disabilities (i.e., dyslexia, dyscalculia, and dysgraphia).

To design educational video games that provide situational context for training, I sought to leverage entertainment technologies (e.g., storytelling technology) and AI technologies (e.g., computer vision technology). I tackled two specific projects. The first project was to improve anti-phishing training by simulating actual phishing attacks in a role-playing game. The key design challenge was to design an intensive workplace atmosphere that is easy for learners to make mistakes but also keeps the pleasure of completing tasks. My approach was to design an interactive storyline about building business as a banker and

to design tragic endings that are realistic but with a sense of black humor. The second project was to teach vocabulary for objects located in the player's immediate vicinity. The key design challenge was to guide learners to interact with learning materials in the physical world. To meet this challenge, I designed a new selection highlighting mechanics for AR scenes and designed an AR progress bar to indicate players' selection progress. Evaluations showed that my two educational video games for training were fun and also improved learners' post-test performance in practice.

To help learners with SLDs use math e-learning tools for math skills training, I started by conducting an interview study with teachers for SLDs to study the difficulties that learners with SLDs faced in using e-learning tools. According to the interview study findings, learners with SLDs needed teachers to help them manage negative emotions during drills and practice. However, teachers were often unavailable during students' independent exercises. Therefore, I designed an intelligent tutoring system that detects and mitigates learners' negative emotional behaviors. This system analyzes eye-gazing data together with other traditional input data to detect learners' negative emotional behaviors. I also designed four intervention methods to mitigate the negative emotional behaviors: (1) praising for correct steps to solve the problem, (2) providing hints, (2) switching to a simpler problem, and (4) offering brain breaks. I conducted a formative study with teachers for SLDs to refine the design of the intelligent tutoring system. Teachers agreed that this system would help learners with SLDs reduce negative emotional behaviors. They also suggested that the system, in the future, should personalize the detection of negative emotional behaviors to help students who have more severe learning disabilities.

Overall, my dissertation answered three main research questions: (1) how to design educational video games that provide situational context to practice skills; (2) whether and how learners with SLDs have difficulties in using the existing e-learning tools to practice math skills; and (3) how should we design the e-learning tool to intelligently tutor learners with SLDs. At the end of my dissertation, I discuss the limitations of my doctoral work and propose potential future directions for designing educational video games and intelligent tutoring systems for drill-based training.

BIOGRAPHICAL SKETCH

Zikai Alex Wen (温子剀 in Chinese) is a Ph.D. Candidate in the Department of Computer Science at Cornell University working on educational technology and entertainment technology under the supervision of Prof. Shiri Azenkot, Prof. Rene Kizilcec, and Prof. Serge Belongie. Specifically, his research focuses on designing and building educational video games and intelligent tutoring systems for learners with different learning abilities and needs. Before joining Cornell, he received his B.S. in Computer Science from University of Strathclyde, Glasgow, UK, and from Beijing University of Chemical Technology, Beijing, China, in 2014.

To my parents, 温 少华 (WEN Shaohua) and 金 琨 (JIN Kun), who have been supporting me wholeheartedly throughout my Ph.D. journey.

ACKNOWLEDGEMENTS

First, I would like to acknowledge Prof. Shiri Azenkot for her incredible support. During my Ph.D. study, she has been giving me the freedom to work on the research projects that I liked, helping me to identify important research questions, and teaching me how to explain my work clearly. She opened the door to HCI research for me and taught me how to do good HCI research.

I would also like to thank Dr. Erik Andersen for introducing me to the exciting educational technology research field. I was deeply inspired by his passion for designing engaging video games for educational purposes. He showed me that learning should be fun and that using and developing computer technologies to facilitate learning is the key to make it happen.

I am also super lucky to have the support from Prof. Rene Kizilcec, and Prof. Serge Belongie. Rene can always point out interesting connections between my research. He also encouraged me and gave me inspiration whenever I was stuck in my research. I learned a lot from Serge about how my research would be useful for other applications in AR/VR.

I have been very lucky to work with many smart and inspiring researchers. In particular, I want to thank Prof. Yuhang Zhao, Dr. Lei Shi, Dongfang Gaozhao, Erica O Silverstein, Anjelika Lynne S Amog, Prof. Katherine Garnett, Zhiqiu Lin and Rowena Chen. It has been a great pleasure collaborating with them. Working together with them has made my Ph.D. life an unforgettable joyful memory.

Last but not least, I would like to dedicate my dissertation to my parents, WEN Shaohua and JIN Kun, for supporting me wholeheartedly during the rough time of my Ph.D. It would not be possible for me to earn my Ph.D. degree without them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

To increase fluency in content knowledge and skills, learners need to repeatedly and persistently practice the skills. Drill and practice is a well-received approach to repeatedly train learners to practice skills through a series of exercises and to reward them with corrective feedback [116]. However, drill-based training may not improve learners' performance if its exercises are badly designed (e.g., not fun, not relevant to the learning goal, or becoming too difficult or too simple) [104, 82].

To improve the training outcomes of drill and practice, recent research has studied using computer technology to design drill-based training that is engaging and personalizable. Researchers designed two types of computer-based drill and practice: educational video games (e.g.,[175, 197, 144]) and intelligent tutoring systems (e.g., [39, 93, 83]).

The primary goal of designing educational video games is to engage learners in repetitive training, especially to motivate them to practice independently. Prior research [134, 44, 138] has suggested that learners love playing educational video games more than training in the traditional way (e.g., using flashcards). Their studies also showed that the gaming group performed better than the traditional learning group in post-tests.

However, research studies [119, 159] found that learners could not fluently use the in-game skills to solve real-world problems. For example, learners enjoyed playing educational video games to quickly memorize and recall the pattern of malicious URL addresses but they still could not recognize the suspicious URL addresses during daily online communications [159]. One plausible reason is that the practice problems in the games (e.g., [14, 175, 127]) are similar to paper-based quizzes. But situated learning theory [113] has emphasized that practice problems must be authentic and relevant to the situation in which learners use the skills. Otherwise, learners may find it difficult to apply previously learned knowledge and skills to real-life situations. Many educational video games, however, did not provide situational context.

Therefore, my first key research topic is to study ***how to design educational video games that provide situational context for skills training***. Prior research on educational video games [75, 19, 50] has designed situational context for games for K-12 education subjects. For example, ST Math [125] for math skills training designed cute characters and animations to visualize math word problems. However, we still do not have a good design in situated learning games for two training programs: anti-phishing training [159] and vocabulary learning [132]. So I tackled two specific educational video game design projects, one for anti-phishing training and one for vocabulary learning.

The key challenge of designing a situated learning game for anti-phishing training is to create an intensive workplace atmosphere in a virtual world that makes learners fall for phishing attacks. Meanwhile, it is equally important to design incentives for players to enjoy practicing the exercise in such an intensive environment. My approach was to design an interesting story-line about acquiring business contracts as a banker. I also adapted the document inspection game mechanics in *Papers, Please* to the anti-phishing context of processing business emails. In case of learners being phished in the game, I designed tragic endings that are not only realistic but also have a sense of black humor.

I explain the gameplay design details and the implementation details in Chapter 2. I also present in Chapter 2 the result of a user study that compared my game with a standard form of training and a competing training game (which does not simulate phishing attempts through role-playing). The results suggested that my game is more engaging and effective in improving learners' performance than the other two existing training methods.

My second educational video game design project was to design a vocabulary learning game based on what the learner can see. Although recent work has used AR technology to design vocabulary learning games that can recognize the learner's immediate vicinity [158, 92, 185], their games relied on using QR-Codes or pre-installation. As a result, learners were not able to use the existing AR games to learn vocabulary anywhere at any time. To meet the learners' needs, I researched using computer vision technology, specifically object detection and recognition neural networks, to design an AR vocabulary scavenger hunt game.

I present in Chapter 3 the details of adapting computer vision technologies for my AR game design. I also present my user study results in Chapter 3, which showed that players found my AR game to be more engaging than a traditional on-screen game and were more likely to remember learned vocabulary after a five-day delay.

It is inadequate to only explore the design space of gameplay for situated learning because individuals' skill level and learning ability affect their learning experience and performance. To personalize drill-based training, prior research has developed intelligent tutoring systems [39, 93, 83] that can personalize exercise problems by analyzing learners' learning activities and estimating their proficiency level. Take math skills training, for instance, many drill-and-practice websites (e.g., ASSISTments and Lynnette) analyze the correctness rate of historical answer submissions to choose the suitable math problem to present next. Research [190] showed that intelligent tutoring systems effectively improved general education learners' math skills while reducing the total time spent on drill-based training.

However, we do not know whether these e-learning tools are also effective for learners with specific learning disabilities (SLDs). SLDs predominantly affect learners' abilities to understand or use language, especially in reading (dyslexia), math (dyscalculia), and writing (dysgraphia) [171]. Math skills training, for instance, is more challenging for learners with dyscalculia by definition, and learners with other SLDs may also struggle with math problems that involve reading and writing.

Since many existing math e-learning tools were not designed for learners with SLDs, my second key research topic is to study ***whether and how learners with SLDs have difficulties in using the existing math e-learning tools to practice math skills***. Furthermore, if the current design of math e-learning tools does not work well for learners with SLDs, ***how should people design the e-learning tool to intelligently tutor learners with SLDs?***

To study the difficulties that learners with SLDs faced in using math e-learning tools, I conducted semi-structured interviews with teachers for learners with SLDs. I first talked to the teachers because they are the stakeholders who incorporate e-learning tools into the curriculum. Teachers would evaluate the effectiveness of e-learning tools then recommend their students the suitable tools to use.

By coding the semi-structured interviews, I found that learners with SLDs had difficulties in using math e-learning tools because of the following reasons: text-intensive user interfaces, inability to adjust difficulty levels, and not reacting to learners' anxiety and negative emotional behaviors. Based on the findings, I distilled design implications to help shape the design of more inclusive and effective e-learning tools for students with SLDs. I present the details of my findings and discuss the implications of my findings in Chapter 4.

To improve the usability of math e-learning tools for students with SLDs, I designed an intelligent tutoring system that automatically detects and mitigates learners' negative emotional behaviors. Assisting learners in self-managing negative emotions is an essential accommodation for learners with SLDs according to the special education studies [73, 64]. However, as suggested by my interview study, many existing math e-learning tools did not offer this accommodation, but relied entirely on the help of teachers. Therefore, I sought to design an intelligent tutoring system that can automate the teacher's help with self-managing emotions. To form this design, I conducted formative studies with five teachers for learners with SLDs.

I designed the intelligent tutoring system to automatically detect negative emotional behaviors by analyzing eye-gazing activities, inputs on the touchscreen, and response time. To confirm the system's speculation, I also designed a dialogue system that asks learners for their emotional state to decide how to intervene in the exercises. To mitigate learners' negative emotional behaviors, the teachers and I designed four intervention methods: (1) praising for correct steps to solve the problem, (2) providing hints, (2) switching to a simpler problem, and (4) offering brain breaks. I elaborate the preliminary design and report teachers' opinions of my system in Chapter 5.

To summarize, my doctoral dissertation has three main contributions: (1) explored two different approaches to design educational video games that provide situational context for drill and practice: one approach is to design a simulated virtual world using storytelling technologies; and the other approach is to use computer vision technologies to augment the learning materials in the real world; (2) studied the difficulties that learners with SLDs faced in using e-learning tools to practice math skills; and (3) proposed the design of an intelligent tutoring system that can detect and mitigate learners' possible negative emotional behaviors.

This paragraph describes the outline of the rest of my dissertation. Chapter 2 presents the design of *What.Hack*, which is the first type of anti-phishing training game that simulates actual phishing attacks in a role-playing game. Chapter 3 presents the design of *CollectiAR*, which is the first type of vocabulary learning game that uses computer vision to create vocabulary scavenger hunt activities in learners' immediate vicinity. Chapter 4 reports a semi-structured interview study with teachers for SLDs to study their experiences and their students' experiences in using existing math e-learning tools. Chapter 5 presents the design of an intelligent tutoring system for students with SLDs, which helps the learners self-manage their negative emotional behaviors. Chapter 6 concludes the dissertation with a summary of the contributions and the limitations of my research and proposes potential future directions for designing educational video games and intelligent tutoring systems for drill-based training.

CHAPTER 2

WHAT.HACK: ENGAGING ANTI-PHISHING TRAINING THROUGH A ROLE-PLAYING CYBER DEFENSE SIMULATION GAME

2.1 Introduction

Phishing is the act of deceiving people into divulging information or unintentionally installing malware on their computers by sending the victim(s) counterfeit emails [88]. These counterfeit emails work by misleading the victim into thinking they come from a legitimate source. For example, a phishing email can link to an imitation of the PayPal login screen. Victims who believe the link is legitimate will enter their login credentials to the fake site, unwittingly giving the hackers access to their PayPal account. In addition to financial gain, government-backed hackers may disrupt elections by phishing specific persons who are affiliated with powerful institutions. In the 2016 US election, John Podesta, the chairman of Hillary Clinton's campaign, clicked on the change password link in a phishing email intended to look like a Google warning [200]. His action immediately unlocked some or all of his emails to the hacker.

To repel phishing attacks, phishing defense technology has evolved rapidly. Recent automatic systems apply machine learning to classify phishing emails, but these automated approaches are not foolproof [18]. There remains a non-negligible probability of users receiving phishing emails and these users must decide whether an email in their inbox is phishing or legitimate.

Research [62] shows that hackers can effectively and efficiently target end users due to the public's general lack of awareness regarding information security. People who are prone to taking risks are more likely to be phished [164]. Even if people are aware of phishing, simply knowing does not provide useful strategies for identifying phishing attacks [65]. On top of this, phishing attacks often convey a sense of urgency or utilize threats to pressure the recipient into responding [112]. Due to these factors, people judge the legitimacy of incoming message by visual cues, which can be easily copied by phishers [106]. Therefore, user education is a vital approach to protect users against phishing. While many organizations provide materials on how to defend against phishing attacks, such as email bulletins or information security websites [48], studies [108] found that these kinds of materials only work if people keep paying attention to them.

To better engage learners and change user behaviour, several anti-phishing games have been proposed. Although evaluations of these games have demonstrated an improvement in their players' ability to identify phishing websites, existing games leave out email context that hackers often leverage to demand immediate attention and encourage rash decision making. A person who can quickly parse a URL might hurriedly click on the hyperlink syntax without hovering on it because it seems to be an urgent request from the boss. Moreover, these game designs are not particularly effective at teaching players how to detect combined phishing techniques. For example, some who have played those games might not fall for malicious URLs in a phishing email, but they might click on the malware attachment enclosed in the same email. Incorporating these combined phishing techniques into our game's design can lead to more engaging challenges and more practical knowledge.

To develop a comprehensive anti-phishing game, we designed *What.Hack* (pronounced what dot hack), an online simulation game that features an engaging sequence of puzzles. Each puzzle requires players to study the anti-phishing rules in a rulebook that help evaluate whether an email is legitimate or phishing. Phishing emails in the game are generated by templates collected from real phishing emails. Players need to carefully identify phishing emails or they will encounter a bad ending (e.g. loss of integrity in the game).

To analyze the effectiveness of *What.Hack*, we tested the impact of the game on players' ability to recognize the real incoming phishing emails that came from a database of emails collected by the authors' university. We compare it against the current non-gamified training that the university uses and a good competing anti-phishing game, *Anti-Phishing Phil* [165]. We found that *What.Hack* achieved a 36.7% improvement in players' correctness in identifying incoming phishing emails from pretest to posttest, but did not find a statistically significant improvement for the training materials or *Anti-Phishing Phil*, which indicates that our context-based approach is effective for training people to recognize and defend against phishing emails. Moreover, examination of the feedback from players and logs of their play activity shows that players found the game engaging.

2.2 Related Work

2.2.1 Games for Learning

Educational games can improve learners' performance, especially knowledge that is best learned actively through experience rather than passively. For example, *Crystallize* [53], a 3D video game for learning the Japanese language, showed that situated learning can be effective and engaging. *Reduct* [23] demonstrated that novices can learn programming concepts by playing a game. This suggests that gamifying education is potentially beneficial.

Anti-phishing education often struggles to capture the interest of end users. Materials commonly used for cybersecurity training include notes, videos, and email bulletins. However, these materials are often not very engaging and separate the learning material from the context in which employees routinely apply this information (e.g. email clients). Staff interviews by Conway et al. [48] revealed a continued desire for engaging cybersecurity materials that tie into daily experiences and practices. An investigation into the current state of cybersecurity education in industry produced similar conclusions [159]. According to both reports, anti-phishing education is working and more people are aware of the concept of phishing, but more work needs to be done. Our goal is to replace training programs that typically emphasize readings on cybersecurity with a role-playing game that mimics the actual situation of being phished.

2.2.2 Simulating Situational Context through Role-playing

Using the role-playing approach to engage students and improve their learning transfer performance has been a well-known design strategy in gamified education. For example, *Quest Atlantis* [28] is a 3D virtual learning environment that allows students to work together to perform educational activities that are known as Quests. There are other successful role-playing task-solving simulation games designed for the purpose of science and ethics education [100, 162], engineering internships training [43], etc. These games indicate the potential for supporting increased levels of engagement and learning across different domains.

Research in learning theory also supports the idea of presenting information in context. The theory of situated learning [113] stipulates that "the potentialities for action cannot be fully described independently of the specific situation " [20]:6. Shaffer [162] drew upon situated learning literature on communities of practice [113] when introducing epistemic frames for supporting learning transfer. According to encoding specificity theory [183], recall is highest when the context in which something is learned is perceptually similar to the context in which it is used. As Gee suggested, games that engage players in authentic situated problem solving facilitates learning can be transferred out of the game [75]. Experiments on tutorials in games [19] also showed evidence for the efficacy of presenting information in context in the case of the protein-folding game *Foldit* [50].

Game Type & Examples	Description	System Attack	URL Phishing	Spear Phishing
Board Games [59, 169, 147, 191, 80]	Teach high-level security concepts.	✓	✓	✓
Capture-The-Flag [198, 155, 41, 24, 10, 139]	Let coders compete for scores by defending their systems and hacking others'.	✓		
Sys-Attack Sim RPG [47, 9, 13, 177]	Teach players to defend against computer system attacks in a realistic system attacks simulation game.	✓		
Non-Phishing RPG [22, 165, 127, 206]	Teach players to identify phishing URLs in a cartoon-like game without phishing attempts.		✓	
Phishing Sim RPG *What.Hack*	Teach players to defend against URL and spear phishing attempts in a realistic phishing simulation game.		✓	✓

Table 2.1: Game-based Cybersecurity Training Designs Comparison

2.2.3 Game-based Cybersecurity Training Designs Review

Recent surveys [141, 169, 178] have summarized the current state of game-based cybersecurity training designs. We categorized the list of games mentioned in these surveys and compared their game type, target audience and design objectives with *What.Hack* in Table 2.1. We excluded cybersecurity games for children [8, 12, 15, 98] because they have a slightly different game design goal. In general, these games focus on introducing basic tips of staying safe online. Tips for kids are easier to practice in general and some might even be considered inapplicable under most real-world situations such as a corporate workplace. For instance, Carnegie Cadets [8] suggests that one should look at the sender's username and the subject title to identify spam emails, which is a habit that is actually exploited by some phishing attacks. In the following subsections, we discuss each of the game types listed in Table 2.1.

Board Games. *Control-Alt-Hack* [59] is a board game that teaches players high-level security concepts such as phishing, social engineering, etc. While this does help to increase awareness and understanding of cybersecurity topics as a whole, it is not sufficiently specific enough to simulate the low-level decisions required for anti-phishing strategies in practical contexts. In general, board games in information security [169, 147, 191, 80] are not meant to teach hands-on security skills, such as how to identify phishing attacks, which are the main focus in our game.

Capture-The-Flag. Capture-The-Flag (CTF) is a game-based computer security competition for students to practice skills of defending against hackers. There are two types of cybersecurity CTF: attack-defend and Jeopardy-style [56]. In attack-defend CTF, each team attacks other teams' servers and protects their own server. The "flags" are files in the defending computer that the attack team attempts to retrieve as they compromise the computer. In Jeopardy-style CTF, teams solve puzzles by using knowledge like cryptography, coding, etc. Solving a puzzle means they capture a flag. The team with highest number of flags wins. Researchers [56, 177, 126] have shown that CTF is an engaging and effective way to motivate coders to learn how hackers think and how to defend against them. However, these competitions usually require basic coding background to learn about hacking techniques. Therefore, they are not directly applicable for teaching the general public about anti-phishing techniques.

System-Attack Simulation Role-playing Game. To engage the general public to learn how to defend a computer system or network settings from being compromised, CyberCIEGE [177] and similar approaches [47, 9, 13] adopted an attack simulation role-playing design that is similar to our game but focuses instead on system attacks. A user study found CyberCIEGE to be engaging even when students knowingly fail, but did not assess learning effectiveness [177]. In this work, we further examine whether the potential for increased effectiveness and self-efficacy in handling real-world cyber threats justifies the effort of designing a role-playing game to simulate situated phishing attempts.

Non-Phishing Role-playing Games. Recently, new designs for anti-phishing games [22, 127] have drawn inspiration from the popular game design framework for anti-phishing training [21] and *Anti-Phishing Phil* [165], which do not simulate situated phishing attempts. *Anti-Phishing Phil* is a representative Non-Phishing RPG that teaches players how to identify phishing URLs. In each round, players act as a fish to "eat" the worm that shows safe URLs and "reject" the bait that shows phishing URLs. Visualizing URLs as worms makes the

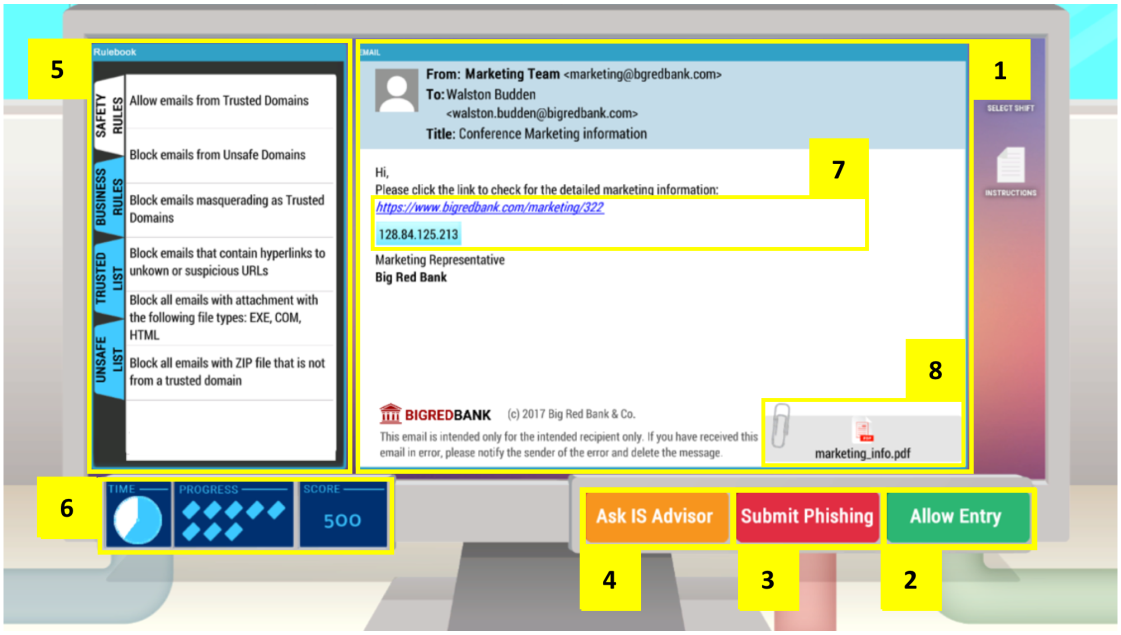

Figure 2.1: In *What.Hack* players process emails to acquire contracts and to protect their network from cyber criminals (1). Player either select *Allow Entry* button to let the email reach the recipient (2) or select *Submit Phishing* button to block it from doing potential harm (3). If players cannot determine whether the email is dangerous, they can click *Ask IS Advisor* to take some time to analyze the email (4). Players can refer to the rulebook to decide whether this email is legitimate or phishing (5). Players need to process a number of emails within a time limit (6).

game fun, but it also takes the URLs out of context and does not simulate the real experience of detecting phishing emails.

The authors of *Anti-Phishing Phil* mention two limitations. One is their focus on URLs and domain names, which are only two of the phishing attack templates, leaving the player vulnerable to content-based attacks like spear phishing attacks. The other limitation is their exclusive focus on URL syntax without addressing URL semantics, meaning their players are still vulnerable to URL semantics attacks (safe URL syntax redirects to forged URL address). These limitations also apply to recent games [22, 127] that are extended from it. One of our main goals with *What.Hack* is to achieve more holistic anti-phishing education.

An anti-phishing game design framework [21] includes design elements such as safeguard effectiveness and perceived susceptibility. Multiple Non-Phishing RPGs [22, 127, 206] as well as *What.Hack* address all of these elements. That said, we argue that this framework is sound but does not sufficiently address the identification of real-world phishing threats if design elements are not implemented in a context that resembles the one used by hackers. Evidence from a psychological study [148] showed that certain types of phishing emails, such emails that sound like they come from a friend or warn of some kind of failure, are more effective than other types, such as emails that offer a deal, regardless of the specific phishing templates the hackers use. Our goal is to implement phishing attempts in a way that closely matches how they occur in the real world.

2.3 Gameplay Design

We have developed a prototype game, *What.Hack*, which facilitates role-playing for learning email phishing defense. A screenshot of *What.Hack* is shown in Figure 2.1. *What.Hack* was designed with three primary learning goals: 1) teach email phishing defense in context by replicating as many real-life conditions as possible, 2) engage the player by setting clear goals and tasks that become more difficult over time, 3) provide immediate feedback about the consequences of decisions the player makes.

What.Hack attempts to teach phishing emails defense throu-gh the game mechanics themselves. The core game mechanics encourage players to screen incoming emails to determine whether they might be malicious, and provide a set of constraints ("rules") that players can use to evaluate an email. Over time, the rules become more specific and combine to create a complex rule set, simulating the large amount of constraints that must be applied in real life.

The key challenge in constructing situated learning approaches for anti-phishing centers on how to replicate the perceptual characteristics of the situations in which real people defend against phishing attacks in a way that is fun, educational, and not tedious. To do this, we designed a training module that provides rich visual and situational context. In *What.Hack*, the player assumes the role of an employee working for a bank. To win, players need to help their bank acquire contracts by processing business-related emails and avoiding being phished at the same time.

2.3.1 Porting a Subset of *Papers, Please* Game Mechanics to a Phishing Simulation Game

Since processing emails is not generally considered compelling, we investigated existing games with game mechanics centered around document inspection. We observed that *Papers, Please* [121], a popular and highly engaging game released in 2013, focused heavily on document inspection. In this game, the player acts as a border patrol officer. The player must review the passports and other supporting documents of entrants seeking admission to a fictional communist country. To do this, the player must apply a set of rules specified by the border control office, which are expressed to the player through a rulebook. The player is directed to accept only passports that come with valid paperwork and reject or detain those with improper forms.

While designing *What.Hack*, we adapted some of these document inspection mechanics to the anti-phishing context of processing business emails to foster motivation for learning email safety rules. In our game, players can allow their colleagues' business emails to go through or to block them, which will affect their company's success. Learning email safety rules becomes necessary for players to make progress in the game. With the shift of purpose to anti-phishing training, players must now prevent their company from being hacked and help it to prosper.

Papers, Please includes a range of game mechanics that we did not adopt in *What.Hack*. *Papers, Please* makes a critique of Orwellian communist bureaucracy and dehumanizing processes [131]. In addition to the narrative, the game introduces several ethical dilemmas, forcing the player to choose between obeying oppressive laws and taking risks to advance moral causes. Since we did not incorporate critiques or ethical dilemmas into *What.Hack*, a significant portion of what makes *Papers, Please* compelling and unique was lost in this conversion. However, we consider this acceptable since our primary goal was to adapt the document inspection mechanics for anti-phishing, not to raise questions about the nature of the organization in which the player operates.

2.3.2 Simulating Email Processing Context

The following sections highlight four different ways in which *What.Hack* simulates the real-world context of email processing: 1) workflows, 2) time pressure, 3) interactions with IT support and 4) harmful effects of phishing.

Simulating email processing workflows

A major goal in our game design is to simulate the real-life trade-offs in processing incoming emails. Although deleting every incoming email is a solid defense against phishing attacks, this is implausible in many situations because legitimate important emails will be missed. Therefore, in *What.Hack*, the player must consider each email and decide whether to accept or reject it. The player must make this decision without performing a dangerous action (such as opening an email attachment that could potentially contain a virus).

The player interacts with a simulated operating system, as shown in Figure 2.1. From this screen, the player can launch other application windows. The application that plays the biggest role in the game is the email client, in which the player can view a stream of emails that have been sent to the player. Some of these emails are related to legitimate business interests of the bank, and others are not. Some emails are malicious phishing attempts.

The player indicates their intention to accept or reject an email by clicking on the buttons in the lower right of the screen. If the email is business related and safe then the player should select the "Allow Entry" button (highlight (2) in Figure 2.1). Otherwise, the player should select the "Submit Phishing" button (highlight (3) in Figure 2.1) to "shred" the email. If the player "shreds" too many legitimate emails, the player's manager will show up and inform the player that they underperformed because their customers complained that the player is unresponsive. This mechanism mimics the real-life decision process in which every employee has to decide whether to trust an incoming email and figure out the appropriate response.

Simulating time pressure

A qualitative investigation by Conway et al. [48] indicated that employees are more vulnerable to phishing links and attachments when they are swamped by a large amount of work. Due to time pressure, they need to scan incoming emails and react quickly if the email is relevant to their work so that they can finish tasks in time. Therefore, it is critical to simulate the time pressure that phishing victims often experience. This is implemented through a time limit, indicated by highlight (6) in Figure 2.1. The game is organized into five "shifts," or levels, which require the player to process an average of six emails in eighty seconds. Therefore, if the player takes too long to consider each email, the player will not pass the shift. This also incentivizes the player to work fast - the player will obtain a higher score if the player can accurately process more emails than required.

Simulating interactions with Information Technology support

Since phishing remains such a prevalent problem, many IT departments have developed mechanisms for staff to report phishing emails and they encourage people to do so. This can sometimes help prevent other people from falling victim to the same email. Since this is a key component of many people's email processing workflow, we simulated it in our game. The button for *rejecting* an email is called "Submit Phishing" (highlight (3) in Figure 2.1).

Furthermore, many IT departments will provide consultation on whether or not an email is safe. However, this takes time, creating a tradeoff for staff members who experience pressure to process many emails quickly. Therefore, we

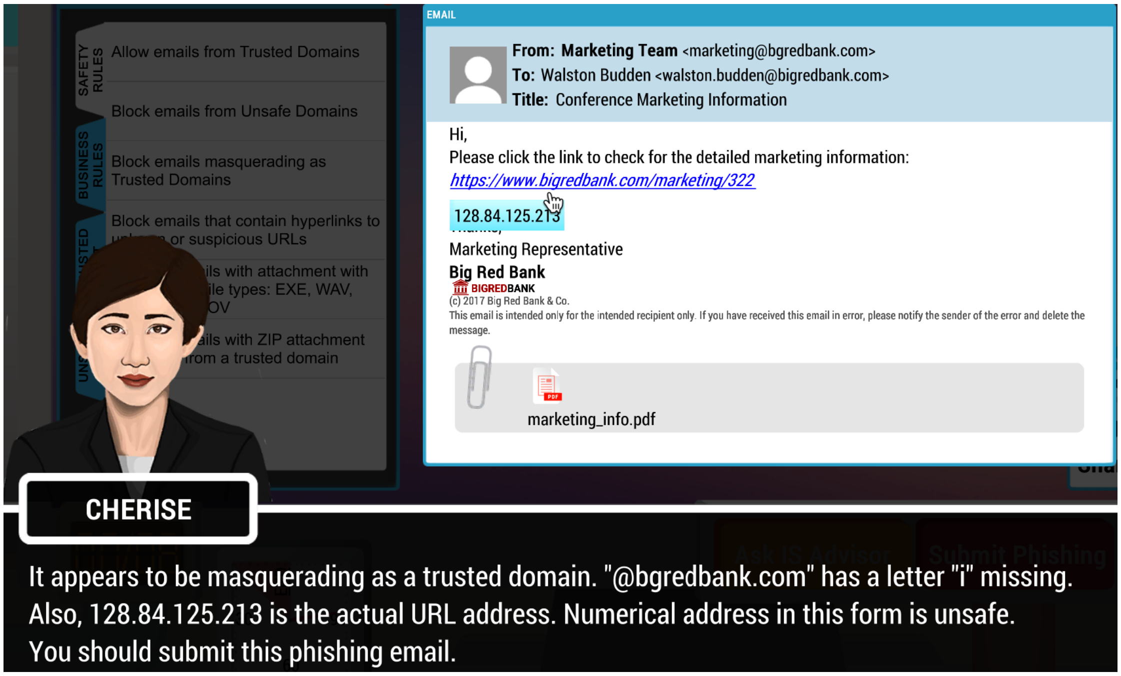

Figure 2.2: After clicking on the "Ask IS Advisor" button, Cherise, the information security advisor, appears and says: "It appears to be masquerading as a trusted domain. "@bgredbank.com" has a letter "i" missing. Also, 128.84.125.213 is the actual URL address. Numerical address in this form is unsafe. You should submit this phishing email.". The corresponding email is highlighted.

implemented this tradeoff through the addition of a button that lets the player refer the email to the Information Security (IS) advisor for help if they are uncertain whether email is safe. After players click on the "Ask IS Advisor" button (highlight (4) in Figure 2.1), the cybersecurity advisor will appear and tell the player whether this email is safe or unsafe (shown as in Figure 2.2). If the email is unsafe, she will also explain the reason. A small amount of limited game time (2-4 seconds) are deducted to simulate the cost of communication. This mechanism encourages player to think actively whether this email is a phishing email before asking for help.

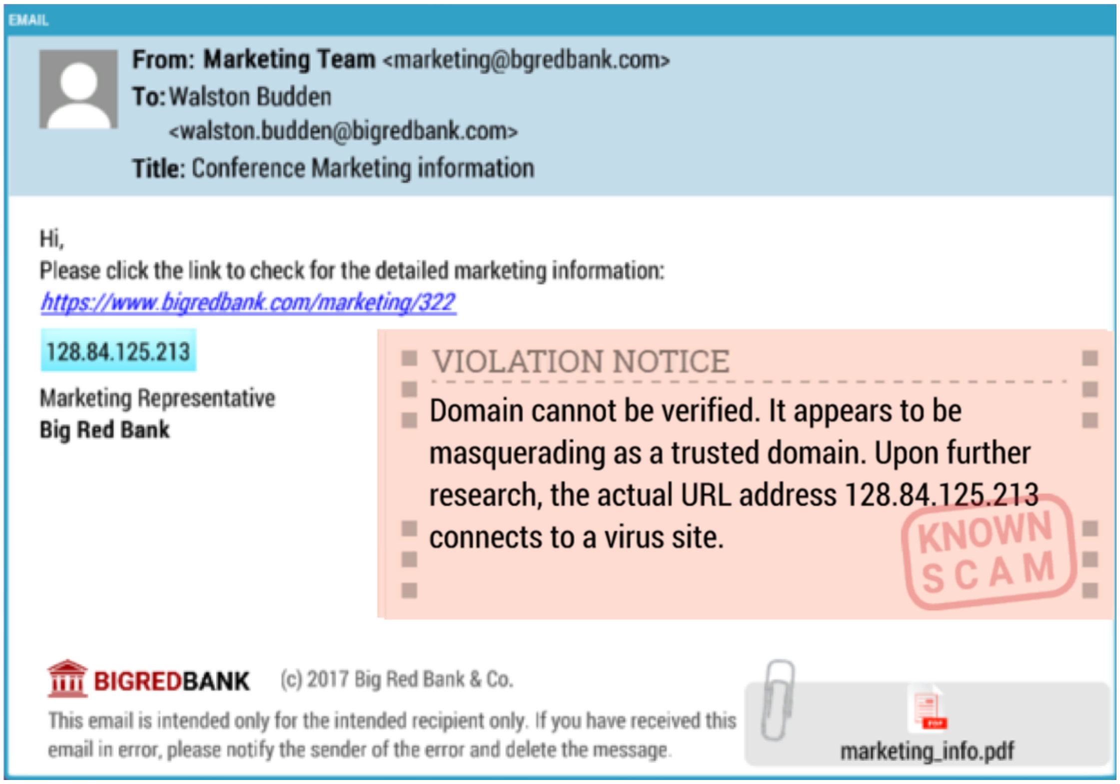

Figure 2.3: *What.Hack* provides immediate feedback when the player makes a decision. Here, the player clicked "Allow Entry" for an email that contained a malicious link. The player has now received a violation explaining the mistake. Violations affect the player's overall progress and score and acquiring too many violations in a shift will result in the player being required to repeat that shift.

Simulating Harmful Effects of Phishing

A big challenge in cybersecurity training is teaching learners to appreciate the risks of a bad decision. Existing training materials often state these risks at a high level but do not really *show* the learner what will happen if something goes wrong. Therefore, *What.Hack* simulates how the player's decisions lead to various outcomes, both positive and negative. From the standpoint of maximizing engagement, a key consideration is how to give the player sufficient *freedom* to make meaningful decisions that will have both short-term and long-term consequences on the game, while still ensuring that the players understand what they do is right or wrong.

If the player makes a wrong decision, a violation note will appear, as shown in Figure 2.3. If the player labels an unsafe phishing email as safe, the note will state the specific rules that the email violates. Similarly, if the player thinks a safe email is a phishing email then the note will remind the player that this email is safe. This approach helps players reflect on unfamiliar anti-phishing knowledge, which helps retain knowledge [51]. If the player fails too much, the player's bank loses trust and the player gets fired.

2.3.3 Structuring the learning content

Our goal for the level design of this first *What.Hack* prototype was to measurably improve the player's ability to recognize potentially malicious emails within a short amount of play time. Therefore, we focused on three popular phishing attack templates [159]: 1) similar domain attack, 2) URL manipulation, and 3) malicious attachment.

In order to maximize engagement, we constructed the shifts so that they combine concepts in order to form a progression that starts easy and grows more difficult. The progression design motivated by the theories of flow [52], elaboration [152], and the Zone of Proximal Development [189]. This contrasts with existing anti-phishing training games that typically focus on only one concept at a time, and only combine concepts on the last level [165, 22].

The progression generally increases one or two attack templates per shift (level), and continually combines these attack templates with other templates introduced in previous shifts. The similar domain attack is first introduced at Shift 1. The URL manipulation attack happens at Shift 3; and the malicious attachment attack appears at Shift 4.

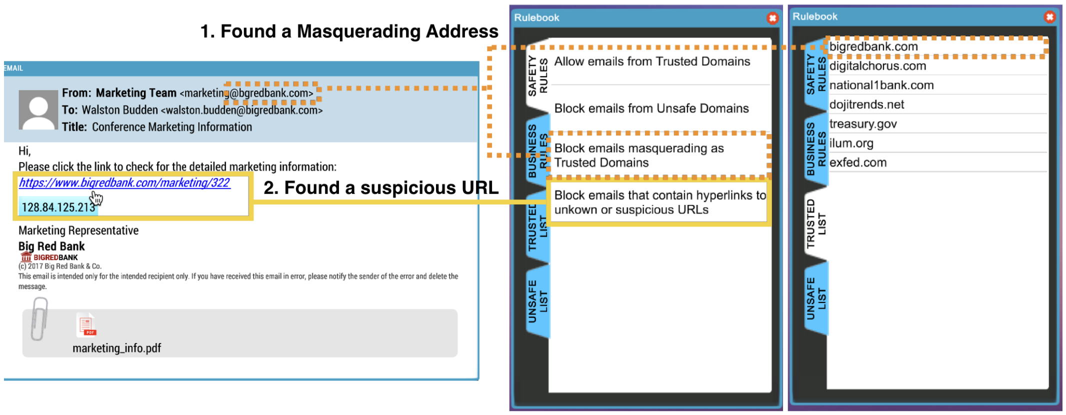

Figure 2.4: Players apply the rules from the rulebook to the email to determine if an email is a phishing attempt or not.

Balancing Freedom and Learning through the Rulebook

Traditional email embedded training often relies on absolute rules like “never click on links”, which are impractical in real life. To improve, the rules taught by our game should be meaningful rather than purely prescriptive. Furthermore, we designed the game so that players can perform actions at will but also learn which actions lead to negative consequences. A key innovation of *Papers, Please* is that the target skills (related to border control document verification) are communicated to the player through a *rulebook*, to which rules are gradually added over the game. In *What.Hack*, we explored whether the same rulebook mechanism could be used to deliver learning content related to phishing defense. The rulebook is shown in highlight (5) of Figure 2.1. It can be viewed at will, thereby serving as reference material for the learning progression.

Figure 2.4 shows an example of how a player might use the rulebook to determine whether an incoming email is malicious. Table 2.2 shows how the rules are introduced in the learning progression.

Shift 1	Allow emails from Trusted domains Block emails from Unsafe domains Block emails masquerading as Trusted domains
Shift 2	Block any email with inappropriate content Block non-business related emails from unknown domains Allow business-related emails from unknown domains
Shift 3	Block emails that contain hyperlinks to unknown or suspicious URLs
Shift 4	Block all emails with attachment with the following file types: EXE, COM and HTML Block all emails with ZIP attachment that is not from a trusted domain
Shift 5	Block all emails with DOC attachment that is not from a trusted domain but DOCX attachment is acceptable Ask IS advisor about emails from unknown domains that has a business-related attachment

Table 2.2: The rules that are introduced in each shift. The ruleset for each shift gradually grows in complexity, requiring players to make more realistic decisions in later shifts.

2.4 Performance Evaluation

Attack Templates	URL Only	URL + Domain	Attachment Only	Attachment + Domain	Total
# Phishing Emails	2	4	3	2	11

Table 2.3: Attack Templates in the Selected Phishing Emails for the User Study

Existing anti-phishing games that do not simulate phishing attempts through role-playing have shown that they were more engaging and thus provided better learning outcomes than non-gamified training [165, 22]. However, as pointed out by the criticisms of gamified education [135], not all gamified approaches are equally effective. To the best of our knowledge, there is a real lack of user study comparing two different approaches to the same gamified problem. Therefore, we ran an in-lab study to compare our game with a representative non-phishing role-playing game in addition to a non-gamified training.

2.4.1 Study Design

In order to evaluate the effectiveness of *What.Hack*, we conducted user studies that measured the impact of the game on players' ability to recognize phishing attacks on a pretest and a posttest. In the process, we compared the game to two anti-phishing training approaches.

We compared our game with *Anti-Phishing Phil* [165], a competing non-phishing role-playing game that educates players to identify similar domain and URL manipulation attacks, which has been cited for over 300 times. Its latest version is also provided for cybersecurity training in Carnegie Mellon [14]. Its gameplay covers the key elements that are addressed in the most popular game design framework for anti-phishing training [21]. The recent new designs [127, 22] of anti-phishing games followed the fashion of *Anti-Phishing Phil* gameplay. These evidences show that it is a representative anti-phishing game with which we decided to compare.

In addition to *Anti-Phishing Phil*, we also compared with *PhishLine* training materials [16] that are currently used by Cornell. *PhishLine* training materials are the typical form of fact-and-advice training that have been widely used and studied [107, 193, 105]. *PhishLine* and *What.Hack* cover three popular phishing attack templates: 1) similar domain attack, 2) URL link manipulation and 3) malicious attachment. *Anti-Phishing Phil* only covers the first two attack templates.

What.Hack targets young professionals and adults. We recruited target players through fliers, face-to-face interactions and Cornell's experiment sign-up system. Participants who signed up using the university's system received 2 experiment credits upon completion. We required the participants to be at least 18 years old and that they had never taken a cybersecurity class or participated in anti-phishing training. We recruited 39 students at Cornell and randomly assigned 13 people to each group.

The user studies consisted of an effectiveness evaluation session and an engagement evaluation session.

In the first evaluation session, which targeted effectiveness, participants were given a pretest in which they identified whether an example emails were phishing or legitimate. After the participants completed the pretest, we randomly assigned them to play *What.Hack* or play *Anti-Phishing Phil* or study *PhishLine* training materials with equal probability. All participants were able to finish their assigned game within half an hour. Participants were then given a posttest using the same emails in the pretest. The posttest also asked players to describe the knowledge that the participants thought they learned from the game that they played.

In the second evaluation session, which targeted engagement, participants played the game that they did not play in the first session. After finishing the game, participants were asked to complete an exit survey about engagement.

2.4.2 Test Design

The goal was to confirm whether *What.Hack* can improve the correctness of identifying phishing emails with a statistically significant result. Therefore, we presented participants with emails in the pretest and the posttest. We asked them to decide whether an email is phishing or legitimate. We ask participants to choose whether it is phishing or legitimate. In addition, they need to rate how confident they are about the answer using the 5-point Likert scale from 1 (random guess) to 5 (very confident).

We selected and fine-tuned 11 real phishing emails from a database maintained by Cornell's IT security office, which is a similar approach that [193] took to select the first 12 phishing emails for their user study. The types of messages include suspicious account activity warning, financial document review, university president announcement, shipment notification, etc. Table 2.3 shows the number of phishing emails that use certain types of attack templates. These phishing emails can be seen in our online appendix: `osf.io/pven9/`.

In addition to the 11 phishing emails, we also chose 9 authentic emails from the communications database verified by Cornell's IT security office. They are not publicly available due to the privacy policy.

We examined the conceptual knowledge and procedural knowledge that participants retained. After they completed the quantitative test in the pretest and in the posttest, participants were asked to answer the question: "What is the strategy you used to process these emails? Please write in bullet points." This question allowed us to better understand how they make the decision before and after playing the game.

In addition to participants' strategy for identifying phishing emails, we asked them the following question at the end of the posttest to identify what else they had learned from the game:

"Did you learn any new concepts or skills from this training that will help you prevent yourself from being hacked by unsafe or phishing emails? Please write in bullet points."

To evaluate engagement, we asked participants the following 5-point Likert scale rating questions in the exit survey:

"On a scale from 1 (very boring) to 5 (very engaging), how would you rate the engagement of each training?"

"On a scale from 1 (strongly disagree) to 5 (strongly agree), how likely are you to recommend this training if your friends want to learn how to defend attacks from phishing emails? "

2.4.3 Results

We measured the effectiveness of our game by examining the correctness percentage, the false negative rate and false positive rate before and after the game. A false positive is when a legitimate email is incorrectly regarded as a phishing email. A false negative is when a phishing email is incorrectly regarded as a legitimate email. In our case, the false negative rate is more important because it would expose people to the danger of phishing if they mistrusted the phishing email.

Evidence that the game improves correctness. We derive a measure for the

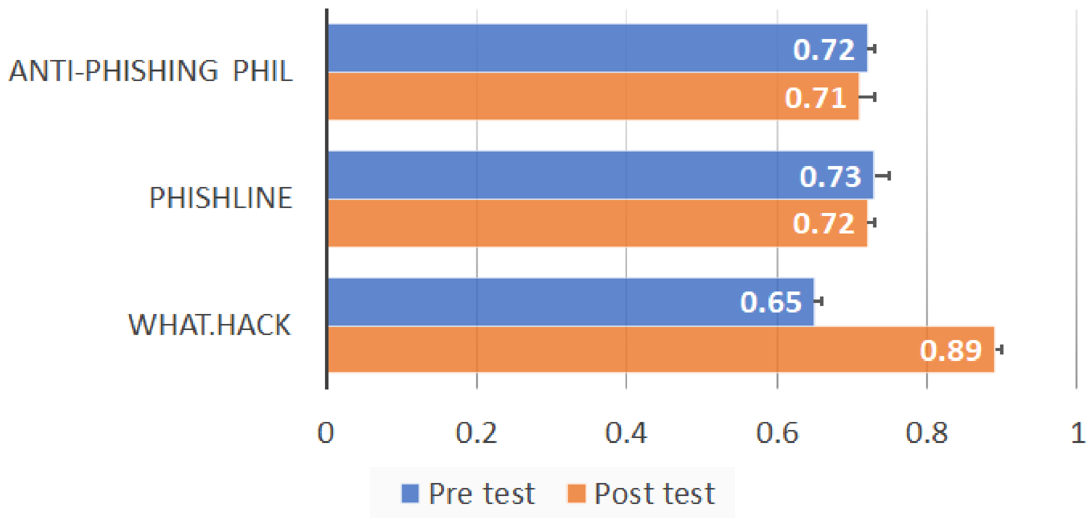

Figure 2.5: The correctness percentage. Our game improved players' correctness in identifying phishing emails by 36.7% **(statistically significant), whereas neither of the control groups achieved a statistically significant improvement.**

change of correctness percentage between the pretest and the posttest using one-way ANOVA. There was a significant effect of training for the three conditions (F(2, 36) = 18.53, $p < .01$) with a large effect size ($\eta^2 = .507$).

Post hoc comparisons using the Tukey HSD test indicated that the score increase for *What.Hack* (M = .239, SD = .106) was significantly different than *Anti Phishing Phil* (M = -.004, SD = .107, $p < .01$) and different than *PhishLine* (M = -.007, SD = .138, $p < .01$). However, *Anti Phishing Phil* condition did not significantly differ from *PhishLine* condition ($p = .90$).

Evidence that the game enhances anti-phishing self-efficacy. The result shows that participants became more confident about their judgments after playing *What.Hack*. The average confidence rating increased from 3.33 (variance = .368) to 4.08 (variance = .329). The distributions between the two sets differed significantly (Mann-Whitney U = 32.5, Z = -2.64, $n_{pre} = n_{post} = 13$, p = .0083, two-tailed). While participants made a wrong decision after playing *What.Hack*, their confidence rates did not significantly change. This evidence meets our overall educational goal that players should be more confident of identifying phishing emails correctly while not making hasty decisions. We did not observe the control groups enhancing participants' confidence in a statistically significant way: the average confidence rating of *Anti-Phishing Phil* was 3.38 pretest (variance = .332) and 3.51 posttest (variance = .517); the average confidence rating of *PishLine* was 3.51 pretest (variance = .307) and 3.58 posttest (variance = .415).

Evidence that the phishing simulation game facilitates learning transfer. To compare each group of participants' ability to transfer the knowledge they have learned from the game or materials to identify whether an email contains malicious content, we removed 3 email examples that requires knowledge of attachment file types to make informative decisions because *Anti-Phishing Phil* does not teach how to identify malicious attachments. We run the analysis on the rest 17 emails that can be determined by validating the domain address and/or the URL hyperlink. The training games/materials still makes a statistically significant impact on the change of correctness percentage between the pretest and the posttest for the three conditions (F(2, 36) = 13.34, $p < .01$) with a large effect size ($\eta^2 = .426$) using one-way ANOVA.

Post hoc comparisons using the Tukey HSD test indicated that the score increase for *What.Hack* (M = .199, SD = .11) was still significantly different than *Anti-Phishing Phil* (M = -.013, SD = .12, $p < .01$) and different than *PhishLine* (M = -.02, SD = 0.14, $p < .01$). However, *Anti-Phishing Phil* did not significantly differ from *PhishLine* ($p = 0.90$).

Evidence that the game is engaging. There was a statistically significant difference between the engagement ratings of three training games/materials

(Kruskal-Wallis H(2) = 44.121, $p < .01$), which a median rank of 4 for *What.Hack*, 3 for *Anti-Phishing Phil* and 3 for *PhishLine* (a score of 4 is positive, 3 is neutral). The median rank of rating of recommending three games/materials to friends who want to learn about defending phishing email were *What.Hack*: 4, *Anti-Phishing Phil*: 3, *PhishLine*: 3; the distributions in the three groups differed significantly (Kruskal-Wallis H(2) = 28.75, $p < .01$). Figure 2.6 and 2.7 are the bar charts of the engagement ratings and the recommendation ratings respectively.

Evidence that players learned new anti-phishing skills from What.Hack. We find evidence from the posttest questionnaire that some participants learned to defend the classic URL semantic phishing attacks after playing *What.Hack*: *'you need to hover your mouse to check the actual link'*; *'the actual link will appear after hover my mouse on it'*. And some participants learned to not rely solely on the sender's identity or the content of the email to identify phishing emails: *'do not download exe even if it's forwarded by Cornell ppl'*; *'the content is not most reliable way to identify phishing emails'*.

2.5 Discussion

We believe that *What.Hack* is the first anti-phishing game that simulates phishing attempts through role-playing. Our evaluation indicated that *What.Hack* is effective and engaging, when compared to *Anti-Phishing Phil* and *PhishLine*. The results showed strong evidence that leveraging situated learning theory can also enhance anti-phishing training, in addition to domains such as science and ethics education [28, 100, 162].

Design implications. The user study participants thought that *Anti-Phishing*

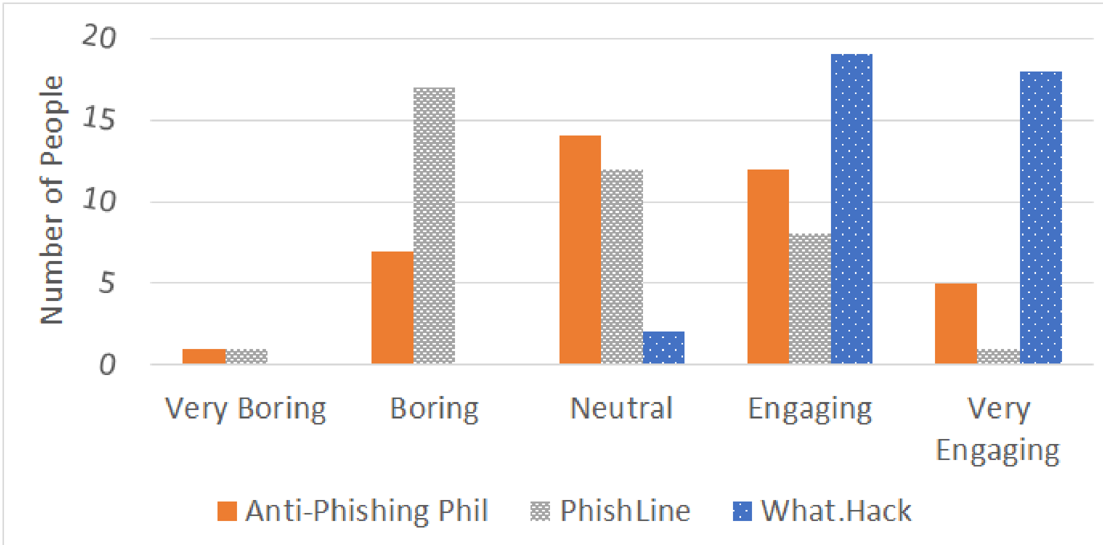

Figure 2.6: The engagement ratings. 95% of the participants find *What.Hack* being engaging or very engaging. The number is 44% for *Anti-Phishing Phil*, and 23% for *PhishLine*.

Phil was more engaging but less recommended than watching videos. The qualitative measures showed that participants retained more conceptual knowledge by playing a game than watching videos. Unlike *What.Hack*, however, *Anti-Phishing Phil* players could not practice using the conceptual knowledge to vet phishing contents. While *What.Hack* could enhance anti-phishing self-efficacy in the user study, *Anti-Phishing Phil* failed to provide the same result. This feeling of "I know it, but I am unsure how to use it" could be the reason for the disparity in ratings for *Anti-Phishing Phil*. *What.Hack* avoided this issue, which implies that gamified cybersecurity education should teach both conceptual knowledge and how to use it.

Another implication is that gamified cybersecurity education should do the best to simulate all factors that affect learning outcomes to avoid unwanted side-effects. In our case, the cognitive load impacts susceptibility to phishing [202]. *What.Hack* prepares players for a heavy cognitive load that they would experience in handling real threats, but *Anti-Phishing Phil* does not. The *Anti-Phishing Phil* paper [165] reports that some players did not look further to avoid phishing attacks if the URL text did not raise suspicion, which did not happen to *What.Hack* players in the user study.

Limitations and future work. *What.Hack* did not fully utilize social engagement in situated learning like other role-playing games did for language learning [54] and science education [100]. Phishing, especially social engineering attacks, is a crime of deceiving people through online social interactions. Therefore, a multi-player extension of *What.Hack* could potentially enhance anti-phishing training.

What.Hack has a broader design goal than *Anti-Phishing Phil* and *PhishLine*, which may introduce unfairness when comparing our game with these two conditions. We integrate conceptual knowledge into an email processing game and let the player internalize what kinds of URLs are dangerous, whereas the other two conditions only focus on conceptual knowledge, such as URL mechanics. However, we believe the most important consideration is the ability of anti-phishing training to prepare people to handle real email threats, and we found that our game is more effective in this regard.

Speaking of the tests result, we noticed that the pre-test performance differs across three conditions. However, the score difference is not large enough to claim that the *What.Hack* group on average made one more mistake than the other two groups. Therefore, we are still confident that three groups share the same anti-phishing skill level. That being said, the pre-test performance indicates that the *What.Hack* group may be able to improve more than the other two groups after the training, which potentially affects the study's effectiveness.

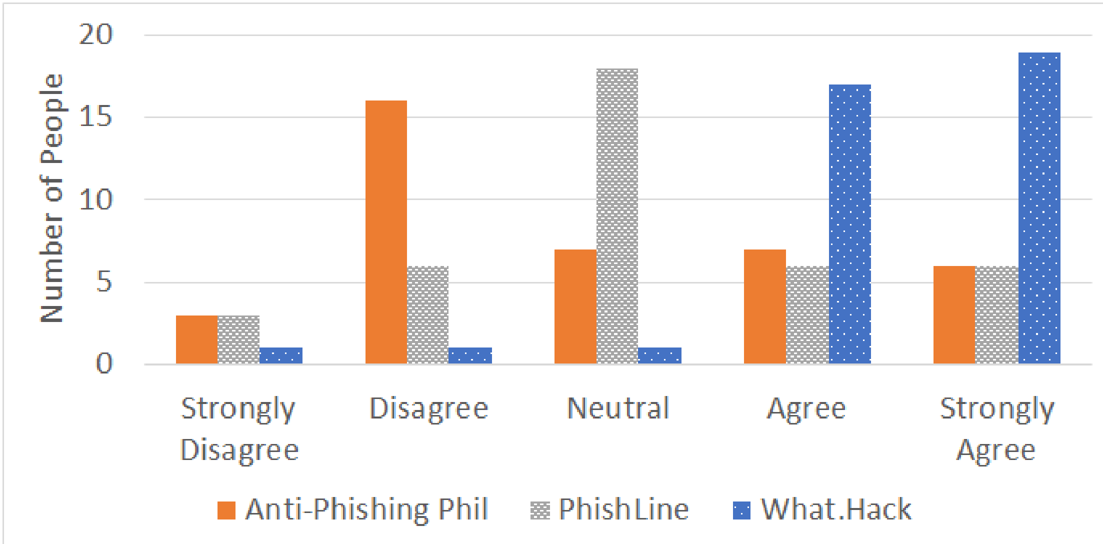

Figure 2.7: The recommendation ratings. 92% of the participants agree or strongly agree that they will recommend *What.Hack*. The number is 33% for *Anti-Phishing Phil*, and 33% for *PhishLine*.

To further determine the impact of our game, especially on long-term retention of phishing attack methods and defenses, we would like to conduct a field study in the future.

What.Hack offers a good starting point for the development of other similar game-based experiences in the field of cybersecurity education. For example, fake news is a trending global problem and new game designs for training people to identify fake news are rare and ad hoc [11]. Our game design can be re-skinned to introduce this issue. We would like to explore in the future whether such training is also more effective by using our game design.

2.6 Conclusions

The main goals of *What.Hack* are to 1) teach players anti-phishing techniques by simulating some of the circumstances in which people routinely defend against phishing attacks, 2) engage players by offering them freedom to experiment and observe narrative consequences, and 3) deliver a set of learning content through a task progression that starts easy and gradually grows more complicated. We presented results from a lab study demonstrating that *What.Hack* was able to improve players' correctness in identifying incoming threats by 36.7%, whereas a control group that played a different game did not achieve a statistically significant improvement. The results indicated that situated learning plays an important role in improving learning outcomes by engaging learners in a relatable simulation world.

CHAPTER 3

COLLECTIAR: MAKING VOCABULARY LEARNING FUN THROUGH A COMPUTER VISION-BASED AUGMENTED REALITY GAME

3.1 Introduction

Vocabulary acquisition is crucial in language learning. However, learning and retaining vocabulary is a formidable challenge. Long-term retention of vocabulary typically involves constant repetitive retrieval practices [194]. How to help learners develop a growing love for vocabulary learning has been a long-standing research topic. Although virtual flashcard applications, such as Anki [1] and Quizlet [7], are widely perceived as convenient and effective tools for vocabulary learning, studies [111, 29, 209] found that students are not engaged and there is still vast room for improvement. To help people learn vocabulary in a more productive way, recent work [181, 68, 182, 37, 158, 92, 185, 53] has focused on context-based learning. Context is valuable to language learning because associating a word's meaning with learners' context helps them recall the word [34, 183].

Previous designs for vocabulary learning use contextual information, such as websites [181], GPS locations [68] and chat history [37], to pick relevant foreign words for rote memorization. The user studies of these projects have shown collective evidence that learning assisted by contextual information provides better outcomes. However, these projects did not completely solve the problem of long-term engagement, which is critical to vocabulary learning because it requires persistent effort.

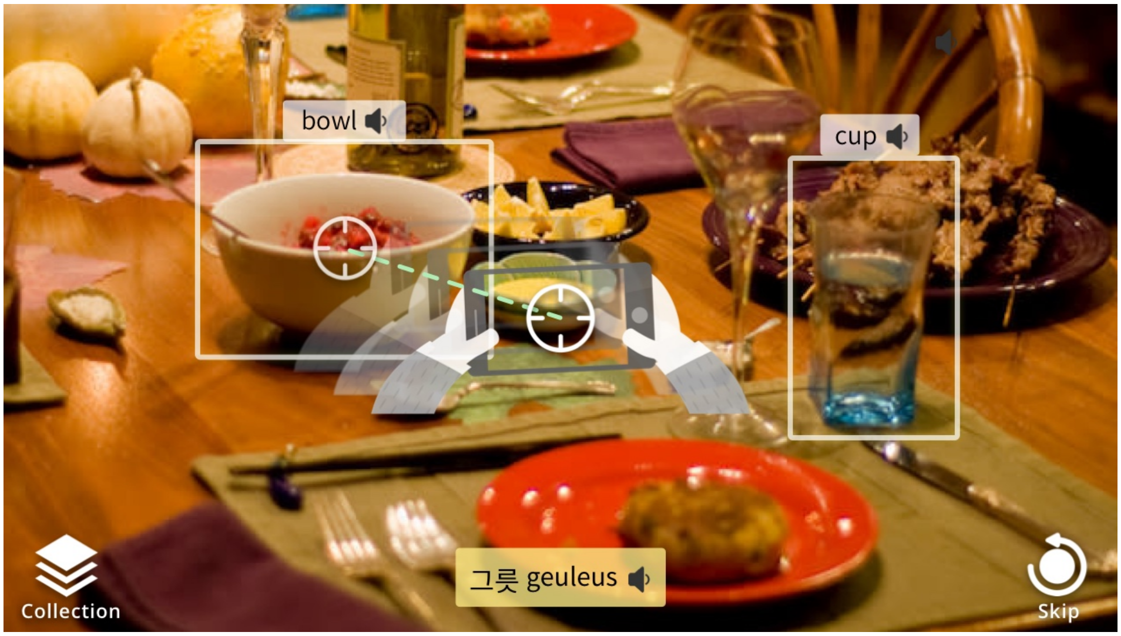

Figure 3.1: A screenshot of onboard experience of *CollectiAR* when the player is trying to collect 그릇 *geulues*, which means bowl in Korean.

With advances in augmented reality (AR) technology, recent projects [158, 92, 185] attempt to further engage learners by overlaying virtual information on the learner's view. These projects share a similar design approach. They display the form, the pronunciation, and the meaning of a foreign word in the vicinity of its corresponding real-world object. They aim to engage learners by connecting new materials to the learner's spatial context, which provides a more immersive experience in which people can naturally infer the meaning of words by analyzing their surroundings. A user study [71] also finds that participants have more accurate recall of new information if it is visually associated with specific locations.

The study results of these AR vocabulary learning tools align with the hypotheses that learners are more engaged compared with flashcards and they are able to build a stronger memory as opposed to the approach that has no direct connection to the player's physical environment. That being said, previous work leveraged physical context for learner engagement without focusing on playful interaction design, which may be inadequate in order to motivate people to explore and learn a second language for long periods of time.

In our view, teaching language in context creates two demands. First, learning applications should immerse the learner to the greatest extent possible in order to utilize the learner's real-world context. However, this may not be enough to sustain motivation by itself. Therefore, a second demand is how to design incentive structures or other game-like elements [61] to engage learners for a prolonged period. With an aim of motivating learners to stay and play, we would like to simulate an immersive experience similar to what people experience in the real world as they actively analyze their surroundings and look for visual cues that relate to the foreign word.

We drew inspiration from existing video games that teach a second language in a virtual environment. Influent [89] and Crystallize [53] have the player explore a virtual world to collect new words that are associated with the objects in that virtual world. The idea of sending players off on a scavenger hunt to discover vocabulary words has been an exciting way to engage students in learning [55]. In order to implement this game in the real world, we need to solve two critical challenges. First, the game should be able to recognize objects in the environment. Second, the game should utilize its knowledge of the environment to engage the player.

With this in mind, we present *CollectiAR*. We explore how AR can be used together with computer vision in order to identify objects located in the player's immediate vicinity and to engage them in finding these objects by generating clues for scavenger hunt in a foreign language. Computer vision alone does not solve our first challenge. Given that computer vision is inaccurate, we need to ensure that recognition accuracy is as high as possible and that the game is playable even when objects are incorrectly recognized. The second challenge is to alert the player the presence of detectable objects in their environment and to design a scaffolded learning process that associates physical objects with corresponding foreign words. We discuss our solutions to these challenges, which will help AR game designers because these two challenges are specific not only to our game but also occur generally in any AR game design that aims to engage players with context-sensitive content.

To evaluate the impact of our game, we conducted an in-lab user study with 42 players. We compared *CollectiAR* to a popular flashcard-based learning approach, *Quizlet*. The key findings show that: 1) *CollectiAR* is more engaging than *Quizlet*; 2) participants perform better on a vocabulary quiz after one week if they learn these words by playing *CollectiAR* instead of *Quizlet*. The result indicates a strong potential future for computer vision-based AR games in learning.

The core contributions of this paper are: 1) a computer vision-based AR game design that explores how language learning games can engender long-term engagement, 2) a general solution to reduce the impact of computer vision inaccuracy in an AR learning game, and 3) a user study suggesting that a game that leverages physical context is more engaging and may improve long-term retention of vocabulary.

3.2 Related Work

3.2.1 Contextual Learning

Learning theories have emphasized the positive impacts of using real-world context in education [173, 149]. Particularly in individuals' language learning processes [75], real-world experience plays a more important role compared to abstract prepositional reasoning. Situated cognition theorists also highlight the presence of context in content learning. Likewise, empirical evidences suggest that the recall rate is the highest if the context and the use are perceptually similar [183].

Using language with reference to specific instances and objects, namely grounding, is important for all parties in communication. There are four levels of semantic grounding, that are attending, identifying, understanding, and considering, identified in a game in which a participant is instructed to assemble LEGO blocks [46]. A common language with a shared reference system allows the instructor and the participant to ground their communication to the identical objects.

Mainstream popular language learning tools do not have the ability to leverage physical context yet. For example, DuoLingo [188] and Rosetta Stone [5] do not currently allow learners to use a smartphone camera to learn language using objects in the vicinity of the smartphone. The academic community has started to notice the importance of context in language learning. For example, MicroMandarin is a smartphone tool that uses the learner's real-world location to suggest relevant vocabulary words [68]. Dearman et al. used a desktop wallpaper to teach vocabulary drawn from interactions with a mobile phone [57]. Trusty et al. augmented the website to teach second-language vocabulary [181].

Although these tools do utilize "the real world'," the use of context is passive and does not involve manipulating the environment. Ideally, learners could learn by *directly interacting* with objects in their environment. We envision learning experiences that allow vocabulary (e.g. "bowl") to be taught in a plausible physical context (e.g. with an actual bowl).

3.2.2 Multimedia Learning

Multimedia learning [123, 124] is an important cognitive theory that suggests guidelines for designing multimedia applications for education that can effectively improve people's learning outcomes and reduce people's cognitive load in the meanwhile. We believe that multimedia learning theory should guide our design because AR is a type of multimedia technology used for education [115].

Multimedia learning makes three assumptions [123, 124]. First, humans can process auditory and visual knowledge separately at the same time. Second, humans can only process a limited amount of knowledge at a time. Third, humans learn by actively engaging in processing this knowledge. Based on these assumptions, multimedia learning identifies the following cognitive processes that take place in learning: (1) selecting relevant words and images from the presented information concurrently, (2) organizing the selected words and images into a coherent representation in separate channels, and (3) integrating the visual and verbal representations and connecting them to prior knowledge.

Multimedia learning has served as the theoretical basis for many studies in technology for language learning. Studies [45, 146, 207, 118, 117] on the effectiveness of vocabulary learning and retention have shown that annotating words with pronunciation audio and illustration images helps students learn vocabulary better than either one alone. The majority of flashcard software [201] also supports adding audio and images to the flashcard to enhance learning. Websites like FluentU [3] make use of videos with subtitles to teach language.

In addition to the traditional forms of multimedia, recent studies on optical see-through AR displays [71, 192] observe that participants who are assisted by AR are able to utilize the spatial factor to improve their learning process. The spatial information refers to specific locations or the physical features of the environment. This observation provides additional motivation to use AR to design a game that engages people learning vocabulary.

3.2.3 Augmented Reality for Language Learning

The affordances of AR in learning have attracted a growing number of researchers to introduce AR into pedagogical scenarios, especially language learning. Language learning applications that are based on the GPS location sensors [163, 143, 130] or embedded sensors [160, 79] claim to have used AR technology. They collect the virtual real-world information nearby the user to suggest relevant learning materials. However, these applications do not directly manipulate the physical environment perceived by the user.

Ideally, users could directly interact with physical objects in the AR learning environment, which is a demanded feature suggested by several literature reviews of AR in education [208, 109]. This feature can integrate learners' visual-spatial processing into the learning process. It would provide more immersive experience than sensor-based AR because people can naturally infer the meaning of words by analyzing their surroundings.

Recently, new AR designs started to address this need by projecting the form, the pronunciation and the meaning of a foreign word nearby its corresponding real-world object [158, 92, 185]. Two projects [158, 92] evaluated their AR designs against basic digital flashcards. They found positive results that participants performed significantly better on the delayed vocabulary recall test when they learned through optical see-through AR displays.

However, it is arguable whether students are able to focus on internalizing the vocabulary on their own for a long period of time without any active interactions with the augmented environment. If the student becomes distracted and bored then the benefits of contextual learning and multimedia learning brought by AR would become ineffective.

A set of AR games for language learning have been developed to better engage students by using incentive structures or other game-like elements. Electric Agents [153] teaches vocabulary by interacting with a TV show through AR and mobile device sensors. Hsu [90] designed an AR English spelling game that is related to physical objects. The AR Ole Cierraojos [179] is a story-based AR game for practicing reading comprehension that pops up from a detectable figure in the book. The evaluations of the games show collective evidences that the games are engaging to the target learners and significantly enhance the learning outcomes.

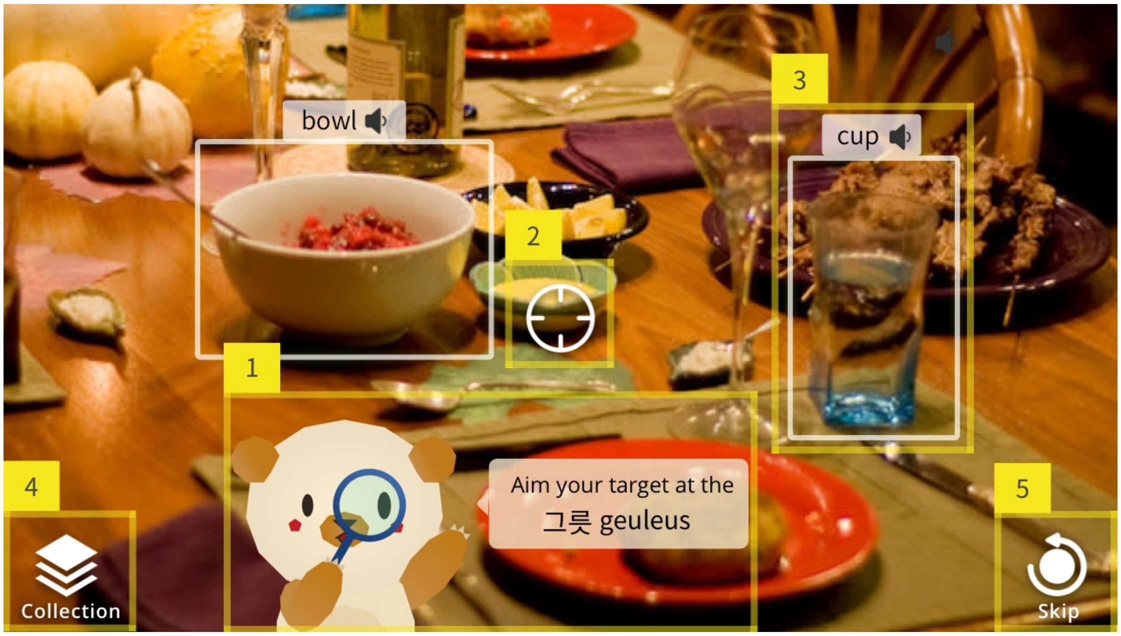

Figure 3.2: In *CollectiAR*, players solve a scavenger hunt puzzle in their immediate vicinity where they can explore and collect the highlighted objects. The spying bear tells players to find the target object that is described in a foreign language (1). Players can move their handheld device to focus on an object (2). The detectable object is highlighted by a box bounding around it and has its label and pronunciation in foreign language attached to it (3). Players can review the collections of scavenger hunt puzzles in the form of flashcards (4). Players can skip and report the detection error if they could not find such an object in the scene (5).

The limitation on these AR games is that they rely on marker-based tracking or position tracking that may not always work well in outdoor environments or rooms that do not allow tagging. This approach also requires initial setup and maintenance effort that makes it hard to deploy.

Our goal is to combine the advantages of both incentive structures and optical see-through AR so that the learner can be engaged in playful interactions and can play the game anywhere at anytime.

3.3 Vocabulary Scavenger Hunts

The primary goal of *CollectiAR* is to engage players in vocabulary learning by providing an immersive learning experience that involves more than repetitive memorization. Ideally, it should maximize the sense of immersive learning by taking the advantage of presence. Presence has been explored by the VR game for language learning, Crystalize [42]. Their study found that the players feel more involved in Japanese culture while playing the game.

CollectiAR is inspired by the design of scavenger hunt that uses the player's presence as the gaming environment. If we are able to associate the meaning of target vocabulary with players' presence, then we can use target vocabulary to generate scavenger hunts that motivate learners to explore and to learn target vocabulary in their immediate surroundings.

To implement *CollectiAR* that makes language vocabulary learning fun, we need to: (1) generate clues for scavenger hunt items in the real environment, (2) provide immediate feedback in the game, (3) implement collection system for long-term engagement, and (4) mitigate the negative impact of computer vision inaccuracy from the interfaces design perspective.

3.3.1 Clues Generator for Scavenger Hunts

Our primary objective is to make language vocabulary acquisition fun. Bartle [30] proposed several player types that he observed. One of these player types is the explorer who finds great joy in exploring the world and discovering unknown objects or hidden information. The design of promoting curiosity and hiding “treasures” from players is found to be an effective intrinsic motivator [122].

These insights lead us to design a scavenger hunt game in which words are hidden and the learner is motivated to explore their immediate vicinity and figure out the hidden words. To make the game challenging and fun, we need to make a trade-off between the words that the game is going to show and the words that the player is going to find. In order to make a successful trade-off that turns the learning experience into a fun exploring game, we need to embed our game design on the player's real world. The design of player interactions with the game environment should also be really easy to understand which object is referred to.

Our first goal is to understand what is in the environment. Our second goal is to scaffold the process of communicating to the player what objects are interactable and helping the player to find the target object. Accomplishing the first goal requires using computer vision to recognize objects in the environment. Accomplishing the second goal requires using augmented reality to overlay feedback on top of the physical views in the camera.

Before the game starts, *CollectiAR* requires the player to scan their surroundings so that it can use computer vision to recognize a number of objects that are available for scavenger hunts. Upon completing the environmental analysis, *CollectiAR* chooses a target from the list of recognized objects according to the word frequency and the player's historical performance in the game. To tell the player which objects are interactable, as shown in Figure 3.2, *CollectiAR* highlights a bowl and a cup on the dining table in front of the player by using white bounding boxes.

After choosing a target, the *CollectiAR* bear tells the player to find the target by displaying and pronouncing the target word in the foreign language. In the example shown in Figure 3.2, *CollectiAR* chooses the bowl as the target. Hence, the *CollectiAR* bear tells the player to find that bowl by pronouncing it as ***geulues***, which means bowl in Korean. By providing candidates' bounding boxes and the target information in Korean, *CollectiAR* bear asks the player to point at the object that they think ***geulues*** refers to by turning the mobile phone to aim the object's bounding box with the telescopic sight ((highlight (2) in Figure 3.2).

In the example shown in Figure 3.2, learners can play *CollectiAR* in a dining room. In addition to the objects on a dining table, *CollectiAR* should be able to detect and teach objects that the majority of people would typically have. It is highly likely that players can be immediately surrounded by home interior objects every day. People learn a language and communicate in various environments. One of the primary environments is people's house. Therefore, we choose 79 words for the current game prototype that are related to home living, electronics and kitchen appliances, etc so that the *CollectiAR* can utilize the most common physical context in which learners would like to play.

3.3.2 Short-term Feedback Design

To implement an immersing and engaging scavenger hunt game that enhances good vocabulary retention, we must address three challenges: (1) we need to limit the set of objects that the player needs to search through in order to find the target; (2) we need to verify whether the player finds the correct object and communicate the information to the player clearly; (3) we need to solidify the

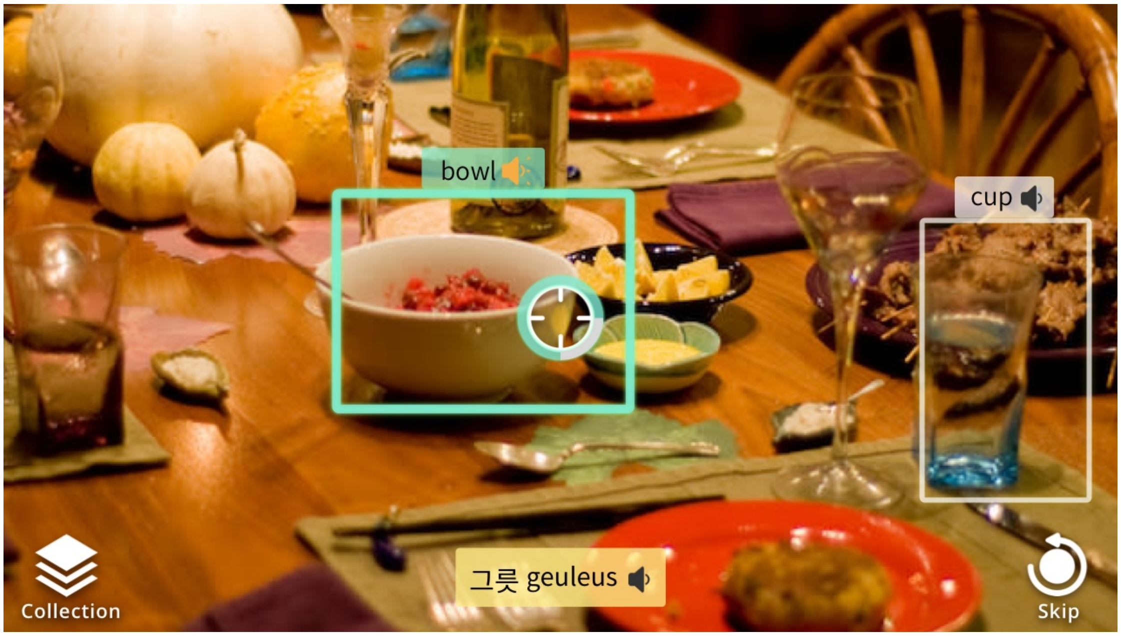

Figure 3.3: ***CollectiAR*** **gives a hint to the player by automatically pronounce the foreign word of the focused object.**

knowledge of foreign words in the player's head. We try to solve these challenges by using augmented reality.

When the player is surrounded by hundreds of objects, it becomes tedious to the player to go through all objects one by one. Instead, one of the design goals that *CollectiAR* must meet is to limit the set of objects that the player needs to search through in order to find the target object. The upper bound of the number of interactable objects is set to three in *CollectiAR*. First, the game filters out detected objects whose computer vision detection confidence levels are lower than 80%. Second, the game sorts the list of candidates according to the player's historical performance. Finally, the game picks the top three objects that the player is least familiar with. For example, an object that the player got it wrong on the first try or an object that the game detects for the first time.

The second challenge is to provide the mechanism to which the player can indicate which object they think is the target then the game can recognize that the player is correct or not and communicate its decision back to the player. A strawman solution is to immediately reveal the result to the player when it detects an object being selected by the player. However, during the iterative design process, we observed that players are uncertain of their decisions. They hoped for more time and hints to help them recall and confirm. Therefore, after the player selects an object, *CollectiAR* immediately changes the color of the bounding box into cyan and pronounces the selected object in the foreign language. After two seconds, *CollectiAR* confirms the result of the object that the player finds.

Figure 3.3 shows a particular scenario where the bowl is selected by the player. The bowl's bounding box color turns into cyan and its pronunciation icon is highlighted which means the player is listening to the pronunciation of bowl in Korean. The cyan arc that grows around the open sight is the progression bar. *CollectiAR* will reveal the result when the cyan arc grows into a circle.

Introducing auditory hint to the short-term feedback design not only serves the purpose of engaging learners but also serves the purpose of improving learning effectiveness. According to previous user studies [118, 117], auditory information plays an important and positive role in long-term memory. The learning effects of the audio reported to be maintained for two weeks with minimal attrition. Because the foreign word pronunciation is the key information that the player uses to correctly identify the target object, we expect that *CollectiAR* can potentially lead to better retention of words.

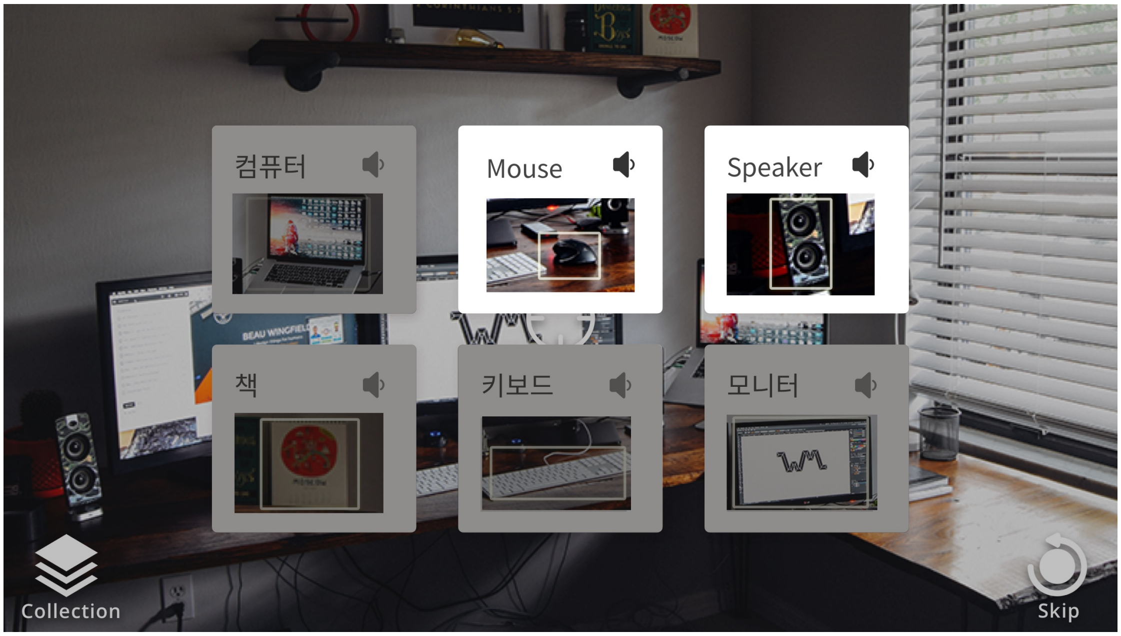

Figure 3.4: ***CollectiAR*** **provides the collection mode to allow the player to review the words they collected from the real world in the flashcard form.**

3.3.3 Long-term Incentive Design

Because one of the game design goals is to keep players playing for an extended number of times, the additional challenge is how to encourage players to return and play *CollectiAR* again. In the Bartle taxonomy of player types [30], he also discusses the achiever who will devote themselves into achieving rewards that confer them little or no gameplay benefit simply for the prestige of having it. Pokemon Go is one of the most successful augmented reality game that attracts this type of players.

Our long-term incentive design drives inspiration from the Pokemon Go collection system. Pokemon Go players don't necessary need to use all of their collected monsters in a gym combat but they are satisfied to see that they are filling up the Pokemon collection and especially when they catch rare Pokemon. Driven by the sense of satisfaction, Pokemon Go players constantly return to the game when they get to a new place to search for new or rare Pokemon. The collection of Pokemon is categorized into different types. For instance, the types include fire, water, grass, etc. Similarly, we categorize 79 foreign words in *CollectiAR* into 9 types: including home living, electronics, kitchen appliances, etc. *CollectiAR* keeps track of the collection status and allows the player to check whether they complete any of the collection types. Figure 3.4 shows the collection mode that allows the player to review the items and new words that they found in their office.

3.3.4 Mitigation of Computer Vision Inaccuracy

It is worth notice that *CollectiAR* replies on the computer vision to make many critical decisions like deciding which foreign word should be the scavenger hunt item and whether the player discovers it. However, unlike traditional video game implementation that assumes the game logic is consistent and correct if it is bug-free, computer vision has high inaccuracy rate.

For the current game prototype, we use the computer vision model, *SSD MobileNet V1* [91], for objects segmentation and recognition. The top-1 accuracy rate is reported to be 70.6% [157], which suggests that in average 3 out of 10 times of recognizing an object provide a wrong answer. The inconsistency of recognition results will cause players' frustration and become a negative effect to the gaming experience.

The biggest challenge in user experience design is to work around the technical constraints posed by current state of computer vision technology: (1) the object recognition is unstable and falsely identifies objects with a high probability, and (2) The player's motion makes visual annotation anchoring a hard problem. A high enough frame refresh rate is required to respond to the player's movements.

The first technical constraint is noticeably annoying while the player chooses an object and waits for confirmation. In order to reduce the chances of missing frames or changing labels during the confirmation phase, we designed an algorithm shown as Algorithm 1 using the temporal smoothing technique.

Because computer vision is unreliable and can sometimes make mistakes, it is also crucial that *CollectiAR* can inform the player when mistakes occurred. Because we need to prevent players from forming incorrect connections between physical objects and vocabularies in the learning process. When actively challenging players to look for the matching object for the target foreign word, we added native language labels for each identified objects on screen. This would allow players to quickly discover falsely recognized objects by computer vision.

To circumvent the second technical constraint, we designed the foreign word pronunciation to be played automatically and entirely during game play. One scenario with unstable computer vision technology is that it successfully recognized an in one frame, but failed in the next frame; therefore, players would observe the bounding box surrounding that object appears and disappears on screen. In order to allow players continue playing our game despite the blinking bounding box caused by unstable object recognition, we designed Korean word pronunciations to be played immediately and automatically as players point the camera focus inside the bounding box of the object. The pronunciations would be played entirely even if the object failed to be recognized by computer vision in the next frame. Therefore, hearing the Korean pronunciations only requires

Algorithm 1: Temporal Smoothing to Reduce Glitches During the Confirmation Phase

```
Input: nowDetection, nowImage
Output: resultDetection
Data: cachedDetection, labelsDictionary, hasMissed
if nowDetection is not null then
    // show the highest frequency label
    resultDetection = nowDetection;
    labelsDictionary[nowDetection.label] += 1;
    Replace resultDetection.label with labelsDictionary's key that has the
      max value;
    cachedDetection = resultDetection; hasMissed = false;
else
    if hasMissed is true then
        resultDetection = null ;
        Clear labelsDictionary and cachedDetection;
    else
        // check if previously detected object still stays there
        hasMissed = true;
        cachedBlock = the center 4x4-pixel block of cachedDetection's
          image;
        nowBlock = the 4x4-pixel block in nowImage at the same location
          of cachedBlock;
        backgroundBlock[2] = two 4x4-pixel blocks in nowImage that are
          outside of cachedDetection's bounding box;
        Calculate sum of absolute differences (SAD) between
          cachedBlock and nowBlock;
        Compare it with the SAD between cachedBlock and 2
          backgroundBlocks repectively;
        if SAD(nowBlock, cachedBlock) is min then
            resultDetection = cachedDetection;
        else
            resultDetection = null ;
            Clear labelsDictionary and cachedDetection;
```

the objects to be recognized by at least one frame in our game. It is not necessary for players to tap the sound icon inside the labels. The sound icon mainly serves as an indicator that pronunciation is being played and players may choose to tap the icon to hear the pronunciations again. Playing the pronunciations automatically and entirely enables players to make decisions on whether an object matches the target word and continue playing our game without solely relying on visual feedback.

3.4 User Study Design

We conducted a user study to determine whether people find *CollectiAR* engaging and evaluate how it compares in terms of efficiency and convenience to basic digital flashcards software. We proposed the following 4 hypotheses for the study: **H1:** Immediately afterwards, participants who played *CollectiAR* will recall more words than participants who used digital flashcards; **H2:** After a period of 5-7 days, participants who played *CollectiAR* will recall more words than participants who used digital flashcards; **H3:** Participants will find *CollectiAR* more engaging than digital flashcards; **H4:** Participants will pay more attention when playing *CollectiAR* as opposed to using digital flashcards.

For this study, we chose Korean as a target language. In order to compare *CollectiAR* to digital flashcards, we used the popular *Quizlet* digital flashcards software. Participants used the same 10.5-inch iPad pro (2018 model) to use both *CollectiAR* and *Quizlet*.

3.4.1 Lab Setup

Participants were randomly assigned to one of two groups: Group A and Group B. Group A used *Quizlet* and then played *CollectiAR*; Group B did this in the reverse order: they played *CollectiAR* first and then used *Quizlet*. In order to directly compare the correctness rate under a 2x2 within-subjects experimental evaluation, we generated two different word lists each containing 9 words that were assigned to the first and second tool participants encountered, respectively. In other words, the set of words that Group A could learn from *CollectiAR* were the same as those that Group B could learn from *Quizlet* and vice versa. These 18 words were a subset of the 79 words that *CollectiAR* could teach. The other 61 words were not included in this user study.

In order to test how *CollectiAR* helps players learn words from their environment, we constructed four different environments, which we will refer to as "zones". The set of words were separated into these 4 zones, each of which mimicked one of the following real-life learning scenarios: living room, dining room and office. Table 3.1 described each zone's learning scenario and the physical items that were placed in the zone. Participants in Group A played *CollectiAR* in Zone 1.1 and 1.2. Group B learned the words in Zone 1.1 and 1.2 from *Quizlet*. Similarly, Group B played *CollectiAR* in Zone 2.1 and 2.2 and Group A learned those words from *Quizlet*.

Zone	Items in the Zone
1.1	*book*, *scissors*, ball, racket, umbrella
1.2	*book*, *scissors*, vase, spoon, bowl, clock
2.1	*chair*, *cup*, knife, apple, cellphone
2.2	*chair*, *cup*, toothbrush, wineglass, bottle, backpack

Table 3.1: The setup of zones for *CollectiAR*.

CollectiAR could teach 9 words in Zone 1.1 and 1.2 and another 9 words in Zone 2.1 and 2.2. Zones 1.1 and 1.2 had two words that appeared in both zones (in italic font in Table 3.1). The same was true for Zones 2.1 and 2.2. These words referred to two a unique physical object in each zone. The cup in Zone 2.1 was a teacup with a handle while the cup in Zone 2.2 was a Chinese teacup with no handle. One of the books was opened while the other was in a stack of books. The scissors and chairs had different colors and sizes.

We also developed two *Quizlet* study sets that were corresponding to Zones 1 and Zones 2. The set of flashcards consisted of the word's forms (in Korean characters and romanization and in English), one object image that depicted the word and the word's pronunciation in Korean.

3.4.2 Procedure

Our study had two learning sessions. In the first learning session, participants were expected to learn 9 Korean words using one app for 10 minutes. In the second learning session, they learned another 9 Korean words using the other app for 10 minutes. The order of the apps depended on the group to which the participant was assigned.

Before each learning session, we asked participants to take a pretest. Upon completing the pretest, participants were instructed to review all 9 Korean words and their corresponding English translations only once. To prepare players for the words they were expected to find in the session, we briefly showed them a list on the iPad of words and translations. Participants were instructed to quickly read through the list and only look at each word once and not memorize. Participants generally completed this preview in less than 30 seconds.

Figure 3.5: *CollectiAR* presents the progress status after participants collect or skip a word in the user study.

When playing *CollectiAR*, participants needed to complete each zone in order. Figure 3.5 shows how *CollectiAR* informed participants about their progress. They were able to skip words if desired. Each time a participant started exploring a new zone, the order of the words was randomized. Each word only appeared once. If players finished exploring all two zones, we instructed them to start over again from the first zone until time was up.

When using *Quizlet*, participants were asked to study in *Quizlet*'s learn mode [7] using the *flashcards* question type. If the learn mode's progress reached 100%, participants were instructed to restart the learn mode and continue reviewing until time was up. Participants could stop using the app if they did not want to continue to study but were not allowed to do anything else until time was up.

Immediately after each learning session, we asked participants to complete a posttest of the Korean words that they just learned. After participants finished both learning sessions, we also asked them to fill in a post-learning survey, where they rated their perception of novelty, engagement level, and attentiveness during the two learning sessions. We also interviewed them and collected data on their previous AR experience, their overall experience and if they have any suggestions on improvement of our game.

In order to evaluate **H2**, we needed to follow up with participants after a delay to measure longer-term retention. Therefore, after a period of 5 days, we sent out emails to participants and asked them to complete a delayed posttest within 2 days.

3.4.3 Participants

CollectiAR targets young professionals and adults. With this in mind, we recruited participants through fliers, face-to-face interactions, and our university's research participation system. A total of 42 university students (20 males and 22 females) participated in this study. Potential participants were screened based on the criteria that they were 18 years or older and could not understand or use simple Korean characters. Four of the 42 participants did not complete the delayed posttest. Participants who signed up through the university's research participation system were compensated with course credits.

We observe that other languages spoken by some participants may have helped them learn the meanings of a few Korean words. For instance, umbrella is pronounced as ***usan*** in Korean and ***yüsan*** in Chinese. They share the same number of syllables and the same ending syllable.

3.4.4 Measures

Pretest, Immediate Posttest, and Delayed Posttest

We used the pretests, the immediate posttest and the delayed posttest to verify hypothesis **H1** and **H2**. These tests asked participants to type the English translation of Korean words using the *Qualtrics* online survey tool [6]. The Korean words were presented in random order and the Korean pronunciations were available. Participants took the pretest and the immediate posttest in the lab within view of the physical zones used for *CollectiAR*. The delayed posttest was the same as the immediate posttest but participants filled it out wherever they were.

We calculated the correctness rate for the knowledge tests by dividing the number of words with correct English translation by the total number of Korean words. We still considered the answer correct even if it had a minor spelling error (i.e. missing double letters). Judging from a significant shift in responses, it appears as though we simplified the English translation shown in both *CollectiAR* and *Quizlet* from *"teacup"* to *"cup"* in the middle of the study. We treated either answer as correct and analyzed the participants as one group because we did not think this significantly affected retention.

Post-Learning Survey

To collect evidence to verify hypotheses **H3** and **H4**, we asked participants to fill out a post-learning survey to rate their learning experience for two learning sessions on a 5-point Likert scale on the following measures:

Novelty. "On a scale from 1 (almost none) to 5 (completely novel), how likely does it provide you new learning experiences that you have never tried before?"

Engagement. "On a scale from 1 (worst) to 5 (best), how much did you enjoy studying the vocabulary?"

Attentiveness. "On a scale from 1 (strongly disagree) to 5 (strongly agree), do you agree that you were focusing on the activity and not affected by any external distractions?"

We also asked participants the following open-ended questions through the survey and interviews about 1) their previous experience with AR apps, 2) their willingness to use our game to learn a foreign language, 3) the pros and cons of our game compared with flashcards, 4) any difficulties they had when playing our game, and 5) any comments or suggestions to improve *CollectiAR*.

3.5 Results

We found evidence supporting hypotheses **H2** and **H3** but did not find evidence to support **H1** or **H4**.

3.5.1 Evidence of Increased Effectiveness

We conducted a two-way repeated measure ANOVA to examine students' knowledge of Korean words before, immediately after and 5-7 days after two learning sessions in the lab. Overall, participants were able to significantly re-

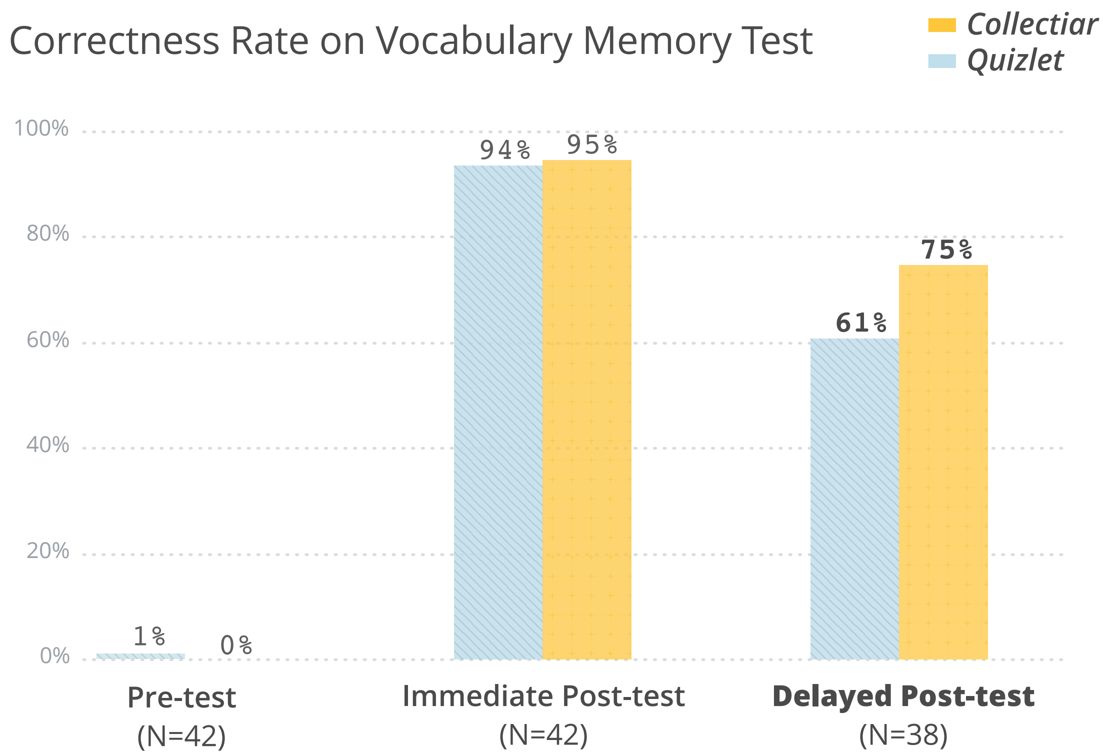

Figure 3.6: The correctness rate. No differences in immediate retention rates between playing *CollectiAR* and using *Quizlet*, but playing *CollectiAR* results in 23% higher retention (statistically significant) than using *Quizlet* after 5-7 days.

call more foreign words after using *CollectiAR* and *Quizlet* ($F(1, 37) = 8.54$, $p < 0.01$). There was a statistically significant interaction between the effects of time and learning methods on vocabulary retention ($F(2, 74) = 11.59$, $p < 0.001$), suggesting that the effects of learning methods differ over time.

As shown in Figure 3.6, The average correctness rate immediately raised from 0% to 95% after playing the *CollectiAR* and dropped to 75% in the delayed post-test after 5-7 days. We observed a similar trend in the case of using *Quizlet*: the average correctness rate rocked up from 1% to 94% in the immediate post-test but dropped further down to 61% after 5-7 days than the case of playing the *CollectiAR*.

Specifically, simple main effect analysis suggested that there was no differences in the correctness rate between playing *CollectiAR* and *Quizlet* immediately after learning ($p > 0.5$). Participants did not recalled significantly more words learned playing *CollectiAR* compared to words learned through digital flashcards in the immediate post-test. This result did not support our hypothesis, **H1**, that our game is not more effective for vocabulary retention than online flashcards immediately after learning. It aligned with previous research that frequency-based method is effective in the short term.

The simple effect analysis supported our hypothesis, **H2**, on long-term vocabulary retention. The result showed that participants recalled 23% more words learned playing *CollectiAR* compared to words learned through online flashcards in the delayed post-tests ($p < 0.001$), which supported our hypothesis that *CollectiAR* is more effective than digital flashcards in the long term. The result also aligned with related work that showed the visual and spatial context is more effective for learning in the long term.

3.5.2 Evidence of Increased Engagement

We compared the scores of the 5-point Likert scale using Wilcoxon signed-rank test. Figure 3.7 visualizes the rating results.

We did not observe a difference in level of attention ($Z = -1.4501$, $p > 0.05$) between *CollectiAR* (median = 4) and *Quizlet* (median = 4). The result does not support our hypothesis, **H4**, that participants will pay significantly more attention while playing our game compared to learning through flashcards.

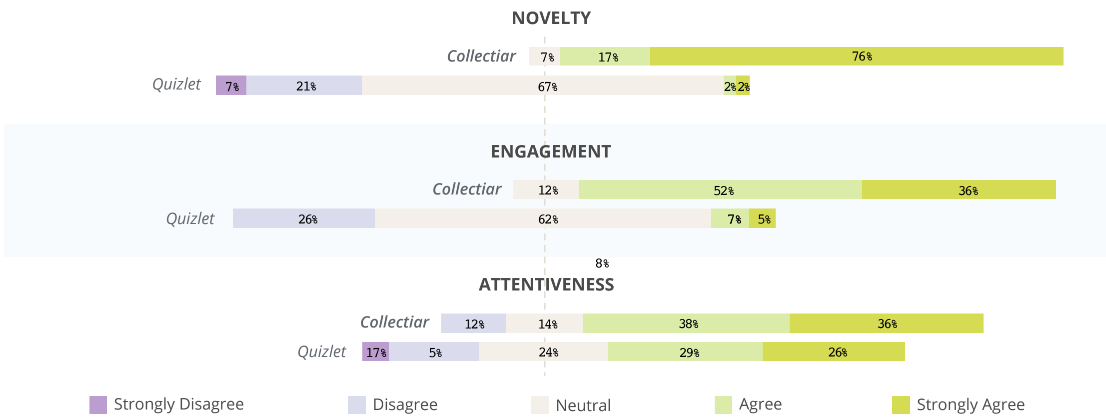

Figure 3.7: 5-Likert Scale ratings by 42 participants. 93% of the participants agree or strongly agree that *CollectiAR* is novel and the number is 4% for *Quizlet*. 88% of the participants agree or strongly agree that *CollectiAR* is engaging and the number is 12% for *Quizlet*. 74% of the participants agree or strongly agree that *CollectiAR* is attentive and the number is 55% for *Quizlet*.

In another study [158], however, researches found that the learners pay less attention to the digital flashcards. One possible reason to explain this phenomenon is that that's because our in-lab study is relatively shorter than their study. In their study, participants are required to use the app to learn for a week, whereas we only required a one-time in-lab study for 30 to 40 minutes.

The data on engagement showed that the level of engagement was statistically higher for the our game (median = 4) than the flashcards (median = 3) group using Wilcoxon signed-rank test ($Z = -4.7752$, $p < 0.01$). Therefore participants were more engaged while learning vocabularies using our game, compared to learning using online flashcards. This supports our hypothesis, **H3**, that our game is more engaging and the result also aligned with existing research [132] that "gamification introduces higher level of engagement than traditional instructional methods".

3.5.3 Overall Impressions

We asked participants open-ended questions in the post-learning survey and interviewed them about their overall experience with our game. We looked into their responses and many participants expressed that our game is fun and engaging. For example, *"More engaging. Can be used as a dictionary to look up things in real world."; "More interesting because you need to actively looking for the objects in the room"; "The content of Flash cards are predesigned and less adaptive but the AR app is more flexible in terms of finding different instances that can be described by the given words."*.

We also asked participants whether they would use our game to learn a new language. Many participants reported they would used our game in the future for language learning. As some of our participants explained: *"It depends on the purpose of learning. I would choose this if I want more fun during my study."; "I will. Quizlet is too static. It's less meaningful and less interactive."*.

We also received constructive criticism about *CollectiAR*'s inconvenience and limitations that we are going to discuss in the next section: *"my main concern is that the vocabulary is limited if I only study words that represent physical objects I can see around me."; "Object detection was not faster enough, and sometimes the objects are identified incorrectly as something else."*.

3.5.4 Limitations of the study

In our study, participants used vocabulary learning tools in a controlled lab setup. These conditions may differ from the usage patterns that players would typically demonstrate in the wild; for example, players might open *CollectiAR* or *Quizlet* multiple times during the day for a short duration. Therefore, further investigation in the field is necessary to understand whether the benefits observed here would translate into more typical use. Nevertheless, these results provide preliminary evidence that an AR game can enhance the vocabulary learning experience.

3.6 Discussion

Our study showed that the user ratings of engagement and performances on delayed vocabulary tests were significantly higher for *CollectiAR* when compared to *Quizlet*. The results suggested that *CollectiAR* met two design goals: a vocabulary learning game that is both engaging and effective in knowledge retention.

To engage players in learning, we used game mechanics to actively challenge users as opposed to passively memorizing predefined words and images. We leveraged computer vision technology so that players can use their immediate vicinity as the gaming environment instead of using predefined markers that are unintelligible outside of a game.

To effectively facilitate memorization, we used contextual learning method in addition to the design of repetitive practice. We enable users to spontaneously associate multiple knowledge of a word, such as pronunciation, text, visual and spatial information.

However, we observed that *CollectiAR* is not so convenient as *Quizlet* for students to learn as many words as possible. players can not use the current prototype of *CollectiAR* to learn an arbitrary set of words that are available in *Quizlet*. Besides, *CollectiAR* is less convenient than *Quizlet* when players need to learn without walking in environment. That said, *CollectiAR* is more convenient than existing AR vocabulary learning games that rely on predefined markers because players are able to play *CollectiAR* anywhere at anytime. Figure 3.8 visualizes the comparison between *CollectiAR*'s overall performances and those of *Quizlet* in our study.

Figure 3.8: Comparison between *CollectiAR*'s overall performances in the user study and those of *Quizlet*.

3.6.1 Design Implications

We summarize two critical limitations from our study that affected the overall player satisfaction and propose solutions to address them: 1) players cannot learn words that do not associate with the physical objects that the AR game can detect; 2) players cannot play the game smoothly if the computer vision AI continuously makes mistakes. These limitations would be imposed on any AR game design that aims to engage players with context-sensitive content.

Need to generate virtual 3D objects in the real environment. Our current *CollectiAR* prototype cannot teach words that do not directly correspond to physical objects located in the immediate vicinity of the learner. In order to break free from this constraint to teach more words, the next version of *CollectiAR* would ideally be able to use AR technology to project virtual 3D objects into the player's immediate vicinity. For example, players who do not keep a cat as a pet could learn foreign words for cat by capturing a virtual cat in their room. Because most of the screen real estate (i.e. the real world) is out of content designers' control, the ability to mix virtual and physical objects in the real environment also provides them the freedom to design progression in gaming and learning.

The biggest technical challenge is to make virtual objects look real in the environment. To achieve this goal, an AR game needs to detect firstly the ground and walls for perspective and secondly the depth of physical objects for occlusion. We were not able to implement occlusion in AR while we developed the current prototype of *CollectiAR* because we did not have a depth sensor for the mobile phone. We found that the deep learning approach [69] is too computationally expensive to run on a mobile phone for the present. Alternatively, we suggest that AR mobile game developers should consider implementing this feature using a depth sensor that directly provides dense point clouds information.

Need to design a feedback loop between the player and the computer vision AI. If the computer vision AI continuously makes mistakes, players can be confused and thus quickly loose interests in the game. In our case, *CollectiAR* constantly mistakes a knife for a pair of scissors. A solution to improve the computer vision AI's accuracy is to use players' feedback to retrain the computer vision AI. An option is to ask players to confirm a bunch of recognition tests during the scanning phase before the game starts. Another option is to allow players to flag wrong recognition instances during the game.

It is worth noticing that privacy concerns need to be addressed when designing such a feedback loop. Asking for the user's permission to send data to the server only once when opening the app for the first time cannot protect users' privacy in this case. Different images (and their metadata for AI training) contain different levels of user's sensitive information. Ideally, players should be informed and able to understand what exact data will be uploaded to the server before they permit or reject.

3.6.2 Study Limitation

The user study for *CollectiAR* game showed that the participants who played the AR game performed better in the delayed test than the participants who used flashcards to learn Korean words. Although the result showed that the AR game was more effective than the traditional flashcard approach, we did not specify the reason behind this phenomenon. It is unclear if the higher delayed recall is due to motivating the participants to spend more time on learning Korean words after the experiment; or it is due to the theory about recalling the details of a learning activity more easily if the activity is kinesthetic rather than on-screen; or both. To distinguish the mechanisms that contribute to learning performance, in the future we would also compare *CollectiAR* game with an AR vocabulary e-learning tool that is not a game.

3.6.3 Applicability to Other Learning Scenarios

We observe that *CollectiAR*'s design can be re-skinned to address other pedagogical scenarios. For example, a similar design could be used to teach people how to identify plants, animals, museum artwork, etc. Existing educational games addressing these topics often teach within pre-designed virtual world. For example, *Quest Atlantis* [28] is a game that encourages students to explore a 3D virtual learning environment to solve puzzles together. A puzzle, called a Quest, can be *"looking for an example of a given species"*. We believe that our game design could be extended to implement something like an outdoor exploration version of *Quest Atlantis*. Instead of asking players to find the object that corresponds to an unknown foreign word, the game could ask the names of plants and animals.

3.7 Conclusion

Facilitating language learning through technology is a challenging research topic. An ideal vocabulary learning app should be engaging, effective and convenient to use. To achieve these goals, especially engagement, we designed a foreign word scavenger hunt mobile game, *CollectiAR*, that uses computer vision and augmented reality.

We conducted a 2x2 within-subjects experimental evaluation (N = 42) to evaluate the performance of *CollectiAR* compared to a passive spaced repetition digital flashcards software, *Quzilet*. The results supported two hypotheses: 1) *CollectiAR* outperforms *Quzilet* on productive recall tests administered after

5-7 days by 23% ($p < 0.01$). 2) the self-reported ratings showed that *CollectiAR*'s learning experience was more enjoyable than *Quzilet*'s. Although the current prototype of *CollectiAR* is less convenient than *Quzilet* because of two critical limitations, there are plausible solutions to overcome the limitations. Our study indicated that computer vision-based AR games can be beneficial for language learning and other pedagogical purposes.

CHAPTER 4

TEACHER VIEWS OF MATH E-LEARNING TOOLS FOR STUDENTS WITH SPECIFIC LEARNING DISABILITIES

4.1 Introduction

According to the National Center for Education Statistics [133], five percent of public school students in the US were identified as having a specific learning disability (SLD) in 2019, although the National Center for Learning Disabilities believes the real figure is even higher [40]. SLDs predominantly affect students' abilities to understand or use language, especially in reading (dyslexia), math (dyscalculia), and writing (dysgraphia). While students with dyscalculia have difficulty with math by definition, students with other SLDs also tend to struggle with math problems that involve reading and writing [40].

Nevertheless, research shows that many students with SLDs can develop proficiency in math skills when teachers provide one-on-one support and personalized instruction that adapts to students' learning styles, abilities, and interests [97, 151]. For example, teachers can present content using various sensory modalities and adjust the difficulty or length of assignments, according to the needs and preferences of each student. However, employing these strategies is time- and resource-intensive [36, 60], making it challenging for teachers to support a large group of students with diverse learning needs.

Recently, new math e-learning tools have gained popularity across the US [142], many of which support personalized math practice. For instance, drill-and-practice websites (e.g., *Khan Academy* [101]) and educational digital games (e.g., *Prodigy* [170]) allow students to practice math exercises independently and receive immediate feedback on their performance. These tools have the potential to provide personalized support to some students while teachers are helping others, or otherwise not available.

A few math e-learning tools (e.g., *The Number Race* [204] and *Calcularis* [95]) were designed specifically for students with SLDs. These tools use interactive visual media and symbols to present problems to students who struggle with processing text-based math problems. For example, *The Number Race* [204] is a game that challenges students with SLDs to read digital symbols that are represented by a car's speedometer in simulated car racing games. A lab study conducted by Wilson et al. [205] showed that students with SLDs made fewer mistakes in whole number calculations after playing *The Number Race*. While these results were promising, they are limited because they were conducted in the lab, removed from the complexities of a classroom setting. In fact, to the best of our knowledge, all prior research evaluating math e-learning tools for students with SLDs has involved only lab studies where students completed tasks using the tools in isolated sessions [95, 205].

While lab studies can tell us whether a tool can help a student learn a particular concept over a short period of time, they cannot tell us about the use patterns, challenges, and effectiveness of these tools in real life. In the context of real life, e-learning tools must be integrated into the student's curriculum as homework or classwork. Teachers must select and assign tasks on these tools as part of their teaching. So, while prior research has shown the potential of e-learning tools for helping students with SLDs learn certain math concepts, we do not know whether they support students' needs in the math classroom. To address this gap, we first turned to the student's teachers. Specifically, we asked the following research questions:

1. Which math e-learning tools are teachers using?
2. Why and how do teachers use these tools?
3. How do teachers perceive the effectiveness of these tools for students with SLDs?

To answer these questions, we conducted semi-structured interviews with 12 US-based teachers who have experience teaching math to students with SLDs in grades five to eight. We focused on this grade range because little attention had been paid to it by prior research [137]. In terms of e-learning tools, we mostly focused on recent websites and applications that could provide personalized exercises, such as drill-and-practice websites and educational digital games, rather than calculators or interactive whiteboards. We asked our participants about their use of math e-learning tools in teaching students with SLDs, their challenges in adopting the math e-learning tools, the accessibility issues that their students faced in using the math e-learning tools, and their teaching strategies for students with SLDs.

We found that our participants used five drill-and-practice websites and six educational digital games for math skills training, none of which were specifically designed for students with SLDs. Overall, participants used these tools for four main purposes: (1) independent practice in the classroom, (2) motivation and engagement, (3) assessment, and (4) tracking student performance. Participants found four critical challenges that made it difficult for them and their students to use these tools for the aforementioned purposes: (1) drill-and-practice websites were text-intensive; (2) e-learning tools provided insufficient feedback to teachers about their students' performance; (3) e-learning tools did not allow teachers to adjust difficulty levels; and (4) e-learning tools were difficult for teachers to setup and maintain. Participants also mentioned that they wanted their students to use assistive technologies (e.g., text-to-speech, closed captioning) within the e-learning tools but they had difficulty acquiring and using them. Based on these findings, we propose design implications for math e-learning tools that address gaps noted by our participants. For example, math e-learning tools should not depend on typing or multiple-choice answers for input but instead allow students to communicate math content using additional modalities, such as using voice-to-text and moving math manipulatives.

In summary, we contribute a study that sheds light on real-life patterns and challenges of using e-learning tools for students with SLDs from their teachers' perspectives. Since teachers are the ones who incorporate e-learning tools into the curriculum, our study represents an important first step into understanding how e-learning tools can support students with SLDs. We further contribute design implications which can help shape the design and creation of more inclusive and effective e-learning tools.

4.2 Related Work

In this section, we present related work on helping students with SLDs succeed in learning math. First, we present research on the learning difficulties that students with SLDs experience and identify pedagogical strategies that support these students. Second, we present research on recent math e-learning tools and their use by students with SLDs. Finally, we present research on the challenges that teachers face when adopting e-learning tools in their teaching.

4.2.1 Math Education for Students with SLDs

Students with SLDs have neurological differences that affect their abilities to understand or use language, especially in reading (dyslexia), math (dyscalculia), and writing (dysgraphia) [176]. Dyslexia is characterized by labored, inaccurate, and slow reading, along with spelling difficulties; dyscalculia causes difficulty with math concepts, calculating, number sense, number language, and problem-solving; and dysgraphia affects writing, including handwriting, spelling, and organizing written language.

Karagiannakis et al. [94] categorized the math learning difficulties of students with dyscalculia into four categories: (1) number sense, (2) working memory, (3) visuospatial skills, and (4) reasoning. Number sense refers to the ability to recognize number symbols, to compare quantities, and to perform calculations; working memory lets people hold on to information temporarily for processing, like a mental sticky note; visuospatial skills involve the ability to mentally process and manipulate visual objects in more than one dimension; and reasoning is the ability to make structured and logical inferences from supporting arguments and evidence. Furthermore, Soares et al. [171] added that students with dyscalculia might have an additional SLD in reading (dyslexia) or writing (dysgraphia) that further affects their ability to solve math problems. Nevertheless, studies [180, 199] have shown that specific instructional interventions can help students with SLDs achieve math proficiency. One such strategy is multimodal instruction, in which teachers both present contents using multiple sensory modalities (e.g., sight, sound, touch, and movement) and also offer students opportunities to express their understanding through multiple modes [199]. For example, a teacher might use math manipulatives, such as base 10 blocks, to engage tactile learners in a lesson on whole number calculations. Another effective strategy for teaching students with SLDs is differentiation [180], in which teachers adjust aspects of their lessons (e.g., difficulty level and length), based on each student's needs. A third strategy is thinking aloud, in which teachers model their thought process while solving problems and prompt students to share their thought processes. All of these approaches require one-on-one attention from teachers, making them both time- and resource-intensive [36, 60].

Although prior work has investigated the difficulties that teachers encounter in employing certain pedagogical strategies, it has not investigated whether and how math e-learning tools can support or detract from use of these strategies. Our study is the first to investigate whether and how teachers are integrating math e-learning tools as they adjust their teaching to their students with SLDs.

4.2.2 Math E-Learning Tools for Students with SLDs

Math e-learning tools serve a wide range of roles in math education, from providing a simple function like calculation to teaching math. A decade ago, researchers found that students with SLDs benefited from light-tech tools such as calculators and interactive white boards [32]. Recently, new math e-learning tools that provide richer features have emerged, such as video lessons [101], math exercises [93], and virtual manipulative objects [33]. Research on recent math e-learning tools for students with SLDs has mainly focused on: (1) personalizing math skills training and (2) making digital math content accessible.

Math Skills Training Personalization.

Recent research on math e-learning tools for students with SLDs [27, 102, 142, 145, 150] has primarily aimed to increase students' academic scores by providing supplemental math skills training without teachers' direct support. Recent math e-learning tools can be divided into two categories: drill-and-practice training and game-based training [120]. We summarize prior work in these two categories respectively.

Drill-and-practice training presents math practice in the form of standardized tests and provides immediate feedback on whether the student's answer is correct or not. Many websites (e.g., *Khan Academy*, *IXL Learning*, *CMP 3*, *BrainPOP*, and *Flocabulary*) provide K-12 math drill-and-practice training. Some of the drill-and-practice websites also provide free video lessons for students to learn corresponding math concepts independently. However, these websites were not designed specifically for students with SLDs. One of the goals of our study is to discover whether and how well teachers and their students with SLDs use these websites in math education.

To our knowledge, *Calcularis* [66] is the only curriculum-based math e-learning tool designed specifically for students with SLDs. This tool visually represents elementary school math problems. For example, it teaches the concepts of whole numbers, quantity, and distance by placing whole numbers on a number line from 0 to 100. It also automatically selects an appropriate practice problem based on the student's correctness rate. Research showed that, after using *Calcularis*, students with SLDs attained higher scores on grade-level math tests [114, 95]. While *Calcularis* shows promise, it is still unknown whether and how teachers are integrating this and similar tools into their classes.

Game-based training aims to transform math practice into fun games with visual representations of math concepts and concrete examples to contextualize math skills. These games do not provide explicit feedback on incorrect answers, but let the player keep trying until they find the correct answer [58]. Some educational digital games, such as *ST Math* [125] and *Prodigy* [170], cover a wide array of math topics, spanning from whole-number calculations to two-dimensional geometry. Other games target a specific math skill. For example, *Slice Fractions* and *Refraction* assist in practicing arithmetic operations with fractions. None of these games, however, were designed specifically for students with SLDs.

Ke et al. [99] studied how three games designed for general education students could improve students with SLDs' performance in math. For example, one of the games, *Ker-Splash*, presented players with a mathematical expression and asked them to create one with a greater numerical value. Ke et al. observed nine students with SLDs playing this game. They found that the game fostered positive attitudes towards math practice, but only four students (44%) received higher scores on a post-test after playing *Ker-Splash*. In addition, when asked what they learned from the game, none of the students mentioned any math concepts or problem-solving procedures. The study was limited by its small sample size, and it did not compare the effects of different game features (e.g., timed task, visualizing math symbols) on engagement and learning , so it is difficult to draw conclusions.

The Number Race [204] and *Number Catcher* [4], two games designed by the French National Institute of Health and Medical Research, specifically target students with SLDs. Both games cover the recognition of whole numbers. These games present Arabic, verbal, and visual representations of numbers together for the player to decide which whole number is larger and which is smaller in a timed game. Wilson et al. [204] conducted a user study with nine participants with math difficulties, aged seven to nine years old, to find out whether these children could gain numerical proficiency after playing *The Number Race* independently for half an hour a day, four days a week, over a period of five-weeks. Their results suggested that *The Number Race* could potentially help children with math difficulties increase number sense over short periods of lab study, but, as with other research in this area, more work is needed to test its use in a real-life context.

Accessible Digital Math Content.

Another research direction related to math e-learning tools for students with SLDs has focused on making digital math content accessible to students who struggle with reading and writing. To achieve this, researchers have followed two main approaches: leveraging speech technologies [31, 72, 174] and providing virtual math manipulatives [184, 167, 49].

Several tools [31, 72, 174] enable students to verbally communicate math content. *Mathtalk* [174] reads algebra through text-to-speech. *MathShare* [31] is a specialized text editor to help students keep math work aligned, type math symbols using shortcuts, and think out loud using speech-to-text. *ViewPlus* [72] is an audio graphing calculator that allows students to plot y-versus-x graphs using speech. To our knowledge, however, no prior work has studied whether and how teachers would use these tools to personalize the math education of students with SLDs.

Research has shown that using math manipulatives allows students with SLDs to communicate math content more easily than using written or verbal formats [129, 17]. Currently, several applications using virtual math manipulatives [184, 167, 49] are available on the market. For example, *The Base Ten Blocks* mobile application [49] presents whole numbers with virtual blocks. *NLVM (National Library of Virtual Manipulatives)* software [184] is a digital library containing manipulatives for numbers, operations, and geometry. *NLVM* provides virtual manipulatives in the toolbar, such as a single block representing 1 and a 10-based block representing 10. If the user generated three single blocks and five 10-based blocks in the central panel, then *NLVM* displayed the number 53.

Nevertheless, virtual manipulatives applications without math exercises might not be effective for students in improving their math skills. Research [128, 203] found that the effectiveness of math manipulatives depended on whether students understood the mathematical meaning of each action they performed. Without the direct support of teachers, however, those virtual manipulatives applications could neither teach students how to use the manipulatives to solve math problems nor ensure students understand what they were doing with the manipulatives.

4.2.3 Adoption of E-Learning Tools by Teachers

Standalone e-learning tools are not effective unless teachers adopt the tools in their teaching [78, 136]. Therefore, some recent research focused on addressing teachers' needs and challenges in adapting e-learning tools into their math instruction. For example, interview studies [38, 85] reported that teachers needed to customize e-learning tools for individual students because teachers wanted to manage different students' learning progress, but could not do so. Vermette et al. [186] identified the challenges that teachers faced in customizing e-learning tools for varying student needs, including setting different difficulty levels of learning content and customizing UI preferences. To enable teachers to quickly identify individual students' performance on drill-and-practice websites during the class, Holstein et al. designed dashboards on computers [87] or on wearable smart glasses [86] that visualized students' statistical data to the teacher in real-time. Prior work has investigated how to help teachers adopt e-learning tools for general education students [87, 86], but comparatively little work considers teachers for special education students, especially students with SLDs.

4.3 Methods

We conducted a semi-structured interview study to investigate which math e-learning tools teachers were using when teaching students with SLDs, why and how teachers used these tools, and how effective teachers felt the tools were for their students with SLDs.

4.3.1 Participants

Pseudonym	Age/Gen	State	Teaching Certification	Teaching Grade(s)	Teaching Environment(s)	Years of Teaching
Carol	55/F	WA	Special Ed.	Grade 3-6	Self-Contained	8
Marcus	32/M	WA	Special Ed.	Grade K-5	Self-Contained	5
Savannah	25/F	NY	Special Ed.	Grade 5	Self-Contained & Integrated Co-Teaching	1.5
Helen	55/F	NY	Special Ed.	Grade K-5	Integrated Co-Teaching[1]	35
James	48/M	NY	Special Ed.	Grade 7	Integrated Co-Teaching	5
Victor	27/M	NY	Special Ed.	Grade 8	Integrated Co-Teaching	6
Lisa	27/F	NY	Special Ed.	Grade 6	Integrated Co-Teaching	5
Amy	45/F	GA	Special Ed.	Grade 6	Integrated Co-Teaching	7
Selena	41/F	NY	General Ed.	Grade 6	Integrated Co-Teaching	16
Darcy	62/F	NY	General Ed.	Grade 8	Integrated Co-Teaching	27
Ian	34/M	NY	General Ed.	Grade 8	Integrated Co-Teaching[2]	10
Jennifer	53/F	NY	General Ed.	Grade 8	Integrated Co-Teaching	15

[1] Helen was working as a math coach together with teachers in the classroom to help students with SLDs learn math.
[2] Ian was teaching students with SLDs together with special education SETSS providers.

Table 4.1: Participant pseudonyms and demographic information.

We recruited via email and social media 12 US-based participants (eight females, four males) who had been teaching math to students with SLDs in public schools. Participants had to be either (1) state-certified in special education and have experience teaching math in grades five through eight, or (2) state-certified in general education and have experience teaching math in grades five through eight alongside a special educator in an integrated co-teaching (ICT) classroom. Two out of the 12 participants were only teaching students with disabilities in self-contained classrooms that contained one special-education teacher; nine participants were teaching students with and without disabilities in integrated co-teaching (ICT) classrooms that contained two co-teachers; and one participant had taught in both types of classrooms by the time of our interview. Their ages ranged from 25 to 62 (mean=42). They had between 1.5 and 35 years of experience in teaching (mean=11.7). In terms of location, participants were living in three states at the time of the interview: One in Georgia, two in Washington, and nine in New York. Table 4.1 shows participant pseudonyms and their demographic information.

4.3.2 Procedure

The study included a semi-structured interview that lasted 40 to 60 minutes. Interviews were conducted face-to-face, over the phone, or via video conference software (Google Hangouts). Participants received a $ 20 Amazon gift card upon completing the interview. We began by asking participants about demographic information and their job duties. Then, we asked questions about participants' use of different e-learning tools, their experiences and challenges with these tools, and their special teaching strategies for students with SLDs. Our questions were grouped in the following categories:

1. Pedagogical strategies for students with SLDs: e.g., Could you give me an example of using a special pedagogical strategy for students with SLDs

and tell me why you use it?

2. Challenges of teaching students with SLDs: e.g., Have you encountered any difficulties when tailoring math exercises for students with SLDs?
3. Use of math e-learning tools: e.g., Which e-learning tools have you used to help students with SLDs learn math and how have you used the tools? What is your purpose for using this e-learning tool?
4. Effectiveness of math e-learning tools: e.g., Do you think the e-learning tool you used is effective for students with SLDs? If so, why?
5. Accessibility of math e-learning tools: e.g., Have your students encounter any accessibility challenges when using the e-learning tool?

4.3.3 Analysis

We audio recorded then transcribed all interviews. The researchers coded the transcriptions using qualitative coding based on constant comparative methods to find common themes across interviews, following the methods outlined by Saldana [156]. The coding process was iterative. Initially, two researchers coded three sample transcripts independently, then discussed the themes and categories together. Then one researcher coded the rest of the transcripts based on the agreed categories. After writing our initial draft and reflecting upon our findings, we repeated the coding process. In the second iteration, two researchers coded three sample transcripts independently, then discussed the themes and categories. In the rare cases when coders disagreed, they discussed the issue until they reached agreement. After that, they coded all transcripts together based on the agreed categories.

4.4 Findings

Website	Topics	Used by	Description
IXL Learning	Grades K-12	James, Darcy, Victor, and Ian (4 people)	**Features:** practice problems with text-based hints;**Input:** choosing the correct answer or typing the final answer.
Khan Academy	Grades K-12	Selena, Amy, Darcy, and Ian (4 people)	**Features:** practice problems with text-based hints, video lessons with a teacher giving a lecture and writing on a blackboard, and an annotation tool;**Input:** choosing the correct answer or typing the final answer.
CMP 3	Grades 3-6	James, Darcy, and Selena (3 people)	**Features:** practice problems without hints, short animations that explain and visualize math concepts, and an optional annotation tool;**Input:** typing the final answer.
Flocabulary	Math Vocabulary	Amy and Lisa(2 people)	**Features:** practice problems without hints and rap music videos that explain the meaning of math vocabulary;**Input:** choosing the correct answer.
BrainPOP	Calculations, Computations	Lisa (1 person)	**Features:** practice problems without hints and video lessons that explain math concepts using real-world examples; **Input:** choosing the correct answer.

Table 4.2: Information for the drill-and-practice websites used by participants including the math topics covered, pseudonyms of participants who used these tools, key features, and methods for student input.

Game	Topics	Used by	Description
Kahoot!	Designed by Users	Jennifer and Victor (2 people)	**Gameplay:** a text-based trivia game with single- and multi-player modes; in the multi-player version, separate groups of players compete against each other;**Input:** choosing the correct answer.
ST Math	Grades K-6	Amy and Marcus (2 people)	**Gameplay:** a single-player puzzle game that uses visual elements to present math problems and requires players to use virtual manipulatives to solve the puzzles;**Input:** dragging-and-dropping virtual manipulatives.
Prodigy	Grades 1-8	Lisa (1 person)	**Gameplay:** a single-player role-playing game that presents text-based math problems that players must answer correctly to defeat enemies;**Input:** choosing the correct answer or typing the final answer.
DareDasl	Arithmetic, Money	Darcy (1 person)	**Gameplay:** a single-player role-playing game that requires players to "drive" a vehicle and calculate the cost of driving from one place to another;**Input:** choosing the correct number of coins (manipulatives).
Jungle Math	Counting, Whole Numbers	Carol (1 person)	**Gameplay:** a single-player puzzle game that requires players to help a monkey collect the correct number of bananas that represents a given whole number;**Input:** choosing virtual bananas (manipulatives).
Slice Fractions	Fractions	Helen (1 person)	**Gameplay:** a single-player puzzle game that requires players to clear a path for a woolly mammoth by slicing up the objects in its way to form visual representations of fractions;**Input:** slicing a virtual manipulative into small pieces.

Table 4.3: Information for the educational digital games used by participants including the math topics covered, pseudonyms of participants who used these games, gameplay designs, and methods for student input.

4.4.1 Use of E-Learning Tools for Math Education

Participants used a total of 11 math e-learning tools, which included five drill-and-practice websites and six educational digital games. Interestingly, we found that our participants were not using any math e-learning tools that were specifically designed for students with SLDs (e.g., *Calcularis* [66]). Instead, participants used tools designed for general education to teach students with SLDs. We report our findings on participants' use of the different e-learning tools below.

Drill-and-Practice Websites. Table 4.2 shows key information for the five drill-and-practice websites used by our participants. These websites incorporated two categories of math content: (1) complete curricula (*IXL Learning, Khan Academy, ST Math, and CMP 3*) and (2) specific topics (*Flocabulary and BrainPOP*). Additionally, four of these websites provided math lessons in the form of videos (*Khan Academy, CMP 3, and BrainPOP*) or music videos (*Flocabulary*). Two websites also provided students with text-based hints (*IXL Learning and Khan Academy*), and two websites allowed students to write, draw, and annotate word problems (*Khan Academy and CMP 3*). Finally, the five websites used two modes of inputting answers: (1) selecting multiple-choice answers (*IXL Learning, Khan Academy, Flocabulary, and BrainPOP*) and (2) typing short answers (*IXL Learning, Khan Academy, and CMP 3*). Figure 4.1 shows two screenshots of *Khan Academy* and *IXL Learning*, which showcases that the drill-and-practice websites presented and received most math content in text form.

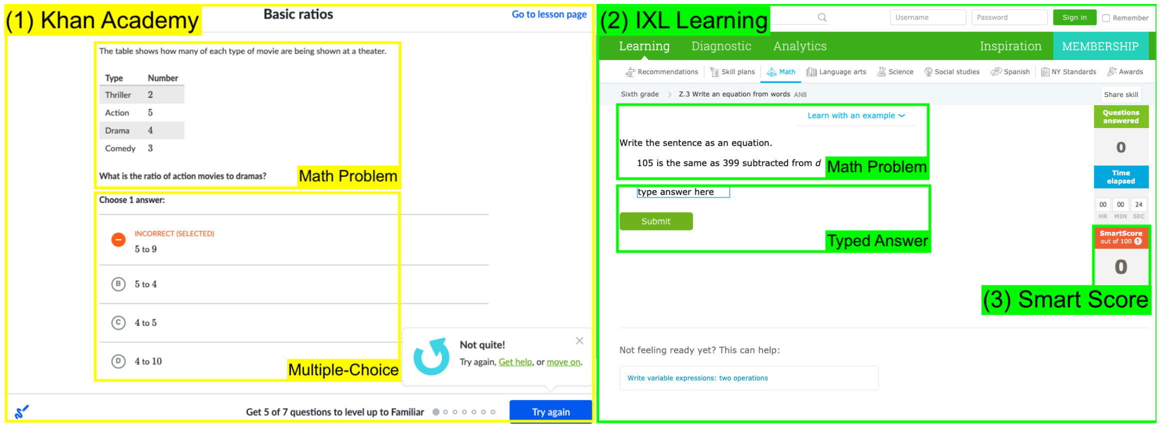

Figure 4.1: These are screenshots of two drill-and-practice websites: Khan Academy and IXL Learning. In Khan Academy (Highlight 1), the student needs to choose an answer to the math problem; In IXL Learning (Highlight 2), the student needs to type the answer. IXL Learning displays the student performance score in a Smart Score panel (Highlight 3).

Educational Digital Games. Table 4.3 shows key information about the six digital games used by our participants. Overall, these games incorporated three categories of math content: (1) user-designed content (*Kahoot!*), (2) complete curricula (*ST Math and Prodigy*), and (3) specific topics (*DareDash, Jungle Math, Slice Fractions*).

Additionally, these games supported three different gameplay styles: (1) trivia competition (*Kahoot!*), in which competitors are asked questions (2) role-playing game (*Prodigy and DareDash*), in which players assume the roles of characters in a fictional setting and (3) puzzle game (*ST Math, Jungle Math, and Slice Fractions*), in which players solve a math puzzle in each task.

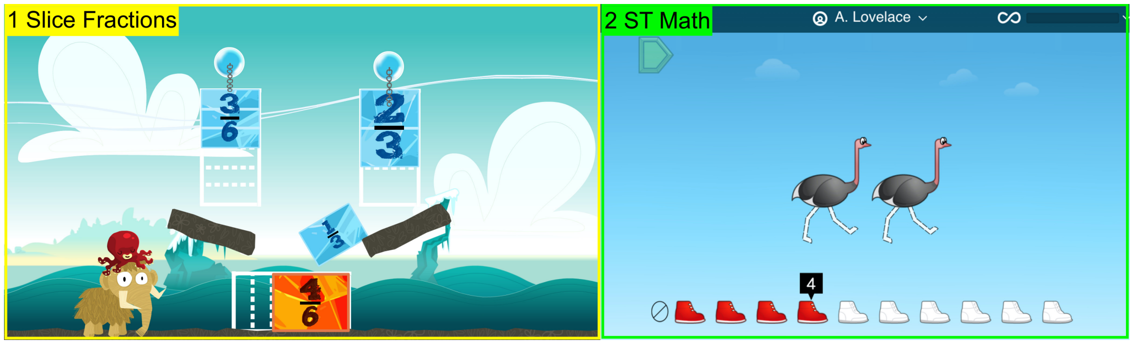

Figure 4.2: These are screenshots of two digital math games: Slice Fractions and ST Math. In Slice Fractions (Game 1), players slice 1/3 of the blue ice cube down to the ground to diminish the red lava cube that is blocking the wooly mammoth's way; In ST Math (Game 2), players select four shoes for two flamingos.

These games also used three different modes of input: (1) selecting multiple-choice answers (*Kahoot! and Prodigy*), (2) typing short answers (*Prodigy*), and (3) interacting with virtual manipulatives (*ST Math, DareDash, Jungle Math, and Slice Fractions*). Figure 4.2 shows how players use virtual manipulatives to solve math puzzles in two games, *Slice Fractions* and *ST Math*. In *Slice Fractions*, players slice the blue ice cube in the air in order to let part of the ice cube diminish the red lava cube then clear the pathway for the wooly mammoth. In *ST Math*, players learn multiplication by selecting the number of shoes that fit the number of flamingos so that all the flamingos can run away from the scene.

Out of all the games, *Kahoot!* was the only one that supported a multiplayer mode.

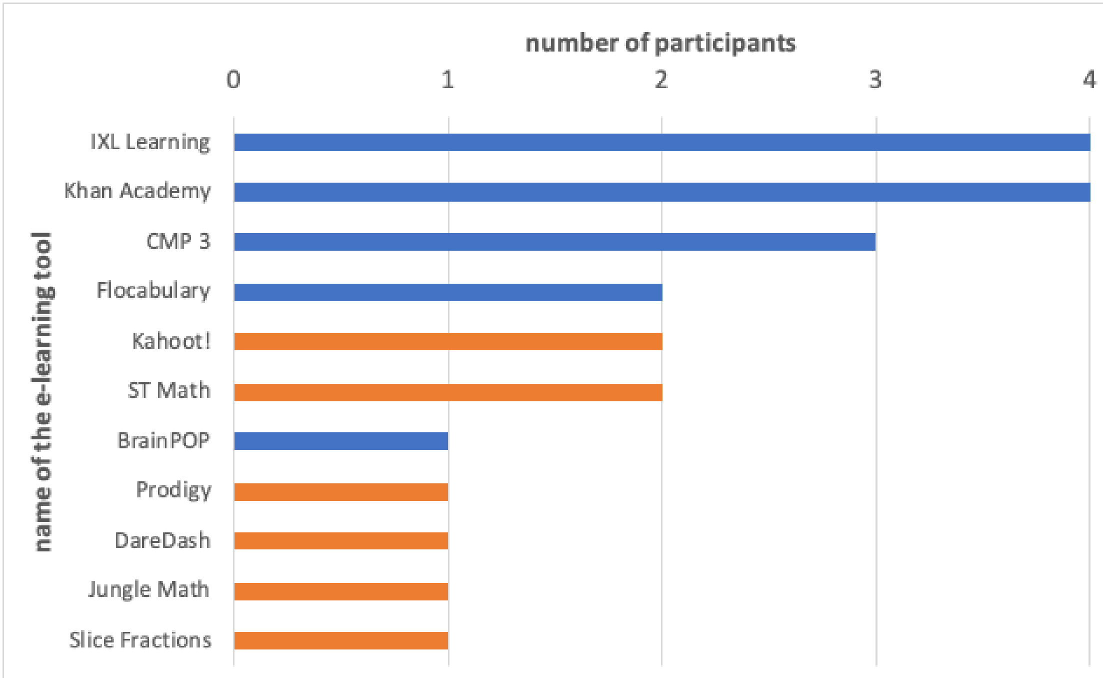

Figure 4.3: The number of participants who used a specific math e-learning tool. The drill-and-practice websites are in blue bars; the educational digital games are in orange bars.

In Figure 4.3, we show the number of participants who used each of the math e-learning tools listed in Table 4.2 and Table 4.3. More participants used drill-and-practice websites than educational games. *IXL Learning* and *Khan Academy* were the most commonly used websites while *ST Math* and *Kahoot!* were the most commonly used games. While most of the websites were used by multiple participants, the majority of the games were only used by one participant each.

Many participants used more than one math e-learning tool since they could not find one tool that fulfilled their multiple needs. For example, Victor and Amy used drill-and-practice websites to prepare students for their traditional math tests, but they needed digital games to reduce students' anxiety in practicing math exercises.

To manage and share resources from different math e-learning tools, four participants (Jennifer, Savannah, Selena, and Darcy) used learning management systems, such as Google Classroom. These systems allowed their students to follow URL links to find the resources or assignments supported by e-learning tools. However, participants did not ask their students to submit math work on learning management systems. Instead, students submitted their work directly to the e-learning tool or to the teacher in person.

4.4.2 Reasons for Using E-learning Tools

Participants used math e-learning tools for four reasons.

Providing Independent Practice in the Classroom. Some participants used the e-learning tools to give students independent math work, so that they could provide one-on-one instruction to other students. For example, Darcy set up *Khan Academy* accounts for every student so she could give them independent tasks to work on while she and her co-teacher provided extra support and attention to students who needed it most, including those with SLDs. As Darcy mentioned, "So we use *Khan Academy* with them. What we do is to have it set up in the class so I can give [students] assignments to work on. The special-ed teacher and I will walk around and help out. The kids with learning disabilities need a lot of support."

Motivating and Engaging Students. Participants observed that students with SLDs were more anxious and much less confident than their peers in solving math problems. As a result, these students would be reluctant to practice math exercises even if they had the ability to solve the problems. As Savannah suggested, "A lot of [students with SLDs] have this preconceived notion that they're not good at math or they don't like math." Therefore, participants used e-learning tools to make math more engaging and less intimidating.

The e-learning tool offered different sensory modalities to present content (e.g., in videos, in songs or using manipulatives). Victor used websites that provided video solutions to problems because "[students] don't want to read [the solution] to see what they did wrong." Amy also used the video lessons on many drill-and-practice websites for her students, and she added that, "[videos] have to be short and to the point because, you know, [students with SLDs] just have trouble focusing for long periods of time." In addition to using videos, Amy engaged her students by playing the songs in *Flocabulary* that explained the meaning of math vocabulary. Participants said their students were more engaged in solving math puzzles when they could use their hands to solve puzzles or play around with the manipulatives. Their students told participants that the digital games were fun to play.

The digital game also engaged students in collaborating and competing as a group. When Jennifer used *Kahoot!* for test review, she found that, "[My students] compete against each other, and the kids love it."

Assessment. James and Savannah reported that they needed to write the Individualized Education Program (IEP) for students with SLDs. An IEP is a legal document that states important information about a student's unique strengths, needs, and current levels of performance, as well as the special education services, supports, and accommodations that the student is entitled to. Without using e-learning tools, James found it difficult to determine at which grade level his students were: "If I do a pretest with [a student] on seventh grade material, [I can determine that a student is] not doing well, but we don't really have a process to go back and say like, 'Oh, you're at a fourth grade level.'" By using the *IXL Learning* diagnostic program, James was able to tell which exact level (e.g., which year and month) his students were at for specific math skills.

Tracking Student Performance. Participants reported that many students with SLDs needed more practice after class. Therefore, two participants (Savannah and Selena) would ask their students to work on math e-learning tools as homework. They chose e-learning tools over traditional homework formats so that they could more easily and efficiently track student performance. For example, Savannah sent students home with drills to do on *IXL Learning*. In this way, Savannah could keep track of "how much time they spend and what type of questions they were working on."

4.4.3 Challenges with Using E-Learning Tools

Participants described four critical challenges that they encountered when using math e-learning tools for students with SLDs.

Text-Intensive User Interfaces. Students with SLDs struggled with reading comprehension, which further affected their ability to solve math problems. Drill-and-practice websites required a lot of reading, with practice problems, hints, and solutions all presented in text form. This was "just like a string of steps and words" (Victor) to the students. They were intimidated and could not make sense of them. Victor found that his students randomly guessed the answers to questions because they did not want to read the words in the questions or in the hints. To help students with difficulty decoding text, three participants

(Savannah, Jennifer, and Marcus) used text-to-speech assistive technology. Although they found that students somewhat benefited from this, reading was only the first step of comprehending a problem. After decoding each word, students needed to determine which information in the problem was important and which math operations they needed to use.

When working with their students one-on-one, teachers read the problems together with their students and used annotations to help students decode the problem. For example, Lisa taught an annotation method called *CUBES* [76]. In this method, students annotated three key pieces of information: (1) **c**ircling math symbols like numbers; (2) **u**nderlining the sentence that asked the question; and (3) **b**oxing the math vocabulary that described the operations. By looking at these annotations, the student could **e**valuate what steps to take, and then **s**olve and check the answer. Using annotations was helpful for students with SLDs, because it could "slow [students] down to really know step-by-step what [the math operation] needs to be" (Carlo) and "help kids to see structural things" (Ian), which referred to the numbers needed in math operations and the math operations needed to produce the final answer. The e-learning tools provided little to no ability to annotate problems. For example, users could only highlight words in the math problem in yellow; they could not circle, underline, box, or draw arrows between the words.

Another common strategy participants employed when working with their students one-on-one was using manipulatives. Manipulatives are physical or virtual objects that students could directly interact with (in other words, manipulate) to understand and communicate math concepts. For example, Carol would translate word problems by using concrete, countable sticks to represent the numbers in the problem. Carol found that, "As [a student] was able to move [the sticks] around, he just caught [the concept]." However, only some of the games and none of the drill-and-practice website provided virtual manipulatives.

Participants also observed that students with SLDs had trouble typing their math work. Many websites required students to type their answers, which was cumbersome. Two participants (Lisa and Marcus) mentioned that many students with SLDs struggled with typing. Students also experienced challenges in figuring out how to communicate special math characters on the keyboard. As Ian said, "There are math characters like π , where's the π key?" Participants further suggested that it would be easier for these students if the e-learning tools allowed them to write or draw.

Insufficient Feedback about Student Performance. Participants reported that they kept track of students' performance and managed students' accounts through the teacher dashboards of the e-learning tools they used. These dashboards typically included summaries of students' performance on the practice problems: (1) aggregated data about the average number of questions a student solved, average time spent on practicing, etc., (2) individual students' problem solving records including their final answers, and (3) prediction of an individual student's math skills mastery level. Although this information helped identify students who made mistakes, participants could not tell from their dashboards *why* their students made those mistakes, because there was no way for students to show their work or explain their final answers. For example, Carol wanted to know how her students annotated word problems in *Khan Academy*, but only their final answers were reported.

Additionally, dashboards provided no way for teachers to know if their students had been guessing randomly. Participants observed that repeated failure in solving math problems would frustrate students with SLDs, leading them to guess answers randomly. Many participants wished the e-learning tools would alert them when their students started to randomly guess. Many e-learning tools, including both games and drill-and-practice websites, allowed students to repeatedly enter incorrect answers to the same question until they answered correctly. As a result, participants had to constantly monitor their students to determine whether they were actually using math reasoning and problem-solving skills, or just guessing randomly. For example, Carol had to sit with her students to make sure they were not guessing or getting frustrated from repeated mistakes. To address this problem, Lisa suggested that the e-learning tools show students messages that encourage them to ask teachers for help when they are unsure how to solve a problem or after they have repeatedly entered incorrect answers.

Inability to Adjust Difficulty Levels. Participants needed to work on more fundamental math skills with students with SLDs, reteaching many math skills that were at lower grade levels than the student's current grade. Three participants (Amy, Darcy, and Carol) emphasized the importance of adjusting difficulty levels of practice problems to adapt to the students' actual abilities.

However, participants could not lower the difficulty level of the e-learning tools, which frustrated the students. Carol remarked that "the kids throw iPads when they get frustrated or shut down." Similarly, Darcy found that when she couldn't find simpler problems on *Khan Academy* for students who needed to practice lower grade-level math, "[students] shut it down and they don't care."

Difficulties with Setup and Maintenance. Participants wanted to use e-learning tools in the classroom rather than assign their use as homework. As Selena said, "I mean homework is good because it's good practice but like is homework a lot of learning? No. I'm pushing a lot of learning in class. So, I want to know more about what kids are doing in class." However, they found it challenging to set up the tools for their students during the limited class time they had. For example, Selena mentioned that it took too long for students to log into the drill-and-practice websites because during her lessons, students only had 12 minutes or so to practice one math skill before they had to switch to another exercise. Marcus also mentioned that using computer-based tools was time consuming because of logging issues. Similarly, Savannah and Amy felt frustrated about not having enough time to set up games for their students. Amy explained: "I mean [students with SLDs] have so much skills that they need to work on, and I'm still supposed to teach the sixth-grade content with everyone else because we have to pass [the state test] at the same time." Participants were already stressed by the need to provide supplemental instruction to their students in addition to teaching the standard curriculum, and there was no time to spare for frustrating technical issues.

Other challenges participants faced included difficulties with the devices themselves. For example, Jennifer remarked, "We had tablets, but they all died. We had Kindles in our rooms [...] but they're like busted." Additionally, Selena explained that "There is a smaller set of 30 iPads, but people don't usually use it because it is hard to keep up. The apps need to be updated. And we have a tech guy, but he didn't do it." Participants did not receive support for resolving these various technical issues.

4.4.4 Needs and Challenges in Using Assistive Technology

As mentioned above, some participants helped their students set up and use text-to-speech to alleviate their reading challenges. Only one e-learning tool, *IXL Learning*, provided text-to-speech by default, so participants showed their students how to use built-in text-to-speech services on their devices. For example, Savannah enabled the text-to-speech accessibility feature on Google Chromebooks for her students. Similarly, Jennifer and Marcus told us that their students used the text-to-speech feature built into iOS.

In addition to text-to-speech, Darcy used closed captioning when her students watched video lessons on *Khan Academy*. She said that closed captioning was helpful because many of her students with SLDs had trouble with auditory processing. Closed captioning allowed them to use visual cues from the captions to support their understanding of the spoken content in the videos.

Participants were unaware of standalone assistive technologies in general. They were also less interested in using these technologies because they would have to add standalone assistive technology tools to a student's Individualized Education Program (IEP), a time-consuming bureaucratic hurdle. Savannah mentioned that "One student be like wanting something for [an assistive technology tool]. It took us like a whole year." Rather than having to go through the IEP process, participants preferred assistive technology that would be available by default.

Another reason to use incorporated assistive technology was to avoid stigma. Jennifer told us that, "Because it's middle school, a lot of [students] are embarrassed about [using a special tool or technology]. They don't want to be different in front of their peers." Savannah shared similar comments, "They're very conscious at this age about every little thing."

4.5 Discussion

Our study answered our three research questions: (1) what math e-learning tools are teachers using, (2) why and how are teachers using these tools; and (3) how do teachers perceive the effectiveness of these tools for students with SLDs. To summarize, we found that each participant used at least one e-learning tool with their students, none of which were specifically designed for students with SLDs. Most participants used educational games. They primarily used these tools to let students practice math problems independently in the classroom and make the work more engaging. In other cases, three participants also used e-learning tools to assess their students' abilities and track their performance in math drill exercises after class. The effectiveness of the tools was hindered by usability problems experienced by both the teachers and the students. The text-intensive interfaces posed major barriers for the students who had difficulty processing language. Meanwhile, the teachers did not receive enough feedback about student performance to be able to adjust their teaching to the students' needs. Teachers were not able to adjust the difficulty level of math exercise on e-learning tools to accommodate students' wide range of ability-levels either. Additionally, mundane setup challenges of the hardware and software made the tools difficult to use in a classroom full of students. To alleviate some of the challenges that students faced, participants encouraged their students to use assistive technology like text-to-speech when using the math tools, but they faced challenges in getting and using the assistive technology.

As e-learning tools become widely used in classrooms, designers have an opportunity and responsibility to also make the e-learning environment more inclusive. Therefore, based on our findings, we discuss design implications for the creation of more inclusive and effective math e-learning tools.

4.5.1 Teacher-Oriented Design Implications

Our study suggested that teachers faced difficulties in using e-learning tools to teach. To help improve the usability of e-learning tools for teachers, we present five design implications, which range from short-term design suggestions to longer-term research directions.

Involve special education teachers in the design of e-learning tools. Existing research on math e-learning tools for students with SLDs (e.g., [27, 95, 102, 137, 145, 150]) rarely included teachers in the design and evaluation process. Our study revealed that teachers for students with SLDs wanted to use math e-learning tools in teaching. However, they faced critical challenges in using the e-learning tools, some of which were related to students' learning disabilities. Nor could they ascertain what confused and even frustrated their students during independent practice. Because students learned most when they studied with their teachers, if teachers could not adopt these e-learning tools in their teaching, students would be unlikely to benefit from the use of e-learning tools. Although recent work started to investigate how to help teachers adopt math e-learning tools [87, 86], it was limited to helping teachers in general education. Future work on designing e-learning tools should also involve special education teachers.

Use manipulatives to present math concepts. According to our study, teachers used manipulatives to explain complex and hard math problems (e.g., word problems) to students with SLDs. However, none of the existing software applications (e.g., web-based [167, 184] and on the iOS system [49]) have combined manipulatives with math problems. The future design of e-learning tools should consider incorporating a set of commonly used manipulatives (e.g., blocks and fraction tiles) and allowing teachers to use the manipulatives to translate the math problem for their students.

Detect and report students' frustration to teachers. According to our study, teachers needed to keep students actively solving math problems on the e-learning tool. However, if the problem was too difficult, students would get frustrated and then pretend to be working on the e-learning tool. As Jennifer suggested, although her students found the games engaging, they did not ask for Jennifer's help when they got stuck on a math problem. Instead, "when students [got] frustrated, [they] started randomly clicking on [things] like 'options'" (Jennifer). One of the plausible reasons was that students might not be willing to let teachers know, possibly out of embarrassment, that they could not solve it like their classmates did. Researchers who studied people with other disabilities using digital tools also discovered this phenomenon [210]. In such situations, teachers needed to understand the root causes of students' frustration and challenges, even if the students did not verbalize them. Therefore, in the future, we should consider different ways to detect the students' frustration and report it to the teachers to help them better track the students' emotional status. One way to do that is by using AI to passively detect a student's frustration level by analyzing the pattern of students' inputs. For example, in order to identify a student's different cognitive-affective states including frustration, Baker et al. developed deep learning technology to analyze the student's mouse activities and facial expressions [110, 84]. D'Mello et al. proposed to analyze the eye gazing data to predict whether a student can no longer actively working on the exercises [70, 67]. Another way is to design friendly user-interfaces to encourage students to share their emotional status to their teachers.

Allow teachers to fine-tune difficulty levels. Based on our findings, teachers sometimes needed to teach students with SLDs math skills at lower grade levels. Prior work [180, 77] also found that differentiating math work for students with SLDs was an effective teaching strategy. However, teachers were not able to assign easier problems for their students on many math e-learning tools. Designers should give teachers the ability to adjust the difficulty levels of problems on the e-learning tool. Recently, learning analytics researchers developed AI algorithms [83, 96] to predict a suitable difficulty level for individual students. Researchers should explore using such algorithms to assist teachers in personalizing the e-learning tools.

Enable quick and easy set up for classroom use. Our study found that teachers wanted to provide independent exercises in class. However, teachers did not have enough time to set up the e-learning tools. According to our study, teachers often wasted too much time on helping students to log in. For example, they needed to help students who forgot their passwords. Our study also suggested that teachers needed to teach multiple types of math skills in one lesson. Thus, they needed to switch between the types of math skill training after every 10 minutes during the class. Based on our discussion, designers should design a classroom mode, which has a quick login process and allows teachers to pre-assign multiple 10-minute short math exercises in the students' e-learning tool account before the class begins.

4.5.2 Student-Oriented Design Implications

In our study, participants reported usability and emotional issues that students with SLDs experienced when using the math e-learning tools with and without assistive technology. To address these issues, we distil five design implications.

Provide a flexible annotation toolkit for word problems. According to our study, teachers taught students with SLDs to reduce mistakes in comprehending word problems by using annotation skills such as CUBES [76]. This was a very powerful skill because it allowed students to slow down in order to find all math symbols and math vocabulary and then see their relationships. While some of the e-learning tools, such as *Khan Academy*, had a simple annotation function, teachers found such features too limited. Designers should implement a flexible annotation toolkit that allows students to circle, underline, and box words, to draw arrows between words, and to highlight words in different colors.

Incorporate game design elements into drill-and-practice websites. Prior work [120] suggested that there were two distinct approaches for designing e-learning tools: drill-and-practice training and game-based training. Teachers found that these two types of tools had their own pros and cons for students with SLDs. On the one hand, game-based learning let students have fun when solving math problems by playing around with virtual manipulatives or working in groups. On the other hand, drill-and-practice websites could prepare students for their exams in solving traditional types of math problems. The current drill-and-practice websites did not incorporate any game design elements that our participants praised in the interviews. As Selena suggested, there were too many kinds of e-learning tools for her to manage. She wished she could have had one tool that helped students prepare for their traditional math tests in an enjoyable way. Therefore, future research should combine the advantages of drill-and-practice training and game-based training in one tool.

Design AR manipulatives for math education. Our study showed that teachers used physical manipulatives frequently in teaching students with SLDs. While physical manipulatives are useful tools for students with SLDs, past research has highlighted that students depend on their teachers to learn how to use the manipulatives to solve math problems [128, 203]. Therefore, we propose using AR technology to enable students with SLDs to use physical manipulatives to solve math problems when their teachers are unavailable. Researchers have used AR technology to detect students' interactions with physical models and to teach students by combining digital information with the physical models [161, 166, 74]. For example, *Talkit* [166] enabled students with visual impairments to listen to information about components of a 3D printed model by interacting with it. Nevertheless, none of the AR-learning tools were designed for students with SLDs in math education. Future work may explore this approach to help students with SLDs use physical manipulatives to solve math problems.

Avoid using words that exacerbate students' preconceived notions of being bad at math. Our study found that students with SLDs were more anxious and much less confident than their peers in solving math problems. While we reviewed the design of websites and games, we found that some e-learning tools might deliver the unintended message to students that they are not good at math. For example, *IXL Learning* used “smart scores” (as shown in Figure 1 (Highlight 3)) to communicate the student’s progress towards completion of the assignment. However, students who received low scores might internalize the message from the website that they were not smart. Therefore, we suggest that designers should carefully consider the words that are used on the dashboard and in the point-and-reward system. A design suggestion given by O’Rourke et al. [140] is that visualizing the progress of earning new points can encourage students to develop a growth mindset in exercises.

Provide assistive technology in e-learning tools for general education. Shinohara et al. pointed out that social acceptability was important in designing assistive technology [168]. We learned from our study that middle school students felt embarrassed when they were using a stand-alone assistive technology tool because the tool revealed their disabilities in learning to their peers. This finding aligned with a Norwegian interview study with teenagers with visual impairments [172]. The teenagers rejected assistive technology if the technology made them look less capable than their peers and wanted mainstream technology to include built-in accessibility. This collective evidence likely explains why stand-alone assistive technologies for digital math content [31, 72, 174] have not been broadly adopted. Therefore, we suggest that, whenever possible, assistive technology should be directly integrated into mainstream e-learning tools. For example, according to our study, students should have the option to listen to a word problem. We were glad to find that *IXL Learning* implemented a text-to-speech feature (only for K-3), but designers should follow this practice moving forward.

4.5.3 Limitations and Future Work

Our study was limited by our convenience sampling method and the relatively small sample size of 12 teachers. While such methods are standard in the field of human-computer interaction, we did not include perspectives from teachers in private schools and other school systems, nor did it represent teachers in most states in the US and other countries. Our study participants only came from three states in the US and taught public school students in fifth through eighth grade.

That being said, we did reach saturation, finding a convergence of participants' opinions regarding use of recent math e-learning tools for students with SLDs. By analyzing the shared background and characteristics of our study participants, we would suggest that they may represent the group of teachers who teach tens of students from middle- and low-income families each year. They often need to work with parents who do not have the ability and resource to tutor their kids after school. To help every student make learning progress even after class, these teachers are happy to try and use different e-learning tools.

As with qualitative studies of this type, our findings should not be used to generalize patterns across all teachers or students, but rather to gain insight into current use patterns and shed light into future design and research directions. Researchers should interview more teachers from various locations (e.g., more US states and other countries) to provide more generalizable results.

Finally, in this study, we focused on teachers, who determine whether and how e-learning tools are used in practice. As seen from our findings, they also have keen insights into student experiences and pedagogical strategies. However, students are, of course, the primary users of e-learning tools and their first-hand experiences must be investigated. The parents of these students also observe and are often involved in after-school educational experiences and their perspectives merit consideration as well. As we move forward with our research, we plan to study the experiences of these two important stakeholder groups.

4.6 Conclusion

To understand the role of e-learning tools in supporting students with SLDs, we conducted semi-structured interviews with 12 US-based teachers who taught math to students with SLDs in grades 5-8. Our study revealed that recent math e-learning tools are not sufficiently effective or inclusive for the large number of students with SLDs. We have provided design implications and advocate future work on math e-learning tools to expand the target users from students to teachers-of-students. We hope that, by meeting both teachers' and students' needs, we can accelerate the adoption of math e-learning tools in the inclusive educational model. As a result, many more students can improve their math skills and engagement when they are using the math e-learning tools.

CHAPTER 5

AN INTELLIGENT MATH E-TUTORING SYSTEM FOR STUDENTS WITH SPECIFIC LEARNING DISABILITIES

5.1 Introduction and Background

Learning math is difficult. To help students learn math when they cannot get a tutor, many e-learning tools (e.g., Khan Academy [101], ST Math [125]) have been developed to provide independent math practice exercises. However, the commonly used math e-learning tools were not designed for students with specific learning disabilities (SLDs). As a result, our prior research study [195] showed that these students had difficulties in using math e-learning tools. For example, the students struggle with reading problems, hints, and solutions that are presented in text form. The students may also struggle when they have to solve a problem that is beyond their math abilities. As the students struggle in practicing math exercises, they build up feelings of frustration and irritation, which leads up to exhibiting negative emotional behaviors (e.g., randomly guessing answers or even damaging the e-learning device).

Recent research on math e-learning tools for students with SLDs [17, 95, 150] has mainly focused on helping students overcome text processing difficulty by designing interactable manipulatives that can represent text-based math problems. For example, Calcularis [95] challenges students to compare the size of two quantities by positioning two numbers on a number line from 0 to 100. French National Institute of Health and Medical Research designed two games, The Number Race [204] and Number Catcher [4], for the students. These games present Arabic, verbal, and visual representations of numbers together to train students to compare which whole number is larger and which is smaller. However, if students struggle in using these e-learning tools, they can only receive two types of help: feedback for wrong answers or step-by-step hints. These types of help are not adequate to help students with SLDs, as mentioned in Chapter 4.

The interviews of teachers for SLDs [196] suggested that teachers would like to have a smarter virtual tutor who not only cares about the student's correctness rate but also the student's emotional state. Otherwise, without the teacher's active monitoring and in-time interventions, students with SLDs were more likely to give up working on the math e-learning tool when they entered a negative emotional state (e.g., getting frustrated or angry).

Although prior work on affect detection has proposed designs of intelligent tutoring systems for student emotional self-management [25, 81], their work did not consider the needs of students with SLDs. Their systems assumed that students are willing to ask the system for help and that students can recover from frustration by learning why they are wrong. As we mentioned in Chapter 4, none of these assumptions can be applied to students with SLDs. As a result, we do not know how we should design an intelligent e-tutoring system for students with SLDs to reduce their negative emotional behaviors.

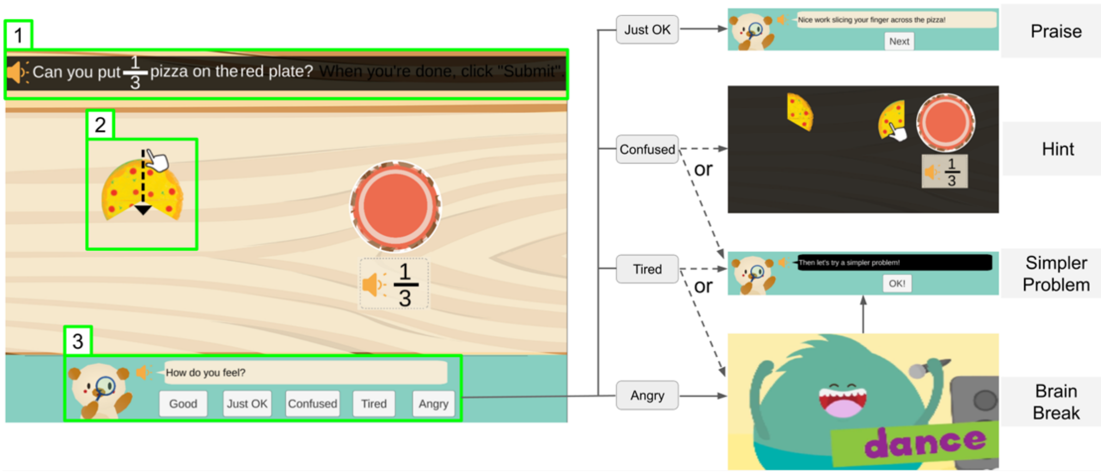

Figure 5.1: An overview of the intelligent math e-tutoring system for students with SLDs. Zone 1 presents the problem in text from (and highlights the sentence that the system is reading). Zone 2 illustrates how students slice their finger across the pizza to make half pizza slices. Zone 3 presents the dialogue system that asks the students about their emotional state.

Therefore, to form our design, we worked together with a teacher for students with SLDs to design an intelligent e-tutoring system (as shown in Figure 5.1) that reduces students' negative emotional behaviors. Our system automatically detects three negative emotional behaviors: (1) staring at the problem without trying, (2) touching the screen agitatedly, and (3) being distracted away from the screen. Our system uses eye-tracking data, inputs on a touchscreen, and response time to model student behaviors. If our system detects that a student exhibits one of these three behaviors, the system identifies that it is time to confirm the student's emotional state through dialogue. Because both teachers and systems are just speculating about the negative feelings of the students, so it is necessary to confirm with the students before deciding how to intervene.

We designed four methods to intervene when students is getting upset. To prevent students from getting upset when they don't have the ability to solve the problem, our system provides hints or to switches to a simpler problem.

To engage students who have been experiencing negative emotions (i.e., just ok, tired, or angry), our system sends personalized encouraging messages or provides brain breaks.

We also conducted a formative study with five teachers for students with SLDs. During the studies, we demonstrated our design prototype to the participants. We asked them to provide feedback about the system design based on their experience in tutoring students with SLDs. The teachers thought it was necessary for e-learning tools to check in with students' emotional status. They reported that our design of e-tutoring intervention methods would effectively reduce the students' negative emotional behaviors. Teachers also suggested that we should personalize the detection of negative emotional behaviors for students who have much severe learning disabilities. We present an intelligent e-tutoring system for students with SLDs that can reduce the students' negative emotional behaviors. We designed this system together with teachers for SLDs and conducted formative study with teachers. In the future, we will conduct studies with students to get feedback from them. Our work will inspire e-learning tool designers and developers to design inclusive e-learning tools for students with SLDs.

5.2 System Design

5.2.1 System Prototype

We developed a math game prototype that uses the intelligent e-tutoring system to manage students' negative emotional behaviors. This prototype trains students in solving fraction problems. Students practice fraction skills by manipulating pizza in two ways: cutting a pizza into different slices and moving the correct amount of pizza slices onto different plates (as shown in Figure 5.1, Highlight Zone 2).

The prototype runs in Windows 10 operating system on a Windows Surface Pro 7 tablet. Students use the touchscreen to play the game. A Tobii Eye Tracker 5 is stuck to the bottom of the tablet to detect where the student gazes at the screen. The game prototype processes the eye-gazing data stream to identify the in-game elements that the student gazes at, including texts, fraction numbers, pizza slices, plates, and UI buttons. The game prototype also logs the running time and touchscreen inputs. All the student data are used for negative emotional behavior analysis.

5.2.2 Triggering Intervention in Outstanding Behaviors

Many students with SLDs are reluctant to seek help when they were exhausted or agitated, but they show specific negative emotional behaviors [195]. Therefore, our system actively detects three types of negative emotional behaviors. First, students are distracted away from the screen if their eyes have not gazed on the screen for more than one minute. Second, students are touching the screen agitatedly if they repeatedly pressed the screen more than three times within 0.1 second. Third, students are hesitating to solve the puzzle if they have not touched the screen for more than two minutes and they have spent more than 70% of the time on gazing UI buttons and interactable but useless manipulatives. After detecting that students are exhibiting one of the three behaviors, the system will confirm students' emotional state through dialogue.

5.2.3 Interventions for Learning Difficulties

Students with SLDs build up frustration and irritation in using math e-learning tools for two main reasons [195]. One main reason is that students have to solve a math problem that is too difficult to solve. The other reason is that students cannot manage their negative emotion when they feel exhausted or angry.

To help students who do not know how to solve a problem, we designed two methods: providing hints or switching to a simpler problem. Providing hints is a traditional method to point the students in the direction of the correct answer. Nevertheless, if the students cannot comprehend the hint due to weak math abilities, switching to a simpler problem that matches their math ability is more useful than providing new hints.

Reducing the difficulty level of exercises, however, may not help students recover from exhaustion or anger. Many students with SLDs, due to cognitive differences, feel exhausted or angry quicker than general education students [35]. Inspired by the guidelines for tutoring students with SLDs [154, 187], we designed another two methods to help the students maintain or restore a good emotion: praising for a correct problem-solving behavior or providing brain breaks. The behavior-specific praise describes the approval of correct student inputs (e.g., "Nice work slicing your finger across the whole pizza! You made half pizza slices!"). The genuine approval is more effective in encouraging students with SLDs than generic encouragement (e.g., "you are doing awesome!") [187]. Nevertheless, students would need brain breaks instead of encouragement when they reach the tipping point of negative emotions [2]. A brain break is a break from the current learning task that students are working on. Students can rest their eyes or take a physical exercise during a brain break. Therefore, we designed a kind of brain break that temporarily hides all e-learning interfaces and guides students to play a kinesthetic game (i.e., a body dancing game that gets the student moving and grooving).

We mapped the e-tutoring interventions in exercises to three negative emotional states (just OK, tired, and angry). If the students feel just ok, they will receive behavior-specific praise. If the students feel confused, they may choose to get a hint or to try a simpler problem. If the students feel tired, they can try a simpler problem or take a brain break. If the students feel angry, they will take a brain break then continue to a simpler problem.

5.3 Formative Study: Teacher Feedback

To form our design, we conducted a formative study with teachers for students with SLDs (3 females, 2 males, between 0.5 and 7 years of experience in teaching students with SLDs in grades 3-6). We started from interviewing teachers because they have been designing and practicing interventions in tutoring the students. We anticipated that teachers would be able to use their teaching experience to offer valuate insights in our current design. During the study, we demonstrated our prototype to the participants. We asked them to give feedback on the design of e-tutoring intervention methods and the specifications to detect negative emotional behaviors. We present two main themes that were summarized from the teacher feedback.

All the teachers mentioned that e-learning tools should intervene in exercises when the student is experiencing negative emotions. One teacher explained, "I know how important the emotional status affects these students' ability to learn. Like, if they are frustrated, or if they are stressed, their thinking brain is not [turned] on." They liked the design of our e-tutoring intervention methods, especially providing brain breaks for students with SLDs. One teacher said, "there are kids who might be on a computer for a little while and they might get, you know, overwhelmed or zoned-out [...] For example, I have a kid who works for like 10 or 15 minutes and then gets to break. That's part of their IEP [Individualized Education Program]." Therefore, our design provides an important accommodation for the students as they practice math exercises independently.

Three teachers commented that the detection of negative emotional behaviors may not work for students who have much severe learning disabilities. These students may take significantly longer time to decode information and comprehend the problem. As a result, the current system may initiate chats with these students too early, and thus interrupt the students' thinking. To detect whether students are struggling, teachers had been providing the students both non-math exercises and math exercises to compare the differences of student performance (e.g., response time and body gestures). Teachers will intervene when the students behave much slower or agitatedly in math exercises. However, there is no known way to automate this detection approach.

5.4 Discussion: Increasing Algorithmic Fairness in ITS

Prior work on affect detection was designed for general education students [26, 25, 81, 63]. Although prior work claimed that they had successfully detected disengaged behaviors and affect, we did not know whether they would work for students with SLDs. As suggested by the recent study on algorithmic fairness in education [103], most educational e-systems designed for the general students would disable disadvantaged students including students with disabilities.

To overcome the unfairness that an algorithmic system like affect detection would introduce to students, Kizilcec et al. emphasized that researchers need to consider the needs of disadvantaged students when designing the system in three phases [103]. The three phases are measurement, model learning, and action. Our work focused on designing for the action phase of affect detection for students with SLDs. Different from prior system designs [25, 81], our system is aware of actions related to SLDs and communicates with the student about why the system believes it should initiate help. This design follows the design guidelines for algorithmic interfaces to improve fairness, as summarized in Kizilcec et al.'s paper [103].

The teacher feedback from the formative study shows that teachers appreciated how we designed for the action step to meet the needs of students with SLDs. Teachers also suggested that we should further explore the design of measurement and model learning so that students with more severe learning disabilities can benefit from the system too. Therefore, we can tell from our work that the design guidelines for improving algorithmic fairness in education are potentially useful and that future system design work should also try to follow the guidelines.

5.5 Conclusion and Future Work

We presented the design of an intelligent math e-tutoring system for students with SLDs. The system provides four types of intervention methods to reduce negative emotional behaviors. It automatically intervenes in exercises by detecting students' negative emotional behaviors then asking the students about their emotional state to eventually determine which intervention method to use. Teachers in the formative study mentioned that the system meets the needs of students with SLDs. They also suggested that the detection of negative emotional behaviors can be personalized to help students with more severe learning disabilities. In the future, we will explore the personalization of negative emotional behaviors detection. We will also recruit students with SLDs to try our prototype and to provide feedback. Our work provides a starting point for e-learning tool designers and developers to leverage our design to develop e-tutoring systems for students with SLDs.

CHAPTER 6

DISCUSSION AND CONCLUSION

6.1 Contributions

My doctoral dissertation answered three main research questions: (1) how to design educational video games that provide a situational context for skills training; (2) whether and how learners with SLDs have difficulties in using the existing e-learning tools to practice math skills; and (3) how should we design the e-learning tool to intelligently tutor learners with SLDs.

To answer the first research question, I proposed two gameplay designs that can provide situational context for game-based training. The first gameplay design is a role-playing game that can simulate the situation in which the learner builds business by processing contracts and avoiding occasional phishing attacks. I adapted the document inspection game mechanism in *Papers, Please* to the anti-phishing context of processing business emails to foster motivation for learning email safety rules. The second gameplay design is an AR game that leverages computer vision technology to analyze the learner's immediate vicinity then use recognizable objects to provide a vocabulary scavenger hunt game. To make this computer vision-based AR game playable, I developed a time smoothing method to reduce the computer vision detection errors and I designed a selection highlight and confirm interface to guide the learner to "select" a physical object that they want and is also recognizable by computer vision.

In addition to research how to embed realistic context into educational video games, the other contribution of my dissertation is to resolve the issues that learners with SLDs face in using e-learning tools to practice math skills. To achieve this goal, I started from interviewing teachers for SLDs to study the pros and cons of the existing e-learning tools used by learners with SLDs. My study found that educational video games were user-friendly for learners with SLDs because the games replaced text-based information with visual information that was also interactable. Nonetheless, the existing games could not help the learners self-manage their learning anxiety and negative emotional behaviors, which has been a big issue for the learners. Based on the needs of learners with SLDs, I designed an intelligent tutoring system that detects and mitigates learners' negative emotional behaviours. This system especially analyzes eye-gazing data together with other traditional input data to detect learners' negative emotional behaviors. I also designed four intervention methods to mitigate the negative emotional behaviors. Two out of four intervention methods, praising learners and providing brain breaks, were designed for learners with SLDs because they need these interventions to take them out of the anxious mood.

Overall, during my Ph.D. study, I sought to make drill-based training fun and effective, not only for general learners but also for learners who have SLDs.

6.2 Limitations

In this section, I discuss two major limitations in my research.

The user studies for *What.Hack* and *CollectiAR* were conducted in an ideal lab environment. Study participants only had one or two chances to play the game within a short period of time (20-40 minutes). However, one of the biggest characteristics for drill-based training is to be persistent in repeated training. Therefore, although the lab studies' results were positive, they did not indicate that people will take the initiative to play educational video games from time to time. In addition, some usability issues and other problems may not appear until the games are deployed in the real world. The anti-phishing training game, *What.Hack*, has been adopted by a large company. The company's report showed that more than 870 employees continuously played the game, which resulted in 90% drops in phishing offenses compared with the traditional training group. However, *CollectiAR* game for second-language learners and the intelligent fraction game for learners with SLDs were not publicly released yet. In the future, I will deploy these games in the app store and to test their effectiveness and usability issues by collecting and analyzing in-game data from hundreds of online players.

The current design of the intelligent tutoring system for student emotion management only detects negative emotional behaviors using a general method. However, by nature, learners' ability to process information is hugely different; and their ways to express anxiety or depression may be different, too. Therefore, this current detection method needs to be personalizable. To achieve this goal, I need to first deploy the version that uses general detection method, so that I can collect student data, which will be used to help us design machine learning technology that can personalize the negative emotional behavior detection.

6.3 Future Directions

I envision that educational video games, in the future, will be available for any type of skill training for both formal and informal learning. My dissertation shows that we can design an effective game for anti-phishing training, which is for informal learning; it also shows that we can design effective educational video games for early language and math skills training, which is needed by formal education.

As long as the skill is used to solve a complex task in real life, then we will be able to design new gameplay or adapt an existing gameplay design to create a fun and effective game to train this skill. Educational game designers should design the core game mechanism based on simulating all factors that can replicate the situation similar to what learners would experience when they need to use the skills to solve problems. To design an effective educational video game for situated learning, researchers should conduct interview studies with the instructors and learners to summarize all critical features that should be replicated in the game. Then the next biggest research challenge will be making the critical features part of the gameplay mechanism so that the replication is not boring. Although I did not find a universal solution for all simulation training programs, I believe that my role-playing attack simulation game design framework would work for training that improves people's ability to process complex language-based information (e.g., fake news detection).

In addition, I envision that future educational video games are accessible for all types of learners because they have an intelligent tutoring AI that not only predicts the learner's current skill level but can also recognize that learners may have learning difficulties including language processing difficulties and emotional management difficulties so that the educational video games can automatically offer personalized help including providing assistive technology or emotional support. This is part of the whole solution for developing a fairer intelligent tutoring system for special education students.

For special education students who have SLDs, I believe that future work on educational video games and intelligent tutoring systems should study how to help the learners learn how to self-manage their negative emotions during drill and practice. My current work shows that to achieve this goal, the system needs to design both the detection algorithms and the intervention methods. My work has proposed a general scheme for the majority of learners with SLDs, but it may not work for learners who have more severe learning disabilities. Therefore, future research should focus on making the system personalizable for both detection algorithms and intervention methods. For example, the system may provide calm brain breaks as a breathing game for learners who have both SLDs and ADHD, but provide active brain breaks such as a dancing game for learners who have SLDs but not ADHD.

6.4 Conclusion

People have been trying to promote learning by playing educational video games. Nonetheless, people enjoyed playing educational video games but had difficulties to apply what they have learned in the game in practice. Therefore, I tackled this problem by stretching the design space of gameplay to create situational and authentic context for training. I also studied the needs of learners with SLDs in using e-learning tools and I designed a new intelligent tutoring system that detects and mitigates these students' negative emotional behaviors when playing educational video games. As a result, my dissertation takes one step further towards the ultimate goal that educational video games are accessible and effective for any types of skill training for both formal and informal learning. In the future, I will continue exploring the design of new types of educational video games and intelligent tutoring systems that can provide enjoyable and useful learning experiences.

BIBLIOGRAPHY

[1] Anki - powerful, intelligent flashcards. https://apps.ankiweb.net/. Accessed: 2018-10-03.

[2] Brain breaks for kids — understood - for learning and thinking differences. https://www.understood.org/en/school-learning/learning-at-home/homework-study-skills/brain-breaks-what-you-need-to-know. Accessed: 2020-09-20.

[3] Fluentu: Language immersion online — learn a language with videos. https://www.fluentu.com/en/. Accessed: 2018-10-03.

[4] The number catcher - teaching arithmetic and treatment of dyscalculia. http://www.thenumbercatcher.com/nc/home.php. Accessed: 2019-09-20.

[5] Official rosetta stone - language learning - learn a language. https://www.rosettastone.com/. Accessed: 2018-10-03.

[6] Qualtrics: The leading research & experience software. https://www.qualtrics.com/. Accessed: 2018-10-03.

[7] Quizlet: Learning tools & flashcards, for free. https://quizlet.com/en-gb/help/2762234/studying-with-learn. Accessed: 2018-10-03.

[8] The carnegie cyber academy - an online safety site and games for kids, 2007.

[9] Mavi interactive. agent surefire. http://maviinteractive.com/mavi_products.asp, 2008. Accessed: 2018-09-20.

[10] Cyber security challenge uk Cyphinx. https://www.cybersecuritychallenge.org.uk/competitions/play-demand-cyphinx, 2016. Accessed: 2018-09-20.

[11] Game sets sights on fake news. https://www.american.edu/soc/news/fake-news-game.cfm, 2017. Accessed: 2018-09-20.

[12] The federal bureau of investigations, "Kids Games.". https://archives.fbi.gov/archives/fun-games/kids/kids-games, 2018. Accessed: 2018-09-20.

[13] Information assurane support environment Cyber Protect. https://iatraining.disa.mil/eta/cyber-protect/launchcontent.html, 2018. Accessed: 2018-09-20.

[14] Information security office cmu "anti-phihsing phil.". https://www.cmu.edu/iso/aware/phil/index.html, 2018. Accessed: 2018-09-20.

[15] Onguardonline. https://www.onguardonline.gov/media, 2018. Accessed: 2018-09-20.

[16] Phishline training. https://www.phishline.com/complimentary-content/, 2018. Accessed: 2018-09-20.

[17] Jugnu Agrawal and Lisa L. Morin. Evidence-Based Practices: Applications of Concrete Representational Abstract Framework across Math Concepts for Students with Mathematics Disabilities. *Learning Disabilities Research and Practice*, 31(1):34–44, 2016.

[18] Ammar Almomani, Brij B Gupta, Samer Atawneh, Andrew Meulenberg, and Eman Almomani. A survey of phishing email filtering techniques. *IEEE communications surveys & tutorials*, 15(4):2070–2090, 2013.

[19] Erik Andersen, Eleanor O'Rourke, Yun-En Liu, Rich Snider, Jeff Lowdermilk, David Truong, Seth Cooper, and Zoran Popovic. The impact of tutorials on games of varying complexity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 59–68. ACM, 2012.

[20] John R Anderson, Lynne M Reder, and Herbert A Simon. Situated learning and education. *Educational researcher*, 25(4):5–11, 1996.

[21] Nalin Asanka Gamagedara Arachchilage and Steve Love. A game design framework for avoiding phishing attacks. *Computers in Human Behavior*, 29(3):706–714, 2013.

[22] Nalin Asanka Gamagedara Arachchilage, Steve Love, and Konstantin Beznosov. Phishing threat avoidance behaviour: An empirical investigation. *Computers in Human Behavior*, 60, 2016.

[23] Ian Arawjo, Cheng-Yao Wang, Andrew C Myers, Erik Andersen, and François Guimbretière. Teaching programming with gamified semantics. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2017.

[24] Suranjith Ariyapperuma and Amina Minhas. Internet security games as a pedagogic tool for teaching network security. In *35th Annual Frontiers in Education*, pages S2D–1. IEEE.

[25] Ivon Arroyo, Beverly Park Woolf, Winslow Burelson, Kasia Muldner, Dovan Rai, and Minghui Tai. A multimedia adaptive tutoring system for mathematics that addresses cognition, metacognition and affect. *International Journal of Artificial Intelligence in Education*, 24(4):387–426, 2014.

[26] Stephanie Baker, Carmen Petrick, Taylor Martin, A Aghababyan, Zoran Popović, E Andersen, and YE Liu. Learning fractions through splitting in an online game. In *American Educational Research Association Annual Meeting, AERA*, 2013.

[27] Marjoke Bakker, Marja van den Heuvel-Panhuizen, and Alexander Robitzsch. Effects of mathematics computer games on special education students' multiplicative reasoning ability. *British Journal of Educational Technology*, 47(4):633–648, jul 2016.

[28] Sasha Barab, Michael Thomas, Tyler Dodge, Robert Carteaux, and Hakan Tuzun. Making learning fun: Quest atlantis, a game without guns. *Educational technology research and development*, 53(1):86–107, 2005.

[29] Blair Barr. *Checking the effectiveness of Quizlet as a tool for vocabulary learning*. PhD thesis, 2016.

[30] Richard Bartle. Hearts, clubs, diamonds, spades: Players who suit muds. *Journal of MUD research*, 1(1):19, 1996.

[31] Benetech. Mathshare.

[32] Emily C. Bouck and Sara Flanagan. Assistive Technology and Mathematics: What is There and Where Can We Go in Special Education. *Journal of Special Education Technology*, 24(2):17–30, jun 2009.

[33] Emily C. Bouck, Christopher Working, and Erin Bone. Manipulative

Apps to Support Students With Disabilities in Mathematics. *Intervention in School and Clinic*, 53(3):177–182, jan 2018.

[34] John Seely Brown, Allan Collins, and Paul Duguid. Situated cognition and the culture of learning. *Educational researcher*, 18(1):32–42, 1989.

[35] Diane Pedrotty Bryant, Brian R. Bryant, and Donald D. Hammill. Characteristic behaviors of students with LD who have teacher-identified math weaknesses. *Journal of Learning Disabilities*, 33(2), mar 2000.

[36] Todd W. Busch, Kari Pederson, Christine A. Espin, and Jacalyn W. Weissenburger. Teaching students with learning disabilities: Perceptions of a first-year teacher. *Journal of Special Education*, 35(2):92–99, jul 2001.

[37] Carrie J Cai, Philip J Guo, James Glass, and Robert C Miller. Wait-learning: leveraging conversational dead time for second language education. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, pages 2239–2244. ACM, 2014.

[38] M. N. Callaghan, J. J. Long, E. A. van Es, S. M. Reich, and T. Rutherford. How teachers integrate a math computer game: Professional development use, teaching practices, and student achievement. *Journal of Computer Assisted Learning*, 34(1):10–19, feb 2018.

[39] Carnegie Mellon University. Lynnette - Smarter way to learn middle school math.

[40] National Center for Learning Disabilities. The State of Learning Disabilities: Facts, Trends and Emerging Issues (Third Edition, 2014). Technical report, 2014.

[41] Peter Chapman, Jonathan Burket, and David Brumley. Picoctf: A game-based computer security competition for high school students. In *3GSE*, 2014.

[42] Alan Cheng, Lei Yang, and Erik Andersen. Teaching language and culture with a virtual reality game. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 541–549. ACM, 2017.

[43] Naomi C Chesler, Golnaz Arastoopour, Cynthia M D'Angelo, Elizabeth A Bagley, and David Williamson Shaffer. Design of a professional practice

simulator for educating and motivating first-year engineering students. *Advances in Engineering Education*, 3(3):n3, 2013.

[44] Carol A Christensen and Michael M Gerber. Effectiveness of computerized drill and practice games in teaching basic math facts. *Exceptionality: A Special Education Journal*, 1(3):149–165, 1990.

[45] Dorothy M Chun and Jan L Plass. Effects of multimedia annotations on vocabulary acquisition. *The modern language journal*, 80(2):183–198, 1996.

[46] Herbert H Clark and Meredyth A Krych. Speaking while monitoring addressees for understanding. *Journal of memory and language*, 50(1):62–81, 2004.

[47] Benjamin D Cone, Cynthia E Irvine, Michael F Thompson, and Thuy D Nguyen. A video game for cyber security training and awareness. *computers & security*, 26(1):63–72, 2007.

[48] Dan Conway, Ronnie Taib, Mitch Harris, Kun Yu, Shlomo Berkovsky, and Fang Chen. A qualitative investigation of bank employee experiences of information security and phishing. In *Thirteenth Symposium on Usable Privacy and Security*, 2017.

[49] Coolmath4Kids. Base Ten Blocks — Manipulatives.

[50] Seth Cooper, Firas Khatib, Adrien Treuille, Janos Barbero, Jeehyung Lee, Michael Beenen, Andrew Leaver-Fay, David Baker, Zoran Popović, et al. Predicting protein structures with a multiplayer online game. *Nature*, 466(7307):756–760, 2010.

[51] National Research Council et al. *How people learn: Brain, mind, experience, and school: Expanded edition*. National Academies Press, 2000.

[52] Mihaly Csikszentmihalyi. *Flow: The psychology of optimal experience*, volume 41. HarperPerennial New York, 1991.

[53] Gabriel Culbertson, Erik Andersen, Walker White, Daniel Zhang, and Malte Jung. Crystallize: An immersive, collaborative game for second language learning. In *CSCW 2016*.

[54] Gabriel Culbertson, Shiyu Wang, Malte Jung, and Erik Andersen. Social

situational language learning through an online 3d game. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2016.

[55] Patricia Cunningham, Sharon Crawley, and Lee Mountain. Vocabulary scavenger hunts: A scheme for schema development. *Reading Horizons*, 24(1):8, 1983.

[56] Andy Davis, Tim Leek, Michael Zhivich, Kyle Gwinnup, and William Leonard. The fun and future of CTF. In *USENIX Summit on Gaming, Games, and Gamification in Security Education (3GSE 14)*, 2014.

[57] David Dearman and Khai Truong. Evaluating the implicit acquisition of second language vocabulary using a live wallpaper. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1391–1400. ACM, 2012.

[58] Muhammet Demirbilek and Suzan Lema Tamer. Math teacher's perspectives on using educational computer games in math education. In *Social and Behavioral Sciences*, volume 9, pages 709–716. Elsevier, jan 2010.

[59] Tamara Denning, Adam Lerner, Adam Shostack, and Tadayoshi Kohno. Control-alt-hack: the design and evaluation of a card game for computer security awareness and education. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, 2013.

[60] Minyi Shih Dennis, Emily Sharp, Jacquelyn Chovanes, Amanda Thomas, Raquel M. Burns, Beth Custer, and Junkoung Park. A Meta-Analysis of Empirical Research on Teaching Students with Mathematics Learning Difficulties. *Learning Disabilities Research and Practice*, 31(3):156–168, aug 2016.

[61] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*, pages 9–15. ACM, 2011.

[62] Rachna Dhamija, J Doug Tygar, and Marti Hearst. Why phishing works. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, 2006.

[63] Sidney D'Mello, Andrew Olney, Claire Williams, and Patrick Hays. Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of Human Computer Studies*, 70(5):377–398, 2012.

[64] Christian T. Doabler, Hank Fien, Nancy J. Nelson-Walker, and Scott K. Baker. Evaluating three elementary mathematics programs for presence of eight research-based instructional design principles, nov 2012.

[65] Julie S Downs, Mandy B Holbrook, and Lorrie Faith Cranor. Decision strategies and susceptibility to phishing. In *Proceedings of the second symposium on Usable privacy and security*, 2006.

[66] Dybuster. Calcularis.

[67] Robert E. Bixler and Sidney K. D'Mello. Crossed eyes: Domain adaptation for gaze-based mind wandering models. In *ACM Symposium on Eye Tracking Research and Applications*, pages 1–12, 2021.

[68] Darren Edge, Elly Searle, Kevin Chiu, Jing Zhao, and James A Landay. Micromandarin: mobile language learning in context. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3169–3178. ACM, 2011.

[69] David Eigen, Christian Puhrsch, and Rob Fergus. Depth map prediction from a single image using a multi-scale deep network. In *Advances in neural information processing systems*, pages 2366–2374, 2014.

[70] Myrthe Faber, Kristina Krasich, Robert E Bixler, James R Brockmole, and Sidney K D'Mello. The eye–mind wandering link: Identifying gaze indices of mind wandering across tasks. *Journal of experimental psychology: human perception and performance*, 46(10):1201, 2020.

[71] Yuichiro Fujimoto, Goshiro Yamamoto, Takafumi Taketomi, Jun Miyazaki, and Hirokazu Kato. Relationship between features of augmented reality and user memorization. In *Mixed and Augmented Reality (ISMAR), 2012 IEEE International Symposium on*, pages 279–280. IEEE, 2012.

[72] John A. Gardner and Vladimir Bulatov. Scientific diagrams made easy with IVEO™. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 4061 LNCS, pages 1243–1250. Springer Verlag, 2006.

[73] Katherine Garnett. Math learning disabilities: Teaching and learners. *Journal of Reading, Writing, and Learning Disabilities International*, 3(1):1–8, 1987.

[74] Zeynep Gecu-Parmaksiz and Omer Delialioglu. Augmented reality-based virtual manipulatives versus physical manipulatives for teaching geometric shapes to preschool children. *British Journal of Educational Technology*, 50(6):3376–3390, nov 2019.

[75] James Paul Gee. What video games have to teach us about learning and literacy. *Computers in Entertainment (CIE)*, 1(1), 2003.

[76] Genia Connell. CUBES Strategy to Tackle Tough Word Problems — Scholastic.

[77] Russell Gersten, David J. Chard, Madhavi Jayanthi, Scott K. Baker, Paul Morphy, and Jonathan Flojo. Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Review of Educational Research*, 79(3):1202–1242, sep 2009.

[78] Ian W. Gibson. At the intersection of technology and pedagogy: Considering styles of learning and teaching. *Journal of Information Technology for Teacher Education*, 10(1-2):37–62, 2001.

[79] Robert Godwin-Jones. Emerging technologies augmented reality and language learning: From annotated vocabulary to place-based mobile games. 2016.

[80] Mark Gondree and Zachary NJ Peterson. Valuing security by getting [d0x3d!] experiences with a network security board game. 2013.

[81] S. L. Happy, Anirban Dasgupta, Priyadarshi Patnaik, and Aurobinda Routray. Automated alertness and emotion detection for empathic feedback during E-learning. *Proceedings - 2013 IEEE 5th International Conference on Technology for Education, T4E 2013*, pages 47–50, 2013.

[82] Nira Hativa. Computer-based drill and practice in arithmetic: Widening the gap between high-and low-achieving students. *American Educational Research Journal*, 25(3):366–397, 1988.

[83] Neil T Heffernan and Cristina Lindquist Heffernan. 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, 2014.

[84] Nathan Henderson, Jonathan Rowe, Luc Paquette, Ryan S Baker, and James Lester. Improving affect detection in game-based learning with multimodal data fusion. In *International Conference on Artificial Intelligence in Education*, pages 228–239. Springer, 2020.

[85] Serena Hillman, Carman Neustaedter, Alexandra Hillman, and Carolyn Pang. "I have a life": Teacher communication & management outside the classroom. In *Conference on Human Factors in Computing Systems - Proceedings*, pages 1–6, New York, New York, USA, 2019. ACM Press.

[86] Kenneth Holstein, Gena Hong, Mera Tegene, Bruce M McLaren, and Vincent Aleven. The classroom as a dashboard: Co-designing wearable cognitive augmentation for k-12 teachers. In *Proceedings of the 8th international conference on learning Analytics and knowledge*, pages 79–88, 2018.

[87] Kenneth Holstein, Bruce M. McLaren, and Vincent Aleven. Intelligent tutors as teachers' aides: Exploring teacher needs for real-time analytics in blended classrooms. In *ACM International Conference Proceeding Series*, pages 257–266. Association for Computing Machinery, mar 2017.

[88] Jason Hong. The state of phishing attacks. *Communications of the ACM*, 55(1), 2012.

[89] Robert Howland, Sachi Urano, and Junichi Hoshino. Sanjigenjiten: computer assisted language learning system within a 3d game environment. In *Advances in Computer Entertainment*, pages 262–273. Springer, 2012.

[90] Ting-Chia Hsu. Learning english with augmented reality: Do learning styles matter? *Computers & Education*, 106:137–149, 2017.

[91] Jonathan Huang, Vivek Rathod, Chen Sun, Menglong Zhu, Anoop Korattikara, Alireza Fathi, Ian Fischer, Zbigniew Wojna, Yang Song, Sergio Guadarrama, et al. Speed/accuracy trade-offs for modern convolutional object detectors.

[92] Adam Ibrahim, Brandon Huynh, Jonathan Downey, Tobias Höllerer, Dorothy Chun, and John O'Donovan. Arbis pictus: A study of language learning with augmented reality. *arXiv preprint arXiv:1711.11243*, 2017.

[93] IXL Learning. `https://www.ixl.com/`, 1998.

[94] Giannis Karagiannakis, Anna Baccaglini-Frank, and Yiannis Papadatos.

Mathematical learning difficulties subtypes classification. *Frontiers in Human Neuroscience*, 8(1 FEB), feb 2014.

[95] Tanja Käser, Gian Marco Baschera, Juliane Kohn, Karin Kucian, Verena Richtmann, Ursina Grond, Markus Gross, and Michael von Aster. Design and evaluation of the computer-based training program Calcularis for enhancing numerical cognition. *Frontiers in Psychology*, 4(AUG), 2013.

[96] Tanja Käser, KR Koedinger, and Markus Gross. Different parameters-same prediction: An analysis of learning curves. *Proceedings of the 7th International Conference on Educational Data Mining*, (Edm):52–59, 2014.

[97] Garnett Kate. Math Learning Disabilities. *Division for Learning Disabilities Journal of CEC*, 1998.

[98] Fares Kayali, Günter Wallner, Simone Kriglstein, Gerhild Bauer, Daniel Martinek, Helmut Hlavacs, Peter Purgathofer, and Rebbeca Wölfle. A case study of a learning game about the internet. In *International Conference on Serious Games*, pages 47–58. Springer, 2014.

[99] Fengfeng Ke and Tatiana Abras. Games for engaged learning of middle school children with special learning needs. *British Journal of Educational Technology*, 44(2):225–242, 2013.

[100] Diane Jass Ketelhut, Brian C Nelson, Jody Clarke, and Chris Dede. A multi-user virtual environment for building and assessing higher order inquiry skills in science. *British Journal of Educational Technology*, 41(1):56–68, 2010.

[101] Khan Academy. `https://www.khanacademy.org/`, 2006.

[102] Elisheba W. Kiru, Christian T. Doabler, Audrey M. Sorrells, and North A. Cooc. A Synthesis of Technology-Mediated Mathematics Interventions for Students With or at Risk for Mathematics Learning Disabilities. *Journal of Special Education Technology*, 33(2):111–123, jun 2018.

[103] René F Kizilcec and Hansol Lee. Algorithmic fairness in education. *arXiv preprint arXiv:2007.05443*, 2020.

[104] Els Kuiper and Martie de Pater-Sneep. Student perceptions of drill-and-practice mathematics software in primary education. *Mathematics Education Research Journal*, 26(2):215–236, 2014.

[105] Ponnurangam Kumaraguru, Justin Cranshaw, Alessandro Acquisti, Lorrie Cranor, Jason Hong, Mary Ann Blair, and Theodore Pham. School of phish: a real-world evaluation of anti-phishing training. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, page 3. ACM, 2009.

[106] Ponnurangam Kumaraguru, Yong Rhee, Alessandro Acquisti, Lorrie Faith Cranor, Jason Hong, and Elizabeth Nunge. Protecting people from phishing: the design and evaluation of an embedded training email system. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2007.

[107] Ponnurangam Kumaraguru, Yong Rhee, Alessandro Acquisti, Lorrie Faith Cranor, Jason Hong, and Elizabeth Nunge. Protecting people from phishing: the design and evaluation of an embedded training email system. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 905–914. ACM, 2007.

[108] Ponnurangam Kumaraguru, Steve Sheng, Alessandro Acquisti, Lorrie Faith Cranor, and Jason Hong. Teaching johnny not to fall for phish. *ACM Transactions on Internet Technology (TOIT)*, 10(2), 2010.

[109] Teemu H Laine. Mobile educational augmented reality games: A systematic literature review and two case studies. *Computers*, 7(1):19, 2018.

[110] Andrew S Lan, Anthony Botelho, Shamya Karumbaiah, Ryan S Baker, and Neil Heffernan. Accurate and interpretable sensor-free affect detectors via monotonic neural networks. In *International Conference on Learning Analytics & Knowledge*, 2020.

[111] Bruce Lander. Quizlet: what the students think–a qualitative data analysis. *CALL communities and culture–short papers from EUROCALL 2016 Edited by Salomi Papadima-Sophocleous, Linda Bradley, and Sylvie Thouësny*, page 254, 2016.

[112] Elmer Lastdrager, Inés Carvajal Gallardo, Pieter Hartel, and Marianne Junger. How effective is anti-phishing training for children. In *Symposium on Usable Privacy and Security*, 2017.

[113] Jean Lave and Etienne Wenger. *Situated learning: Legitimate peripheral participation*. Cambridge university press, 1991.

[114] Laura Lenz, Katharina Schuster, Anja Richert, and Sabina Jeschke. Are virtual learning environments appropriate for dyscalculic students?: A

theoretical approach on design optimization of virtual worlds used in mixed-reality simulators. *2015 IEEE Games Entertainment Media Conference, GEM 2015*, pages 1–8, 2015.

[115] Fotis Liarokapis, Panos Petridis, Paul F Lister, Martin White, et al. Multimedia augmented reality interface for e-learning (marie). *World Transactions on Engineering and Technology Education*, 1(2):173–176, 2002.

[116] Chap Sam Lim, Keow Ngang Tang, and Liew Kee Kor. Drill and practice in learning (and beyond). *Encyclopedia of the Sciences of Learning*, pages 1040–1043, 2012.

[117] CC Lin and YC Wu. The effects of different presentation modes of multimedia annotations on sentential listening comprehension. In *Proceedings of the 21th International Conference on Computers in Education*, 2013.

[118] CC Lin and YC Yu. Efl learners' cognitive load of learning vocabulary on mobile phones. In *Proceedings of the 20th International Conference on Computers in Education*, 2012.

[119] Yanjin Long and Vincent Aleven. Educational game and intelligent tutoring system: A classroom study and comparative design analysis. *ACM Transactions on Computer-Human Interaction*, 24(3):1–27, apr 2017.

[120] Yanjin Long and Vincent Aleven. Educational game and intelligent tutoring system: A classroom study and comparative design analysis. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 24(3):20, 2017.

[121] 3909 LLC Lucas P. *Papers, Please: a dystopian document thriller*. 2013.

[122] Thomas W Malone. Toward a theory of intrinsically motivating instruction. *Cognitive science*, 5(4):333–369, 1981.

[123] Richard E Mayer. Multimedia learning. In *Psychology of learning and motivation*, volume 41, pages 85–139. Elsevier, 2002.

[124] Richard E Mayer. Principles for managing essential processing in multimedia learning: Segmenting, pretraining, and modality principles. *The Cambridge handbook of multimedia learning*, pages 169–182, 2005.

[125] MIND Research Institute. ST Math - Visual Math Program.

[126] Jelena Mirkovic and Peter A. H. Peterson. Class capture-the-flag exercises. In *USENIX Summit on Gaming, Games, and Gamification in Security Education (3GSE 14)*, 2014.

[127] Gaurav Misra, Nalin Asanka Gamagedara Arachchilage, and Shlomo Berkovsky. Phish phinder: A game design approach to enhance user confidence in mitigating phishing attacks. *arXiv preprint arXiv:1710.06064*, 2017.

[128] Peggy L. Moch. Manipulatives work! `https://doi.org/10.1080/00131720108984802`. Accessed: 2020-09-20.

[129] Paul L. Morgan, George Farkas, and Steve Maczuga. Which Instructional Practices Most Help First-Grade Students With and Without Mathematics Difficulties? *Educational Evaluation and Policy Analysis*, 37(2):184–205, jun 2015.

[130] Luca Morini, Koula Charitonos, Sylvester Arnab, Tiziana Cervi Wilson, Billy Brick, Tyrone Bellamy-Wood, and Gaetan Van Leeuwen. Imparapp: designing and piloting a game-based approach for language learning. In *European Conference on Games Based Learning*, page 1005. Academic Conferences International Limited, 2016.

[131] Jason J Morrissette. Glory to arstotzka: Morality, rationality, and the iron cage of bureaucracy in papers, please. *Game Studies*, 17(1), 2017.

[132] Diana Nahmod. Vocabulary gamification vs traditional learning instruction in an inclusive high school classroom. 2017.

[133] National Center for Education Statistics. The Condition of Education - Children and Youth With Disabilities - Indicator April (2018), 2019.

[134] Diana G Oblinger. Games and learning. *Educause quarterly*, 29(3):5–7, 2006.

[135] Casey O'Donnell. Getting played: Gamification, bullshit, and the rise of algorithmic surveillance. *Surveillance & Society*, 12(3):349, 2014.

[136] Teachers Office of Technology Assessment and Technology(OTA-EHR-616. Teachers and Technology : Making the Connection. *Washington, DC: U.S. Government Printing Office*, (April), 1995.

[137] Min Wook Ok, Diane Pedrotty Bryant, and Brian R. Bryant. Effects of

Computer-Assisted Instruction on the Mathematics Performance of Students with Learning Disabilities: A Synthesis of the Research. *Exceptionality*, 0(0):1–15, 2019.

[138] Cynthia M Okolo et al. Learning disabled students' acquisition of keyboarding skills and continuing motivation under drill-and-practice and game conditions. *Learning Disabilities Research*, 5(2):100–9, 1990.

[139] Marc Olano, Alan T Sherman, Linda Oliva, Ryan Cox, Deborah Firestone, Oliver Kubik, Milind Patil, John Seymour, and Donna Thomas. Securityempire: Development and evaluation of a digital game to promote cybersecurity education.

[140] Eleanor O'Rourke, Erin Peach, Carol S Dweck, and Zoran Popovic. Brain points: A deeper look at a growth mindset incentive structure for an educational game. In *Proceedings of the third (2016) acm conference on learning@ scale*, pages 41–50, 2016.

[141] Cas Pars. Phree of phish: The effect of anti-phishing training on the ability of users to identify phishing emails. Master's thesis, University of Twente, 2017.

[142] Zhongxiu Peddycord-Liu, Veronica Cateté, Jessica Vandenberg, Tiffany Barnes, Collin F. Lynch, and Teomara Rutherford. A Field Study of Teachers Using a Curriculum-integrated Digital Game. In *Conference on Human Factors in Computing Systems - Proceedings*, pages 1–12, New York, New York, USA, 2019. ACM Press.

[143] Bernadette Perry. Gamifying french language learning: a case study examining a quest-based, augmented reality mobile learning-tool. 2015.

[144] Ana Cristina Pires, Sebastian Marichal, Fernando Gonzalez-Perilli, Ewelina Bakala, Bruno Fleischer, Gustavo Sansone, and Tiago Guerreiro. A tangible math game for visually impaired children. *ASSETS 2019 - 21st International ACM SIGACCESS Conference on Computers and Accessibility*, 2019.

[145] Nicola J Pitchford, Elizabeth Kamchedzera, Paula J Hubber, and Antonie L Chigeda. Interactive Apps Promote Learning of Basic Mathematics in Children With Special Educational Needs and Disabilities. *Frontiers in psychology*, 9:262, 2018.

[146] Jan L Plass, Dorothy M Chun, Richard E Mayer, and Detlev Leutner. Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of educational psychology*, 90(1):25, 1998.

[147] PwC. Game of threats – a cyber threat simulation, September 2017.

[148] Prashanth Rajivan and Cleotilde Gonzalez. Creative persuasion: A study on adversarial behaviors and strategies in phishing attacks. *Frontiers in psychology*, 9:135, 2018.

[149] Paul Ramsden. *Learning to teach in higher education*. Routledge, 2003.

[150] Pekka Räsänen, Diana Laurillard, Tanja Käser, and Michael von Aster. Perspectives to Technology-Enhanced Learning and Teaching in Mathematical Learning Difficulties. In *International Handbook of Mathematical Learning Difficulties*, pages 733–754. Springer International Publishing, Cham, 2019.

[151] Anna Maria Re, Martina Pedron, Patrizio Emanuele Tressoldi, and Daniela Lucangeli. Response to specific training for students with different levels of mathematical difficulties. *Exceptional Children*, 80(3):337–352, 2014.

[152] C Reigeluth and R Stein. Elaboration theory. *Instructional-design theories and models: An overview of their current status*, pages 335–381, 1983.

[153] Glenda Revelle, Emily Reardon, Kristin Cook, Lori Takeuchi, Rafael Ballagas, Koichi Mori, Hiroshi Horii, Hayes Raffle, Maria Sandberg, and Mirjana Spasojevic. Electric agents: Combining collaborative mobile augmented reality and web-based video to reinvent interactive television. *Computers in Entertainment (CIE)*, 12(3):1, 2014.

[154] Peg Rosen. Tutoring kids with dyscalculia — tutors for math issues. `https://www.understood.org/en/school-learning/tutors/types-of-tutoring/tutoring-kids-with-dyscalculia-what-you-need-to-know`. Accessed: 2020-09-20.

[155] Andrew Ruef, Michael Hicks, James Parker, Dave Levin, Michelle L Mazurek, and Piotr Mardziel. Build it, break it, fix it: Contesting secure development. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 690–703. ACM, 2016.

[156] Johnny Saldaña. *The coding manual for qualitative researchers*. sage, 2021.

[157] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4510–4520, 2018.

[158] Marc Ericson C Santos, Takafumi Taketomi, Goshiro Yamamoto, Ma Mercedes T Rodrigo, Christian Sandor, Hirokazu Kato, et al. Augmented reality as multimedia: the case for situated vocabulary learning. *Research and Practice in Technology Enhanced Learning*, 11(1):4, 2016.

[159] Wombat Security. State of the phish, 2017.

[160] Paul Seedhouse, Anne Preston, Patrick Olivier, Dan Jackson, Philip Heslop, Madeline Balaam, Ashur Rafiev, and Matthew Kipling. The european digital kitchen project. *Bellaterra journal of teaching and learning language and literature*, 7(1):0001–16, 2014.

[161] Kang Seokbin, Shokeen Ekta, L Byrne Virginia, Froehlich Norooz Leyla, Elizabeth Bonsignore, Caro Williams-Pierce, and Jon E. Froehlich. ARMath: Augmenting Everyday Life with Math Learning. *Conference on Human Factors in Computing Systems - Proceedings*, 2020.

[162] David W Shaffer. Epistemic frames for epistemic games. *Computers & education*, 46(3):223–234, 2006.

[163] Andrea Misao Shea. *Student perceptions of a mobile augmented reality game and willingness to communicate in Japanese*. Pepperdine University, 2014.

[164] Steve Sheng, Mandy Holbrook, Ponnurangam Kumaraguru, Lorrie Faith Cranor, and Julie Downs. Who falls for phish?: a demographic analysis of phishing susceptibility and effectiveness of interventions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2010.

[165] Steve Sheng, Bryant Magnien, Ponnurangam Kumaraguru, Alessandro Acquisti, Lorrie Faith Cranor, Jason Hong, and Elizabeth Nunge. Anti-phishing phil: the design and evaluation of a game that teaches people not to fall for phish. In *Proceedings of the 3rd symposium on Usable privacy and security*, 2007.

[166] Lei Shi, Holly Lawson, Zhuohao Zhang, and Shiri Azenkot. Designing

interactive 3D printed models with teachers of the visually impaired. In *Conference on Human Factors in Computing Systems - Proceedings*, pages 1–14, New York, New York, USA, may 2019. Association for Computing Machinery.

[167] Mikyung Shin, Diane P. Bryant, Brian R. Bryant, John W. McKenna, Fangjuan Hou, and Min Wook Ok. Virtual Manipulatives: Tools for Teaching Mathematics to Students With Learning Disabilities. *Intervention in School and Clinic*, 52(3):148–153, jan 2016.

[168] Kristen Shinohara and Jacob O. Wobbrock. In the shadow of misperception: Assistive technology use and social interactions. *Conference on Human Factors in Computing Systems - Proceedings*, pages 705–714, 2011.

[169] Adam Shostack. Security games & resources, September 2017.

[170] SMARTeacher. Prodigy — fun math learning platform for kids. `https://www.prodigygame.com/main-en/`. Accessed: 2020-09-20.

[171] Neelkamal Soares, Teresa Evans, and Dilip R Patel. Specific learning disability in mathematics: A comprehensive review, jan 2018.

[172] Sylvia Söderström and Borgunn Ytterhus. The use and non-use of assistive technologies from the world of information and communication technology by visually impaired young people: A walk on the tightrope of peer inclusion. *Disability and Society*, 25(3):303–315, 2010.

[173] David Stein. Situated learning in adult education. eric digest no. 195. 1998.

[174] Robert Stevens and Alistair Edwards. Mathtalk: The design of an interface for reading algebra using speech. In *Information Technology and Disabilities*, volume 1, pages 313–320. 1994.

[175] INC STUDYPAD. Splash Math - Fun Math Practice for Kindergarten to Grade 5.

[176] The 101st United States Congress. Individuals with Disabilities Education Act, 1990.

[177] Michael F. Thompson and Cynthia E. Irvine. Cyberciege scenario design and implementation. In *USENIX Summit on Gaming, Games, and Gamification in Security Education (3GSE 14)*, 2014.

[178] Jin-Ning Tioh, Mani Mina, and Douglas W Jacobson. Cyber security training a survey of serious games in cyber security. In *Frontiers in Education Conference (FIE)*, pages 1–5. IEEE, 2017.

[179] Hendrys Tobar-Muñoz, Silvia Baldiris, and Ramon Fabregat. Augmented reality game-based learning: enriching students' experience during reading comprehension activities. *Journal of Educational Computing Research*, 55(7):901–936, 2017.

[180] Nelly Tournaki. The Differential Effects of Teaching Addition through Strategy Instruction Versus Drill and Practice to Students with and Without Learning Disabilities. *Journal of Learning Disabilities*, 36(5):449–458, 2003.

[181] Andrew Trusty and Khai N Truong. Augmenting the web for second language vocabulary learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3179–3188. ACM, 2011.

[182] Kathleen Tuite, Timothy Pavlik, Sandra B Fan, Tyler Robison, Alexander Jaffe, Yun-En Liu, Erik Andersen, and Steven Tanimoto. Picard: A creative and social online flashcard learning game. In *Proceedings of the International Conference on the Foundations of Digital Games*, pages 231–234. ACM, 2012.

[183] Endel Tulving and Donald M Thomson. Encoding specificity and retrieval processes in episodic memory. *Psychological review*, 80(5), 1973.

[184] Utah State University. Matti Math Online Store.

[185] Christian David Vazquez, Afika Ayanda Nyati, Alexander Luh, Megan Fu, Takako Aikawa, and Pattie Maes. Serendipitous language learning in mixed reality. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 2172–2179. ACM, 2017.

[186] Laton Vermette, Joanna McGrenere, Colin Birge, Adam Kelly, and Parmit K. Chilana. Freedom to personalize my digital classroom: Understanding teachers' practices and motivations. In *Conference on Human Factors in Computing Systems - Proceedings*, pages 1–14, New York, New York, USA, 2019. ACM Press.

[187] Stephanie T Villeda, Brooke C Shuster, Lauren Magill, and Erik W Carter. Behavior-specific praise in the classroom. `www.safeandcivilschools.com/`, 2016. Accessed: 2020-09-20.

[188] Luis von Ahn. Duolingo: learn a language for free while helping to translate the web. In *Proceedings of the 2013 international conference on Intelligent user interfaces*, pages 1–2. ACM, 2013.

[189] Lev Semenovich Vygotsky. *Mind in society: The development of higher psychological processes*. Harvard university press, 1980.

[190] Maaike Waalkens, Vincent Aleven, and Niels Taatgen. Does supporting multiple student strategies lead to greater learning and motivation? Investigating a source of complexity in the architecture of intelligent tutoring systems. *Computers and Education*, 60(1):159–171, jan 2013.

[191] Chad Walker. Cryptomancer: A fantasy role-playing game about hacking, 2015.

[192] Ting-Kwei Wang, Jing Huang, Pin-Chao Liao, and Yanmei Piao. Does augmented reality effectively foster visual learning process in construction? an eye-tracking study in steel installation. *Advances in Civil Engineering*, 2018, 2018.

[193] Rick Wash and Molly M Cooper. Who provides phishing training?: Facts, stories, and people like me. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 492. ACM, 2018.

[194] Stuart Webb. The effects of repetition on vocabulary knowledge. *Applied linguistics*, 28(1):46–65, 2007.

[195] Zikai Alex Wen, Anjelika Lynne S. Amog, Shiri Azenkot, and Katherine Garnett. Teacher perspectives on math e-learning tools for students with specific learning disabilities. *ASSETS 2019 - 21st International ACM SIGACCESS Conference on Computers and Accessibility*, pages 516–518, 2019.

[196] Zikai Alex Wen, Erica Silverstein, Yuhang Zhao, Anjelika Lynne Amog, Katherine Garnett, and Shiri Azenkot. Teacher views of math e-learning tools for students with specific learning disabilities. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–13, 2020.

[197] WeWantToKnow. Dragonbox Math Apps.

[198] Gregory B White, Dwayne Williams, and Keith Harrison. The cyberpatriot

national high school cyber defense competition. *IEEE Security & Privacy*, (5):59–61, 2010.

[199] Nancy Cushen White, Graham Neuhaus, Marilyn C. Beckwith, Carolyn M.A. DeVito, Gloria Trabucco, Virginia W. Berninger, Larry E. Hess, Betsy Duffy, Jean Freyer Schedler, Colleen Uscianowski, Judith R. Birsh, and Suzanne Carreker. *Multisensory Teaching of Basic Language Skills*. Brookes Publishing Company, 4th edition, 2018.

[200] Wikipedia. Podesta emails. 2017.

[201] Wikipedia. List of flashcard software — Wikipedia, the free encyclopedia. `http://en.wikipedia.org/w/index.php?title=List\%20of\%20flashcard\%20software&oldid=866037569`, 2018. [Online; accessed 07-November-2018].

[202] Emma J Williams, Amy Beardmore, and Adam N Joinson. Individual differences in susceptibility to online influence: a theoretical review. *Computers in Human Behavior*, 72:412–421, 2017.

[203] Daniel T Willingham. Ask the cognitive scientist do manipulatives help students learn? `https://www.aft.org/ae/summer1992/ball.pdf`. Accessed: 2020-09-20.

[204] Anna J Wilson, Stanislas Dehaene, Philippe Pinel, Susannah K Revkin, Laurent Cohen, and David Cohen. Principles Underlying the Design of "The Number Race", an Adaptive Computer Game for Remediation of Dyscalculia. *Behavioral and brain functions : BBF*, 2(1):19, may 2006.

[205] Anna J. Wilson, Susannah K. Revkin, David Cohen, Laurent Cohen, and Stanislas Dehaene. An open trial assessment of "the number race", an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions*, 2, 2006.

[206] Che-Ching Yang, Shian-Shyong Tseng, Tsung-Ju Lee, Jui-Feng Weng, and Kaiyuan Chen. Building an anti-phishing game to enhance network security literacy learning. In *2012 IEEE 12th International Conference on Advanced Learning Technologies*, pages 121–123. IEEE, 2012.

[207] Makoto Yoshii. L1 and l2 glosses: Their effects on incidental vocabulary learning. 2006.

[208] Steve Chi-Yin Yuen, Gallayanee Yaoyuneyong, and Erik Johnson. Augmented reality: An overview and five directions for ar in education. *Journal of Educational Technology Development and Exchange (JETDE)*, 4(1):11, 2011.

[209] Xiao Wan Zhao, Eddie T.c. Lam, and Lih Ching Chen Wang. Students perception of quizlet as a chinese learning tool: a preliminary study. *International Journal of Technology Enhanced Learning*, 10(1/2):128, 2018.

[210] Yuhang Zhao, Shaomei Wu, Lindsay Reynolds, and Shiri Azenkot. The effect of computer-generated descriptions on photo-sharing experiences of people with visual impairments. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW):1–22, 2017.