USING VISUALIZATION TO UNDERSTAND THE PROBLEM-SOLVING PROCESSES OF ELEMENTARY STUDENTS IN A COMPUTER-ASSISTED MATH LEARNING PROGRAM

by

Shuang Wei

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Computer Graphics Technology

West Lafayette, Indiana

May 2020

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Yingjie Chen, Chair
Department of Computer Graphics Technology

Dr. Paul Parsons
Department of Computer Graphics Technology

Dr. Yan Ping Xin
Department of Special Education

Dr. Timothy Newby
Department of Curriculum & Instruction

Approved by:
Dr. Kathryne Newton

Dedicated to my beloved parents.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my major professor, Dr.Yingjie Chen, for his dedicated support and guidance from when I was first considering applying for the Ph.D. program through to completion of this degree. Dr.Chen continuously provided encouragement and was always willing and enthusiastic to assist in my study and research.

I would like to thank my committee members, Drs. Yan Ping Xin, Paul Parsons, and Timothy Newby for their guidance, thought-provoking suggestions, and comments on the dissertation. I would also like to thank the peer P.h.D student, Ya-hsin Hung, for her insightful advice for this research work.

I would like to thank the COMPS-A research project team. I am grateful to have been a member of the team. And the work of COMPS-A© (Xin, Kastberg, Chen, and Team including Kim, S., Wei, S. et al. 2015-2020) built the foundation of my dissertation research.

Finally, I would like to express my deep and sincere gratitude to my family for their continuous and unparalleled love, help, and support. Thanks to my parents for all the support and selfless love throughout my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CAL –	Computer Assisted Learning
CAI –	Computer Assisted Instruction
LA –	Learning Analytics
LAK –	Learning Analytics and Knowledge
EDM –	Educational Data Mining
EGG –	Electroencephalo Graph
HTML –	Hypertext Markup Language
SVG –	Scalable Vector Graphics
SQL –	Structured Query Language
MLD –	Math Learning Disability

ABSTRACT

CAL (Computer Assisted Learning) programs are widespread today in schools and families due to the effectiveness of CAL programs in improving students' learning and task performance. The flourishing of CAL programs in education has brought large amounts of students' learning data including log data, performance data, mouse movement data, eye movement data, video data, etc. These data can present students' learning or problem-solving processes and reflect underlying cognitive processes. These data are valuable resources for educators to comprehend students' learning and difficulties. However, few data analysis methods can analyze and present CAL data for educators quickly and clearly. Traditional video analysis methods can be time-consuming. Current visualization analysis methods are limited to simple charts or visualizations of a single data type. In this dissertation, I propose a visual learning analytic approach to analyze and present students' problem-solving data from CAL programs. More specifically, a visualization system was developed to present students' problem-solving data, including eye movement, mouse movement, and performance data, to help educational researchers understand student problem-solving processes and identify students' problem-solving strategies and difficulties. An evaluation experiment was conducted to compare the visualization system with traditional video analysis methods. Seven educational researchers were recruited to diagnose students' problem-solving patterns, strategies, and difficulties using either the visualization system or video. The diagnosis task loads and evaluators' diagnosis processes were measured and the evaluators were interviewed. The results showed that analyzing student problem-solving tasks using the proposed visualization method was significantly quicker than using the video method. In addition, diagnosis using the visualization system can achieve results at least as reliable as the video analysis method. Evaluators' preferences between the two methods are summarized and illustrated in the dissertation. Finally, the implications of the visual analytic approach in education and data visualization areas are discussed.

CHAPTER 1. INTRODUCTION

1.1 Background

Understanding students' problem-solving processes, identifying their problematic problem-solving strategies, and identifying students' difficulties are necessary and important for educators such as teachers and educational researchers. By understanding students' problem-solving difficulties, teachers can provide personalized instruction. By identifying students' problem-solving strategies, educational researchers can propose new teaching methods and work on curriculum revision. However, understanding students' problem-solving processes and identifying their problem-solving strategies and difficulties are challenging and time-consuming. For educators, there is an additional challenge because elementary school students are often too young to describe their thinking and problem-solving strategies. In this case, observation is very important. Teachers observe students' problem-solving processes and students' performance (problem-solving correctness) to determine students' problem-solving strategies and difficulties. However, teachers often don't have the time to work with students individually as they need to respond to dozens of students in the classroom. When teachers have no time to sit beside the student to observe their problem-solving process, a video recording is an applicable method for teachers to identify students' difficulties.

Video recording is also an important method for educational researchers to identify students' problem-solving strategies and difficulties. However, although useful, the video analysis process is tedious and time-consuming. An efficient method that can help teachers and educational researchers to understand students' problem-solving processes is desired.

At the same time, today, the flourish of CAL programs provides a wealth of student problem-solving data that teachers and researchers can take advantage of. CAL programs can record not only the performance data (correctness, time on task, etc.) but also their problem-solving behaviors (mouse movement, eye movement, video recording, etc.).

However, the abundance of data requires teachers' and researchers' data analysis capabilities. Sometimes, teachers even ignore these resources as they don't have time to go through the videos to obtain student learning information. While educational researchers use video data a lot but keep suffering from the tedious and frustrating video analysis. Teachers and educational researchers need a method that can help them understand students' learning/problem-solving processes in CAL programs efficiently.

CAL programs can collect many different types of data. For different data types, there are many analysis methods. Performance data such as correctness and time spent on tasks are often presented using numbers and text in tables (Jacovina et al., 2015). Student mouse movement data is analyzed and used to understand student learning in two methods. One method uses mouse movement features such as click, move speed, mouse move distance in x, and y directions to train the model to predict students' performance (Cetintas, Si, Xin, & Hord, 2009). Another method is combining mouse data feature analysis and trajectory visualization to present mouse movement (Freeman & Ambady, 2010). But as mouse trajectory visualization is hard to understand, it is often used as the supplement to the statistical analysis. For eye movement data, many studies apply eye-tracking technology. Similar to mouse movement, it is also analyzed statistically (Prantner, 2016; Moeller, Neuburger, Kaufmann, Landerl, & Nuerk, 2009) and visually (M. Schneider et al., 2008). Compared with mouse movement, there are many more studies about eye movement data visualization. However, most visualization design only presents eye movement data itself without considering other data sources. For example, the typical visualization method of eye-tracking data visualization - heatmap and gaze plot, can only present the trajectory of eye movement. For student problem-solving data analysis and illustration, a new method that can analyze and present multi-source data is necessary.

The eye movement and mouse movement data collected by CAL programs, that are not typically available in educational settings, to some extent, are able to reveal students' cognitive processes. For small kids who can't articulate themselves correctly, these data are valuable resources. This is one of the reasons eye-tracking technology has recently proliferated. It is widely agreed that eye movements are linked with attention during information acquiring tasks (Rayner, 1998). Therefore, eye movement is regarded as a window to the human mind and brain. They are believed to be able to reveal human's cognitive processes. Susac et al. (2014) collected university students' eye movements to investigate their strategies in simple equation solving. They found the number of fixations is a reliable and sensitive measure to present students' flow of attention in the problem-solving process. The comparison of eye movement analysis results and questionnaire reports showed that eye movement data are more objective and reliable. In the human learning field, eye-tracking methodology is utilized to explore student learning behavior and cognitive process, identify individual differences, and predict performance. Eye movements, as reliable data resources, should be used to promote teachers' instruction and to improve student learning. Mouse movement, another data source, is directly related to the performance of students. When students move their mouse to click, drag, and submit their answers, their confidence, hesitation, and uncertainty can be reflected in their mouse movement. There are studies that employed students' mouse movement to predict their emotion, experience, performasnce, etc. (Azcarraga & Suarez, 2013; Navalpakkam & Churchill, 2012; Cetintas, Si, Xin, & Hord, 2010). In the tasks that require mouse operation, mouse movement can directly reflect students' cognitive processes (M. C. Chen, Anderson, & Sohn, 2001).

CAL programs provide these valuable data to teachers and researchers. In this dissertation, the visualization analysis method is developed to help educational researchers to take advantage of the data collected in CAL programs to understand students' problem-solving processes quickly and accurately.

1.2 Significance

Developing a new visual analytic method to help eeducators comprehend students' learning and problem-solving processes quickly and accurately has significance in both education and visual analytics.

Analyzing and presenting students' problem-solving data in the CAL program to educators is important for improving the problem-solving skills of students, especially for elementary school students who are unable to describe their problem-solving process clearly. After students' problem-solving processes, problem-solving strategies, and difficulties are understood, personalized instructions and necessary assistance could be provided. By involving the educators's knowledge, experience, and judgment in student problem-solving data exploration, combined with visual representations, visual learning analytics significantly eases the process of problem-solving data comprehension, problem-solving pattern identification, and instruction decision making.

Visual analytics, which can facilitate data comprehension and exploration process, have proliferated in the educational data analysis field because it is efficient and intuitive (Mazza & Dimitrova, 2004; Vieira, Parsons, & Byrd, 2018). However, more research is needed to determine how visual learning analytics can be used to analyze and illustrate the learning/problem-solving process of elementary school students. In addition, many studies' visual learning analytics methods only analyze and present a single data type. They often fail to combine multiple data sources to comprehensively present student learning/problem-solving processes. A visual analytic method that combines different data types to assist teachers' instruction and educational researchers' study is desired. The developed visual analytic method also contributes to the whole visual learning analytics domain.

1.3 Statement of Purpose

This study develops a visual analytic method for educators to understand how students process information to solve word problems in mathematics from a computer-assisted math learning program (COMPS-A©, Xin, Kastberg, Chen, & Team, 2015-2020). More specifically, the visualization system shows students' learning data (performance data, eye movement data, and mouse movement data) for educators to make hypotheses and identify insights about the problem-solving strategies students employ and the difficulties students meet.

In this dissertation, the developed visualization system is evaluated compared with the video analysis method. The video analysis method is the traditional method used by educational researchers to comprehend students' problem-solving processes without information loss. The evaluation section of this dissertation compares the developed visualization system with the video analysis method in many aspects, including time, diagnosis reliability, and task load, to validate the efficiency of the developed visual analytic method in analyzing and presenting students' problem-solving processes.

1.4 Research Questions

In the context of a computer-assisted math learning program, how can we help educational researchers to efficiently comprehend and identify the problem-solving strategy that a student uses while solving the mathematical word problems?

To help educational researchers identify students' problem-solving strategies, this dissertation proposes a system that visualizes students' learning data (eye movement, mouse movement data, performance data) for educational researchers to comprehend students' problem-solving processes.

After the visualization development, we invite target users – educational researchers—to evaluate the visualization system and compare it with videos to answer three evaluation questions:

1. Compared with video, does the visualization system save educational researchers' time in identifying students' problem-solving strategies and difficulties?

2. Compared with video, does the visualization system help educational researchers get more reliable diagnoses on problem-solving diagnosis tasks (such as identifying students' problem-solving strategies, difficulties, etc.)?

3. What aspects of the visualization system/video result in high task load for educational researchers?

1.5 Assumptions

The assumptions of this dissertation are as follows:

1. The elementary school students in this dissertation are representative samples of 2^{nd} or 3^{rd}-grade elementary school students in the United States.

2. The elementary school students did their best to complete the tasks in the math computer program.

3. The knowledge that is required to solve the math problems successfully is consistent with the knowledge level of U.S. 2^{nd} /3^{rd} -grade students.

4. The recruited educational researchers in this dissertation are representative samples of educational researchers in student mathematical learning studies.

5. Evaluators did their best in both visualization and video diagnosis tasks.

Students and evaluators were recruited based on the principle of voluntariness. Participants are volunteers who might withdraw from the study at any time and with no ramifications. To ensure participants did their best in their tasks, participants received training tasks. In addition, they were told their performance is not the focus of this research study. The focus of this study is to compare the two analysis methods. Also, anonymity and confidentiality of their information are guaranteed in the study. To make sure assumption 3 - the knowledge that is required to solve the math problems successfully is consistent with the knowledge level of U.S. I selected tasks from a computer-assisted math learning program (COMPS-A©). The program is the produce of National Science Foundation (NSF) founded project COMPS-RTI (Xin, Kastberg, & Chen, 2015) and developed by math education and special education experts with a purpose to improve the problem-solving skills of 2nd /3rd -grade students with learning disabilities or difficulities.

1.6 Limitations

1. Due to the evaluation time limit, there were only 18 mathematical problem-solving tasks (two tasks from each elementary school student, in total nine students) in the evaluation.

2. As we only recruited experienced educational researchers, there were only seven participants in the evaluation study.
3. Students took the mathematical tasks in different environments. Some of the students took tasks in an elementary school computer lab, while the others took tasks in a university computer lab. Throughout, we did our best to keep the experimental settings similar. Students worked on the same model of laptops and eye trackers. All students worked in the afternoon, and the illumination of both labs were similar.
4. There was some noise in the learning data as the elementary school students were too small to sit still for a long time. They might play with the mouse, look around, or move away from the tasks. Where possible, noise in the data was removed. For example, if a student's total fixation duration in a task is lower than a threshold, the student task was removed from the dataset.
5. The students in the study were recruited from the Lafayette and West Lafayette area. While the evaluators were recruited from Purdue University.

1.7 Delimitations

1. The purpose of this project is to develop a visual analytic approach for teachers and educational researchers to understand student problem-solving processes. Other learning data analysis methods, such as educational data mining, are outside the scope of this dissertation.

2. This project is designed to analyze and present students' learning processes to help educators identify students' problem-solving strategies and difficulties. Other visual learning analytic directions, such as predicting student performance, are not within the research scope of the dissertation.

3. The target users of these visualization approaches are educational researchers instead of learners. These visualization approaches will assist educational researchers in achieving a better understanding of students learning behavior, learning patterns, and difficulties.

4. The project uses computer-assisted math learning programs and math learning data. The visual learning analytic method is proposed based on these data. However, the visual analytic approach can also be applied to other educational data with similar data types and structures.

1.8 Summary

This chapter introduced the importance of understanding students' problem-solving processes and comprehending students' difficulties for teachers and educational researchers. Then, the methods of analyzing and presenting different types of students' problem-solving data were briefly reviewed. The importance of developing an efficient learning/problem-solving data analysis method is emphasized. Additionally, the research purpose, research questions, assumptions, limitations, and delimitations of the study were introduced in this chapter. In the next chapter, a state-of-the-art literature review will be presented to introduce the status of CAL programs, CAL data, and visual learning analytics. The visual learning analytic studies analyzing the learning data of students will be reviewed in detail.

CHAPTER 2. LITERATURE REVIEW

In this chapter, I will first introduce the development of CAL and the efficiency of CAL programs in education. The supplementary functions of CAL programs will be introduced. Furthermore, different types of learning data recorded by CAL programs are presented in this chapter.

A promising learning data analysis method, called visual learning analytics, is then introduced. This literature review aims to introduce some related researchers' work in using visual learning analytics to understand student learning processes. Their analysis methods, visualization methods, and conclusions on student learning behavior are summarized. Since eye movement and mouse movement are rarely used in current visual learning analytics research, other related research about eye movement and student learning behavior is reviewed and introduced.

2.1 Computer-Assisted Learning (CAL)

CAL is a field that has and changed with the development of technology. In the 1970s and 1980s, with the widespread of computers, computer software was developed to assist education. Computers were used as a tool for tutoring or instructing students in schools (Handal & Herrington, 2003). In the 1980s, more CAL software was used outside of schools (Robert P. Taylor, 1980). At that time, applications were delivered on CD-ROM disks. The rapid expansion of the internet spurred further development of CAL, making it possible to distribute learning material online. It also facilitated the use of new tools, information sharing, and learning cooperation. Because of the development of computer technology and the emergence of new technologies, CAL environments, platforms, and objectives have changed greatly.

In the early stages of CAL development, CAL programs were used primarily as learning tools and tutorials to present learning material and assist in the completion of learning tasks. CAL applications later provided more aggregated learning material to users, such as hypermedia learning systems. The internet has become very important in schools, and the majority of CAL applications are web-based.

2.1.1 Categories of CAL Programs

Alessi and Trollip (2001) classified CAL applications into the following categories: tutorials, hypermedia, simulations and games, tools, open-ended learning, and web-learning. Tutorials present information to students and guide their learning process. Drills and practice help students to focus on mastering basic skills (Handal & Herrington, 2003). For example, math learning programs, with the method of drills and practice, will present students with math questions, let students enter their solutions, and provide feedback to reinforce correct answers and improve problem-solving strategies. Games and simulations are goal-oriented learning software. Games provide a competitive setting (win/lose) for specific subjects. Simulation modules offer a virtual environment, in which students learn as participating members, instead of observers (Alina Zapalska, Dallas Brozik, 2012). Hypermedia learning is different from other learning styles. It consists of many nodes, including text, audio, or animation. Students can access any of these nodes, depending on their learning requirements (Moos & Azevedo, 2008). Computer applications are also used as tools to assist learners in completing their learning tasks, such as writing, drawing, calculating, etc. The open-ended learning environments are comprehensive, integrated systems comprising learner-centered tasks. In contrast to directed learning, in open-ended learning systems, learners decide what is to be learned and what tasks will be selected. The intents and goals of each learner are uniquely established and pursued (Hannafin et al., 2018). Although CAL applications are generally divided into these categories, there are no strict boundaries. Categories overlap for many programs. Web-based learning only defines the learning platform. Any learning methodology can be combined with a web-based learning environment (Alesi & Trollip, 2001).

Some web-based learning applications, such as massive open online courses (MOOCs), put teaching videos, learning materials, and tasks on the internet to provide greater access to students studying at a distance (Handal & Herrington, 2003). Some applications, such as learning management systems (LMS), provide communication, collaboration and reporting tools to offer an interactive learning experience (Bakhshinategh, Zaiane, ElAtia, & Ipperciel, 2018). Different from some applications that employ a "just-put-it-on-the-web" approach, intelligent CAL applications, such as intelligent tutoring systems (ITS), adjust learning materials and tasks based on the students' performance in order to fulfill personalized demands(Siemens & Baker, 2012). These web-based learning applications accommodate various learning styles and offer many students freedom in their learning methods. Moreover, in web-based education, "large amounts of information about teaching-learning interaction are continuously generated and ubiquitously available" (Romero & Ventura, 2010, p601).

Different types of CAL programs tend to be used at different educational levels. For example, clear and straightforward programs, like tutorial programs, are often used in primary education. However, complex and integrated programs, like MOOCs and LMS, are often used in higher education (Nigh, Pytash, Ferdig, & Merchant, 2015), because of their complexity.

2.1.2 The Efficiency of CAL

There is an ongoing debate about the effectiveness of CAL. Many studies have proven the effectiveness of CAL on student learning (Roschelle et al., 2010; Pilli & Aksu, 2013; De Witte, Haelermans, & Rogge, 2015), while other studies have found that CAL provides no learning advantage over traditional forms of instruction (Gleason, Carnine, & Boriero, 1990; Stultz, 2008). The differences in study results may be caused by a variety of factors, such as experiment design, sampling and experiment duration (Fletcher-Flinn & Gravatt, 1995). There are many CAL-related meta-analysis papers that attempt to obtain generalized conclusions, by reviewing the empirical studies and calculating the mean effect size of these studies. The meta-analysis papers adopted strict paper selection criteria to ensure that the studies presented in the papers are comparable. The criteria include, but are not limited to, well-controlled studies (randomly assigned control and experiment groups), standardized achievement tests, adequate sample size, and appropriate study duration (Murphy et al., 2002). All the meta-analysis papers achieved positive mean effect sizes, which, to some extent, proved the effectiveness of CAL (Ryan, 1991; Christmann & Badgett, 2003;Rayne & Baggott, 2004).

CAL programs can provide students with numerous practice opportunities, flexible access, multiple representation methods, immediate feedback, etc. Some CAL programs can even record the student's learning data and transmit it to teachers, which has an important influence on student learning. These programs enable teachers to understand student learning difficulties and provide specific instruction (Fletcher-Flinn & Gravatt, 1995). Koedinger, McLaughlin, & Heffernan (2010) evaluated a web-based math tutor – ASSISTments. In their interview, some teachers have reported changes in their instruction because of students' performance in ASSISTments. For example, Ms. Metelenis reviewed the students' performance report and found that 70% (14) of her students needed help with a word problem, which evaluated the skill of decimal multiplication. As a result, the teacher spent an extra 15 minutes of class time to discuss a similar problem. “Of the 14 students who originally needed assistance, 50% benefited from the additional instruction in the sense that they solved a related item correctly the next time they used ASSISTments (Koedinger, McLaughlin, & Heffernan, 2010, p.496).” This demonstrates one way in which students benefit from teachers adapting their instructions based on students’ learning data.

The learning data registered by CAL programs provide more possibilities for educators to understand students’ learning behavior and cognitive processes. The data can also facilitate CAL program optimization, since student requirements for the programs can also be reflected in their learning data. In the next section, learning data and learning data analysis will be introduced and defined.

2.2 Data

Learning data reports are important for intervention. However, analysis and presentation of learning data is a major challenge. This is because the data collected by CAL programs is large, complex, unstructured, and uncleaned. How to analyze heterogeneous datasets and intuitively present valuable information to educators to facilitate their instruction is the core problem of learning data analysis. This section introduces different types of learning data and different learning data analysis methods, especially visual learning analysis.

2.2.1 Different Types of Learning Data

Depending on the learning environment (computer-based education or web-based education) and the purpose of educational systems (E-commerce and research-purpose), data can also be categorized as performance data or log data. In addition, other data that can be collected from different resources include eye tracking systems, EEG, assessment, field observation, questionnaires, etc.

In computing, a log file is used to register any events that occur between users and the operating system. In the educational domain, the typical log data includes username, start time, end time, list of actions and their timestamp (Doleck et al., 2016; Hershkovitz & Nachmias, 2012). Student performance data includes correctness, time spent on each question, repeat times and number of errors made. Student activity data, such as eye movement, are also collected in some learning studies. The advanced computer technology enables various types of educational data collection. All the learning data mentioned above is collected in real time, while students interact with the system.

2.2.2 Eye Tracking Data

Eye-tracking technology has been intensively used in many studies related to cognitive neuroscience, psychology, learning, and human-computer interaction (Karatekin, 2007; Yoon & Narayanan, 2004; Poole & Ball, 2005; Tai, Loehr, & Brigham, 2006). In learning, the particular topics that eye trackers have been used to investigate include problem-solving (Tsai, Hou, Lai, Liu, & Yang, 2012; J H Boonen & Jolles, 2015; Moutsios-Rentzos & Stamatis, 2015), information processing (Kabugo, Muyinda, Masagazi, Mugagga, & Mulumba, 2016; Copeland, 2016), learning strategies (Catrysse et al., 2018; Lee & Anderson, 2001), decision making(Renkewitz & Jahn, 2012), and individual differences (Blignaut & Wium, 2014; Bartolotti & Marian, 2012).

The participants of above studies covered all age groups from infants to adults. Most papers did not mention the influence of age on eye-tracking data accuracy and precision when they studied on a particular age group of participants. Here, in the dissertation study, the eye-tracking methodology was applied to second or third-grade elementary school students (ages 7 to 8). Is eye tracking data quality influenced by age? What other factors will influence eye-tracking data quality? The related literature is reviewed.

The eye-tracking data quality is related to four components: 1) spatial accuracy. This means the distance between the eye tracker’s recorded gaze point coordinates and the real coordinates of the gaze. 2) Spatial precision. This means, in repeated eye-tracking cases when the real gaze point coordinates are stable, the variation of the recorded gaze point coordinates. 3) Temporal accuracy. This is the difference between the recorded gaze event time and the real gaze event time. 4) Data loss. This refers to the expected number of gaze data recorded, and the real number of data recorded. For example, blinks and looking away from the screen may cause data loss (Hessels & Hooge, 2019; Dalrymple, Manner, Harmelink, Teska, & Elison, 2018; Niehorster, Cornelissen, Holmqvist, Hooge, & Hessels, 2018).

Hessels and Hooge (2019) reported the distribution of eye-tracking data quality measures for four age groups, including 5-months infants, 10-months infants, 3-years preschoolers, and 9-years school-age children. Except for the 3-years group which only included 31 participants, the other three groups had 500 participants, respectively. They employed Tobii TX300 to collect eye-tracking data. The data analysis results show that the data precision of 5-months infants is worst, while the 9-years school-age children's data had the best precision. They also found that the differences in data quality (precision and accuracy) are more significant between younger participants than the differences between older participants. For data loss, authors pointed out that 5-month-old infants have a higher level of data loss compared to 10-month-old infants. Furthermore, 9-years school-age children had the lowest data loss. The authors did not report the significance of differences among different age groups. However, they pointed out that the research assistants experiment guidance could affect the eye-tracking data quality.

Another paper, by Dalrymple et al. (2018) answered whether the data quality differences among different age-groups are significant or not. Moreover, they included adults in their study. There were eleven university students, eleven school-age children (8 years – 11 years, mean age = 9.9 years), thirty-six 18-month-old toddlers, and thirty-six 30-month-old preschoolers. The authors also employed a Tobii TX300 eye tracker to collect participants' eye-tracking data. According to the user manual of Tobii TX300, the gaze accuracy is 0.4º - 0.9º, and the precision is 0.04º-0.15º, depending on the illumination, gaze angle, and some other factors. Their study results showed the average data accuracy of adults was 0.78º (range: 0.35º - 1.52º), and the data precision (average standard deviation across participants) was 0.11º (horizontal), 0.18 º(vertical). For school-age children, the data accuracy mean was 0.93º (range: 0.37º - 2.70º), and the data precision was 0.14º, 0.19º. The authors conducted two rounds of experiments for 30-month-old preschoolers and 18-month-old toddlers to validate the research results as there were many outliers in these two groups' data. For 30-month-old preschoolers, the data accuracy in the two rounds were 1.29º (range: 0.67º – 2.33º), 1.77º (range 0.81º – 5.58º), and the precisions were (0.19º, 0.21º), (0.17º, 0.20º). For 18-month-old toddlers, the data accuracy in the two rounds were 1.31º (range: 0.18º – 3.85º), 1.28º (range: 0.56º – 2.32º), and the precision was (0.20º, 0.21º), (0.20º, 0.21º). The authors concluded that for adults, the accuracy and precision were within or very close to the Tobii's advertised specifications. The statistical analysis presented no significant data accuracy and precision differences between adults and school-age children. The data accuracies of 18-month old and 30-month old had no significant differences from each other. However, adults and school-age children had significantly higher data accuracy and precisions than toddlers. This difference may be caused by the fact that the subcortical regions mediating prosaccades are mature by age 4-6 (Karatekin, 2007).

The eye-tracking data of the dissertation study came from school-year students (7-8 years old). The study adopted Tobii Pro X3-120 eye tracker to collect student eye movement data. The Tobii Pro X3-120 user manual states that the accuracy of the equipment is 0.4 °, and the precision is 0.24°. The advised operational distance is 50 – 90 cm, and the freedom of head movement is 50 cm *40 cm (width * height). To sustain eye-tracking data quality, I let participants sited within the appropriate distance, consistent illuminance, enlarged AOI (area of interests), and removed low-quality cases or participants, which will be depicted in detail in the next chapter.

As long as the eye-tracking data quality is ensured, the potential for eye-tracking technology in education is promising. More and more researchers employed eye-tracker in their studies. Moreover, the emergence of low-cost eye-tracking devices further advanced the flourishing and development of eye-tracking technology. In the next section, eye-tracking data analysis methods will be introduced.

2.2.3 Learning Data Analysis Methods

There are three main methods that are used to analyze learning data: learning analytics (LA), educational data mining (EDM), and visual learning analytics. These methods are not totally distinct from each other. There is some overlap between the three methods.

According to Papamitsiou and Economides (2014), EDM adopts a reductionist viewpoint by analyzing individual components, seeking new patterns, and modifying algorithms, while LA analyzes a system within a holistic framework and tries to fully understand the system. EDM focuses on developing new computational data analysis methods, while LA focuses on using the known methods and models to address issues related to student learning and the organizational learning system (Bienkowski, Feng, & Means, 2012). Visual analytics address questions of cognition, metacognition, motivation, affect, language, social discourse, etc. (Koedinger et al., 2015). In educational data visualization, researchers answer questions related to student behavior, student performance, student feedback, assessment, curriculum, and domain knowledge (Peña-Ayala, 2014). Visual data analysis includes highly advanced computational methods and graphics to expose patterns and trends in large, complex datasets. The goal of visual analysis is to highlight useful information and support decision-making. Many websites and applications (Many Eyes, Google Charts, Flowing Data, and so on) have provided visualizations to support visual analysis.

The objective of our study is to identify student problem-solving patterns and inform educators about students' problem-solving strategies and difficulties. Therefore, we will use learning analytics to analyze student learning data and visual analytics to highlight our findings and support educators' exploration and understanding process.

2.3 CAL Data Analysis

Many studies have utilized analytics of students' learning data to provide insights into student learning behavior. A basic type of analysis uses simple measures of frequency of access to programs and the number of lessons that have been taken to determine the usage frequency of CAL programs and the learning progress of students. Student correctness data is also analyzed to evaluate student performance and to indicate students' knowledge. Other learning data, such as mouse clicks, eye movements, and videos are also analyzed to understand the student's learning strategies and cognitive processes.

2.3.1 Log Data and Performance Data Analysis

Log data analysis typically utilizes statistical analysis techniques, such as descriptive statistics (Ali, Hatala, Gašević, & Jovanović, 2012), simple correlation analysis (Mc Alister, Dunn, & Quinn, 2005), regression analysis and (M)AN(C)OVA. Simple correlation analysis, regression analysis, and ANOVA are often used to predict, compare, and build a relationship. For example, Y. Chen et al., (2017) developed statistical models to analyze millions of learners' activity logs, including clickstream, forum posts, and assignment records to predict student dropout behavior. De Witte et al., (2015) applied descriptive statistics and correlation analysis to analyze log data for 9898 students (the number of exercises the student took) and performance data (pre- and post-test scores) from a computer-assisted online tool called "Gotit!?" They concluded that more exercises led to higher test results (the coefficient is positive and significant). "Since the average student makes 50 exercises, he/she increases the post-test scores by 0.035 (0.0007*50), which is about 3.5%, as the post-test ranges between 0 and 1" (p.326).

Although many analytical methods are used in learning analysis research, when presenting students' learning status to teachers, descriptive statistics are often employed. For example, in the dropout study of Y. Chen et al., (2017), where the research purpose was to build a model predicting student dropout rate, data mining and complex statistical models were utilized. However, when researchers want teachers to understand student learning patterns and factors that can predict student dropout behavior, they used descriptive statistics to understand student learning activities.

2.3.2 Mouse Data Analysis

Mouse data is another type of log data. Mouse data includes the number of clicks, the duration of the click, the speed of mouse movement, and the moving path. Computer mouse trajectory racking is a behavioral methodology that was developed recently (Hehman, Stolier, & Freeman, 2015). Mouse data, especially for the mouse trajectory, is not as widely used as other types of log data, due to its complexity and heterogeneity.

Mouse trajectory analysis is complex and is still being studied. It is rarely used to assist educators in the understanding of student learning. But it is analyzed to discover the learner's emotions, cognition, and psychology. The efficiency of the mouse trajectory data is still under evaluation. For example, Azcarraga and Suarez (2013) combined student mouse behavior (number of clicks, duration of each click, and distance traveled by the mouse) with brainwave data to predict student emotions during learning. The methods used by researchers to analyze the mouse data are multi-layered perception (a class of feedforward artificial neural network) and support vector machines (supervised learning models in machine learning). Twenty-five computer science undergraduate students were asked to solve four algebra equations of different difficulty levels from a tutoring software "Aplusix". During their operation, their mouse clicks, mouse duration, and movements were registered in two different mouse log files – one for click and duration and another for mouse movement trajectory (x- and y- positions of the mouse). Student mouse data was analyzed through machine learning methods to predict student emotions during their problem-solving. The analysis results were triangulated with EGG data results and students' self-reports.

Mouse clicks and click durations are more often analyzed to depict students' learning progress, especially in learning scenarios that the click is required. Antonenko, Toy, and Niederhauser (2012) applied cluster analysis, which is a group data classification method, to analyze click-stream data from an online learning environment. After analysis and aggregation, click-stream data provided rich information about the sequence of accessing tasks, the rate at which learners advance through the learning environment, and the amount of time spent on tasks/resources.

2.3.3 Eye Movement Data Analysis

Many methods have been employed to analyze eye movement data, including plotting gaze data, plotting heat maps, counting the number of gaze fixations in an area of interest (AOI), timing the duration of fixation in an AOI (de Koning et al., 2010; Yüksel & Yıldırım, 2015), and following the path of student gazes (Lin et al., 2016). Some research also combined the eye-tracking with qualitative approaches, such as think-aloud, living observation, and video analysis in their study (Kekule, 2015; Leow et al., 2014). Eye movement data analysis is typically used to explore student learning behavior and cognitive processes. It is also used to identify learning behavior differences between students.

E*xploring student learning behavior and cognitive process.* Hegarty, Mayer, and Green (1992) noticed that “students have difficulty solving arithmetic word problems containing a relational term that is inconsistent with the required arithmetic operation” (p.76). For example, students may have difficulties solving a problem that requires addition operations, but where in a term “less” is used. Researchers utilized eye movement data to explore the student problem-solving process. Researchers recruited 38 undergraduate students. A total of four sets of 18 arithmetic word problems containing four target problems were given to students. Based on the participant’s performance, the researchers grouped them into a low-accuracy group (two or more errors) and a high-accuracy group (one or no errors). They then explored the eye movements (number of rereads, word fixations in the initial reading, and words fixated in rereading) of the high-accuracy group to identify their problem-solving behavior and patterns. Researchers found that high-accuracy students focused more on numbers than text. They also focused on a progressively smaller proportion of the words on a line when rereading. Researchers also found no significant difference in rereading numbers between consistent and inconsistent problems. However, significant differences existed in the rereading of the relevant background information (such as the variable names and relational terms). High-accuracy students revisited more relevant background information in the inconsistent problems. Researchers interpreted that students were constructing the situation model of the problem when they reread background information. When they reread numbers, they were planning their solution.

Bolden, Barmby, Raine, & Gardner, (2015) explored the different approaches that elementary school students used to interpret the dot arrays. They applied heatmaps (Figure 2.1) to show elementary students' eye fixation distribution and duration. In the top left heatmap, the student looked at every dot in the array to count them. The top right heatmap shows the student looked at x and y axes to count the array while in the bottom left heatmap, there is no clear approach that the student used to count the array. The eye movement heatmap give educators a hint on the student cognition process.

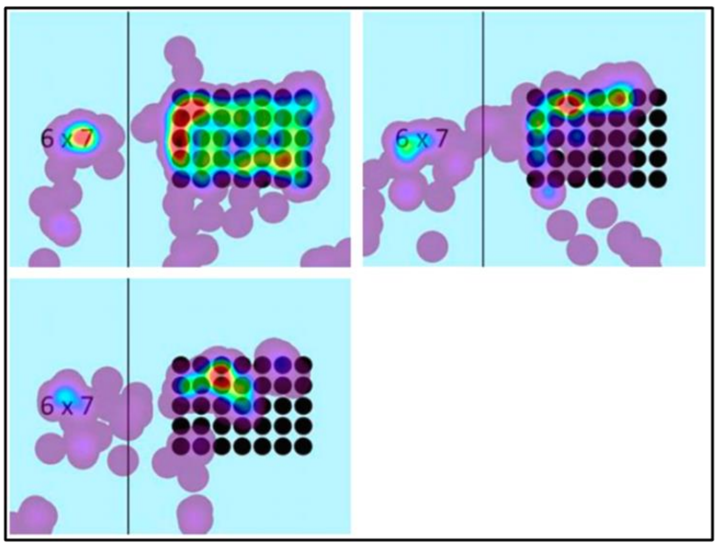

Figure 2.1 Example of a task used in the study of Susac et al. (2014).

There are many other studies that analyzed student eye movement data to explore student decision-making processes in choosing a solution strategy (De Corte & Verschaffel, 1986), and the attention distribution characteristics of elementary school students (Shuang Wei, et al., 2018).

Identify learning behavior differences between students. Susac et al., (2014) applied eye-tracking to reveal students' strategies in simple equation solving. Forty graduate students were recruited. Twenty-two students were studying mathematics, science, and engineering, while the other 18 students were not majoring in math-related fields. These students were required to complete many equation tasks in Figure 2.2. The participants' task was to make x the subject of the equation, while a potential answer was presented below the equation. Students could either look at DA (Yes) or NE (No) to submit their answer. Students' eye movement data, including gaze duration within the screen, fixation duration within areas of interest (equation area and answer area) and the number of fixations located in areas of interest were recorded. Inverse efficiency (inverse efficiency equals fixation duration within AOIs divided by accuracy) was calculated to account for speed-accuracy tradeoffs. The Pearson correlation coefficient, one-way ANOVA, and chi-square test were also used to discover student metacognitive insights eye movement characteristics. Analysis results suggested that the differences in the frequency of checking the offered answer, were a rough indicator of metacognitive accuracy. This is because the students who checked the answer frequently were making more errors than students who rarely looked at the answer. They found a correlation between the number of glances toward the answer and student efficiency (students' response accuracy and speed were taken into account for efficiency), which indicated that the students with higher metacognitive accuracy were more efficient in equation solving. They also found a positive correlation between the task difficulty and the number of fixations.

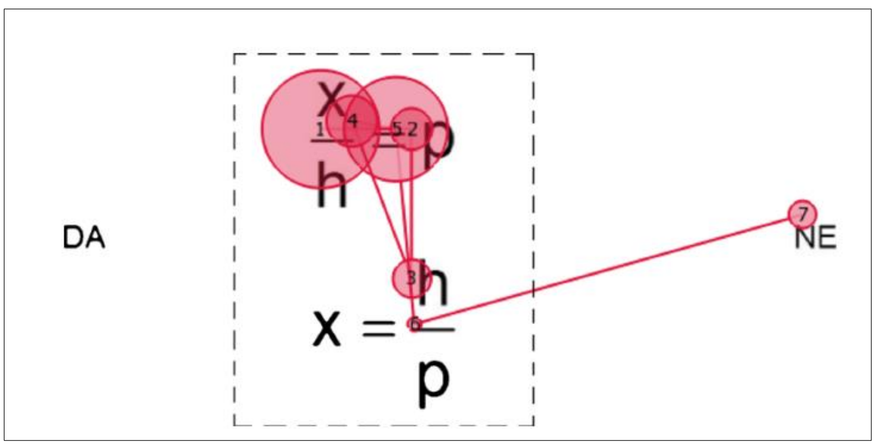

Figure 2.2 Example of a task used in the study of Susac et al. (2014).

Smith et al., (2010) explored student attention to conceptual information. Researchers found that students' gaze jumped between text information and mathematical statements. Despite the large amount of time (about 40%) that students paid on reading the text, they still had difficulties in recalling high-level conceptual information the text contained. Eye-tracking can objectively reflect where students looked to further identify their learning behavior and learning strategies.

Eye movement data is widely used in educational research to explore student learning behavior and to identify student's cognitive processes. However, teachers can seldom get their student's eye movement data analysis results to facilitate their instruction. This may be due to the expense of an eye tracking system and the complexity of eye movement data interpretation. However, with the development of low-cost eye tracking systems, and even no-cost eye tracking plug-ins, eye movement data will be more accessible in the near future. The use of graphics to present complex eye movement data for teachers should simplify data interpretation.

2.3.4 Video Data Analysis

Other learning data, such as video, audio, and questionnaires could also be collected. The methods used to analyze qualitative data are different from the methods adopted in quantitative data analysis. For example, Kaczorowski and Raimondi (2014) analyzed learning videos of four elementary students, using supplicated video-coding software. Researchers coded the student’s attention (where they were looking), navigational habits (how quickly they were swiping through pages), independent work habits (how often they asked for help) and engagement (how often they used support tools) to summarize the learning behavior patterns of students. The analysis results were: 1) students tended to solve the practice questions with immediate feedback, rather than watching review videos, since there was no new visual information on the video page; 2) while watching videos, students were watching other visual stimuli in the room while listening to the video; 3) They used a guess-and-check approach; 4) Although they did not use supporting material, such as demo videos, they still learned since they could use the immediate feedback questions as reference problems to help them with later problems.

The findings of Kaczorowski and Raimondi (2014) benefits the learning data analysis and the learning behavior pattern identification. More specifically, compared with the time-consuming video-coding method, how quantitative learning data analysis can identify student learning characteristics for educators is the research focus of this dissertation.

2.4 Visualization

Visualization uses graphics to display data with the aim of maximizing comprehension. In visual learning analysis, visualization is used to present complex, multidimensional student learning data (Mazza & Dimitrova, 2004). Visualization enables people to use their perceptual abilities and professional knowledge to interpret learning data and understand students’ learning behaviors. “Perception is very powerful. It conveys a large amount of information to our mind, allowing us to recognize essential features and to make important reference" (Koedinger, Baker, Cunningham, & Skogsholm, 2010 p. 10). If visualizations of learning data are provided to teachers, they could use their perceptual abilities, professional knowledge, and teaching experience to understand student learning behavior, comprehend student learning strategies, and identify student learning struggles.

In this section, I will review different types of learning data visualizations adopted by previous studies. The reviewed visualizations are categorized into log data, performance data, and eye movement data visualizations. However, it should be noted that visualizations are just presentation methods. One type of data can be illustrated by many visualization formats, and one visualization format can be used to present many types of data. In this section, the commonly used visualizations for log data and eye movement data are summarized.

2.4.1 Log Data and Performance Data Visualization

In visual learning analytics, the most commonly used data are log data and performance data. The most commonly used visualizations are simple charts such as bar charts, pie charts, scatter plots, line charts, and heatmaps (Vieira et al., 2018).

A *Bar Chart* is the most commonly used chart in log and performance data visualization. It can be used to represent correctness, task completion time, times of repeat, etc. Figure 2.3 is a simple bar chart that depicts the correct answer percentages (performance data) for two groups of students (B. Schneider & Pea, 2014).

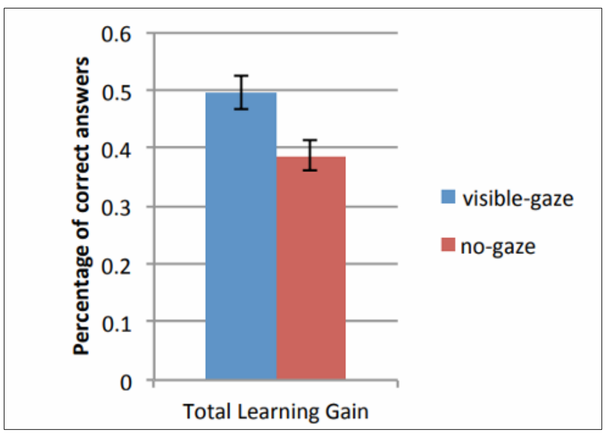

Figure 2.3 A simple bar chart illustrating performance data (B. Schneider & Pea, 2014)

Some alternatives to bar charts can be used to illustrate more complex data. Ahn, Gubbels, Yip, Bonsignore, and Clegg (2013) applied a bar chart to depict student interaction activities (log data). In the study, students contributed elements of inquiry, such as questions, hypotheses, and projects to collaboratively create science projects. Researchers used different colors to represent primary activity types, and secondary actions such as topics, resources, and feedback were represented using the same color with different brightness (Figure 2.4). From the visualization, people can easily identify the different types of activities that a student used to solve a problem.

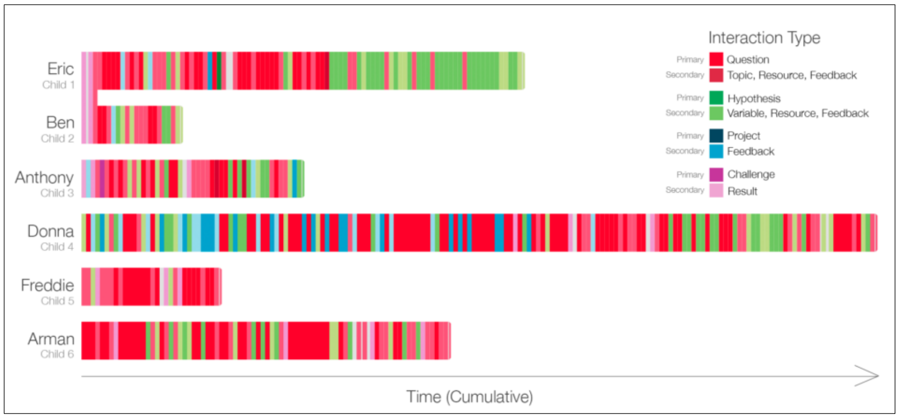

Figure 2.4 An alternative format of the bar chart to illustrate learner activities (Ahn et al., 2013)

A *Line chart* is another popular visualization method that can be used to visualize many types of learning data. For example, Dyckhoff, Zielke, Bültmann, Chatti, & Schroeder (2012) used a line chart to present the number of students accessing different areas of the system (Figure 2.5).

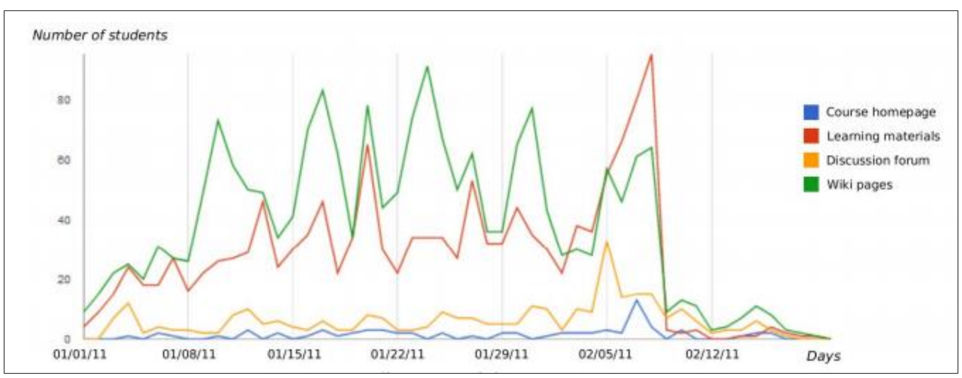

Figure 2.5 A line chart depicting student activity areas (Dyckhoff et al., 2012)

A *Pie chart* is a good visualization method to present percentages. Researchers use it to present the percentage of correctness (performance data) or the percentage of learning material received (log data). Jacovina et al. (2015) used a pie chart to depict the percentage of learning materials a student used (Figure 2.6).

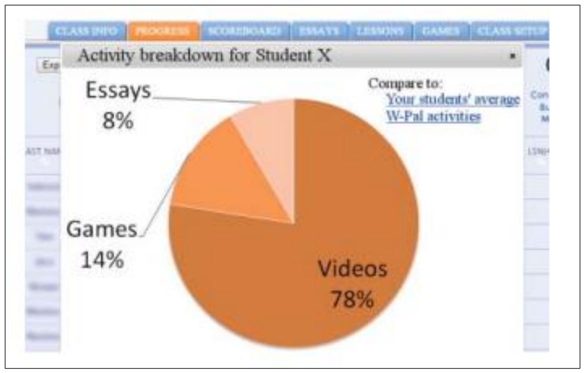

Figure 2.6 A pie chart illustrating the activity breakdown for a student (Jacovina et al., 2015)

Heatmap tables are often used to present student learning progresses and repeat times (Jugo, Kovačić, & Slavuj, 2015; Jacovina et al., 2015). Figure 2.7 is a heat map table (Jugo et al., 2015). The number in each cell represents the number of repetitions and the shades of color corresponds to the number. The larger the repeat number, the deeper the color. The heat map table enables teachers and researchers to easily identify the progress of students and the tasks that are repeated by many students.

Figure 2.7 Heatmap Table (Jugo et al., 2015)

There are some other visualization formats that were used to represent students' learning data. Figure 2.8 is a visualization that depicted the learning paths of students (Jugo et al., 2015). Figure 2.9 is a spider diagram which illustrates a learner's learning profile change after training through a project. "The shaded blue shows the initial profile of the student, while the red outer red profile indicates the 'stretch' on certain dimensions later in the learning project" (Shum, 2012, p.5).

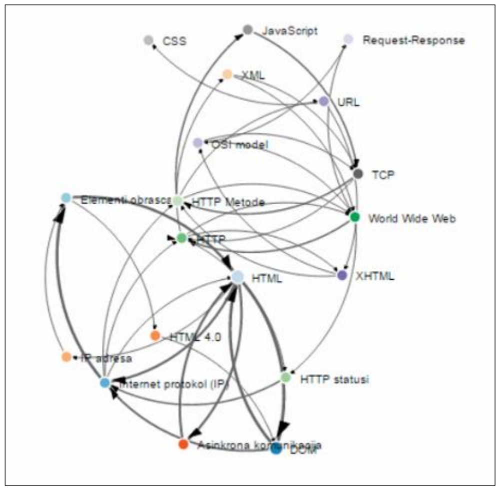

Figure 2.8 Students' Learning Path (Jugo et al., 2015)

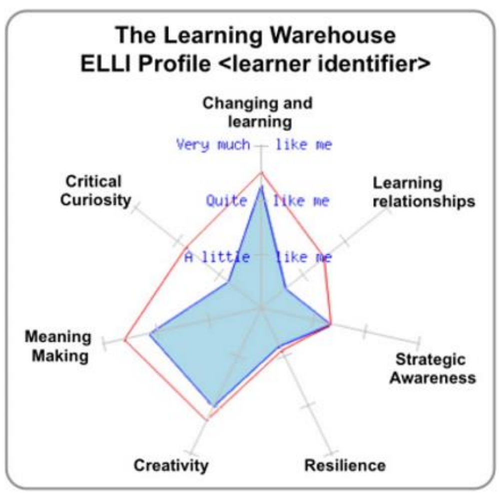

Figure 2.9 A spider diagram depicting a student profile (Shum, 2012)

The charts used in the visual learning analytics of log and performance data are relatively simple. However, efficient use of these charts can be used to develop a powerful visualization system to present student learning behavior and facilitate teacher comprehension and efficient teaching. The “eLAT” learning analytics toolkit is an example. Dyckhoff et al., (2012) developed an integrated visualization system “eLAT” for their learning management system (LMS), which enables teachers to explore the use of learning material, user performance, user properties, as well as user behavior (Figure 2.10). Developers considered the usability, usefulness, extensibility and reusability factors, when adopting charts to illustrate specific learning data. The heuristics evaluation of the system proved the efficiency of “eLAT” in presenting the behaviors and performance of students to teachers.

Figure 2.10 eLAT User Interfaces (Dyckhoff et al., 2012)

2.4.2 Eye Movement Data Visualization

Eye movement data is rarely used in visual learning analytics or visualizations that are designed to present eye movement data to educators. I will, therefore, reference eye movement visualizations used in academic research.

First, one-dimensional eye movement data, such as fixation duration, fixation count, saccade duration, and saccade count can be easily visualized by simple charts (such as a bar chart). The real visualization challenge is how to present the multi-dimensional eye movement data, such as scan-path. How to visualize the spatial and temporal attributes of eye movement data is the key to eye movement data visualization.

"*AOI rivers*" is a visualization format that is widely used in visualizing eye movement data. It is applied to depict where users are looking at a specific time. Burch. et al. (2013) used "AOI rivers" to show the distribution of attention of multiple subjects to different areas of the screen and the transitions between them.

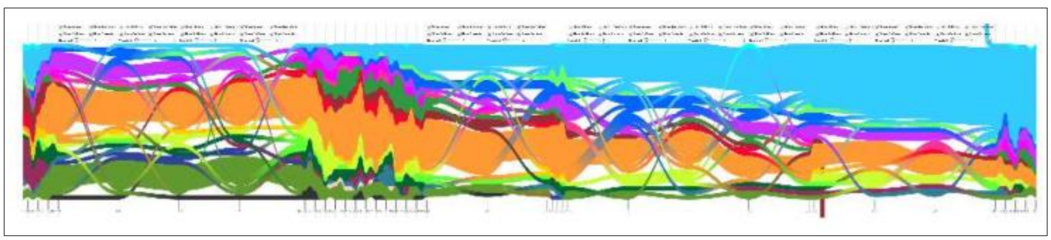

Figure 2.11 AOI River (Burch et al., 2013)

"Timeline Visualization" is another method for visualizing eye movement data. It precisely presents the temporal attribute of eye movements. However, it uses areas of interest (AOIs) to generally present spatial information. For example, in Figure 2.12, the x-axis is time, and the colored bars on the timelines indicate when AOIs are visible. The histograms inside the bars indicate how many participants are looking at the AOI (Kurzhals et al., 2017). Figure 2.13 is another timeline visualization format, which can be used to depict individual participant's eye movements. In Figure 2.13, the x-axis still represents time. Many timelines are stacked together. Each timeline belongs to one individual participant. The colored bars represent AOIs. From this figure, we can clearly see how one participant's fixation moves among AOIs (Kurzhals et al., 2017).

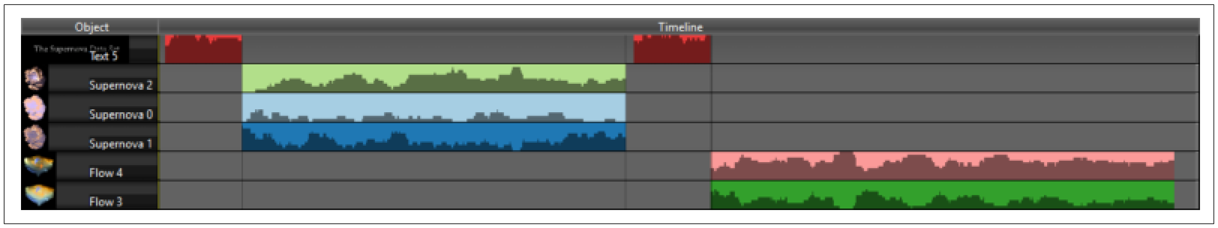

Figure 2.12 Timeline visualization for participant group (Kurzhals et al., 2017)

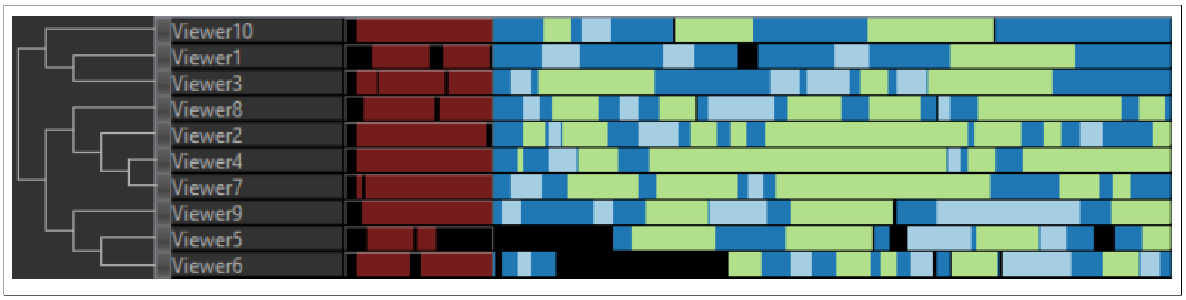

Figure 2.13 Timeline visualization for individual participants (Kurzhals et al., 2017)

These visualizations can present a subject's attention smoothly change over time. However, there is still a lack of eye-tracking data visualizations for student problem-solving procedures, problem context, performance correctness, and student attention simultaneously. Better eye movement data visualizations that fit the learning context are necessary.

2.5 Summary

Most visual learning analytics are conducted for research purposes. Little work has been done to bring visual learning analytics tools into classroom settings (Vieira et al., 2018). Analyzing and visualizing learning data of elementary school students is important for teachers and researchers to understand the learning behavior and learning characteristics of students. Visual learning analytics is also important for educators to provide timely intervention and to make the best use of CAL programs to supplement their instruction. Therefore, learning data analysis and visualizations should be proposed for both primary education and special education. Easy-to-understand eye movement visualizations that can show educators more about the learning context are useful for these ends.

CHAPTER 3. FRAMEWORK AND METHODOLOGY

3.1 Introduction

This chapter outlines the theoretical framework and methodology of the dissertation study. It first introduces the study's research approach and the theoretical framework of problem-solving strategies. Then, the visualization system development process is presented, including CAL learning data collection, learning data analysis, and the iterative process of visualization design. Finally, this chapter introduces the design and execution of the visualization evaluation experiment.

3.2 Research Approach

The research approach for this study is：

1. Identify possible behavior patterns that connect to problem-solving strategies in the math literature. In other words, for a specific problem-solving strategy, the student will be more likely to follow specific mouse/eye interaction patterns with the computer. This research focuses on identifying possible problem-solving patterns while a student reads math problems in a CAL program.

2. Based on the possible behaviors, the goal is to determine what data should be employed in the study, e.g., performance data, eye movement data, or mouse clicks/trajectory. This step is to find out what data we should use to reflect the problem-solving patterns.

3. With this data, a visual analytic approach can be developed to help educational researchers see a student's problem-solving strategies and difficulties. Descriptive statistical methods (e.g., mouse movement speed, fixation duration) are used to prepare the data for visualizations.

4. Develop an effective visualization system by:

a. Identifying appropriate visualization methods, including bar charts, heatmaps, scan paths, etc. Designing appropriate visualization dashboards.

b. Employing statistical measurements such as average, percentage, and standard deviation in the visualization system to provide accurate data values for users.

c. Involving educational experts in the design process as a pilot evaluation to iteratively improve the visualization design.

5. After the visualization development, target users were invited; educational researchers were asked to evaluate the visualization prototype. The purpose of the evaluation is to examine the usability of the visualization system to see compared with video, whether the visualization system can help users identify students' problem-solving strategies and difficulties quickly and accurately.

3.3 Theoretical Framework of Mathematical Problem-Solving Strategies

To understand students' mathematical problem-solving process, it is essential to understand the problem-solving strategies that students often employ. According to the literature review, there are two types of mathematical thinking associated with arithmetic word problems involving solving, sequential and holistic (Polotskaia, Savard, & Freiman, 2015). Researchers pointed out that successful problem-solving requires interplay of both ways of thinking (Sfard, 1991). Sequential thinking enables problem-solvers to understand a situation as a process/event while holistic thinking enables problem-solvers to understand the situation as "a system of relationships or a structure" (Polotskaia, Savard, & Freiman, 2015, p.254).

Different problem-solving strategies are caused by two mathematical thinking types. Hegarty, Mayer, and Monk (1995) contrast two general strategies: direct-translation strategies and problem model strategies. Direct-translation strategies involve a short-cut approach in which the problem-solver "attempts to extract the numbers in the problem and key relational terms (such as "more" and "less") and develops a plan that involves combining the numbers in the problem using the arithmetic operations that are primed by the keywords" (p.19). Model-based problem-solving strategies are different from direct-translation strategy. Using model-based problem-solving strategies, problem-solvers translate the problem statement into a situation model, which is an object-based representation rather than text-based representation in the direct-translation strategy. In this case, students not only look at numbers and keywords but also focus on background information such as pronouns to build problem model (Hegarty, Mayer, & Green, 1992). Also, the study by Hegarty, Mayer, and Monk (1995) revealed that unsuccessful problem-solvers reexamined numbers and keywords significantly more often than successful problem-solvers.

Moreover, some teachers have noticed students solving math word problems linearly, which means they drag/fill numbers in the equation in sequence without any effort to comprehend keywords nor understand the problem. That is another method to solve math problems, which is an error-prone method. ("Why are math word problems SO difficult for elementary school children?" n.d.).

Based on mathematical thinking theory and previous studies, this study summarizes four types of problem-solving strategies and their possible reflections in eye movement and mouse movement data patterns.

The first strategy is the model-based problem-solving strategy. In this stage, students are able to understand the problem as a complete structure and build the corresponding situation model. Students who use the model-based problem-solving strategy tend to pay attention not only to numbers and keywords (relationship words) but also spend time to comprehend background information such as pronouns (variables) (Hegarty et al., 1995).

The second strategy is the direct-translation strategy (keyword strategy). Students may identify keywords, such as 'total' and 'left', and decide based on the keywords which number is whole and which number is a part. In this case, students may pay much more attention to the number tags and keywords, especially in their revisits.

The third strategy is linear drag, which means students drag the tags into the diagram equation in sequence without mathematically thinking. The combination of performance data, eye movement, and mouse movement data may reflect the linear drag strategy. For example, the situation that a student drags number tag sequentially and quickly, with little attention to the problem content may, indicate the student drag tags linearly without thinking.

The fourth strategy is guess and check, in which students solve mathematical problems by guessing the answer and then checking the guess with feedback (Taspinar & Bulut, 2012). In this case, the student may directly use guess and check strategy, or the student may have tried to understand the problem (model-based problem-solving strategy) but failed, then he/she resorted to a guess and check strategy to finish the problem.

In real problem-solving tasks, students may combine multiple strategies to solve the problem. For example, students may use model-based problem-solving strategies to understand the problem. After they have a comprehensive understanding of the problem, they may employ keyword information to answer the problem, which is the keyword strategy. Also, some students may use a model-based problem-solving strategy at first, but meet difficulties in understanding the problem, inhibiting the building of a problem model. Then, they may turn to a guess and check strategy to make the work done.

3.4 Pilot Study of Student Eye Movement

In this dissertation study, students' eye movement was collected when they solve the arithmetic problems from a CAL program. The CAL program is called Conceptual Model-Based Problem Solving (COMPS-A$^{©}$, Xin, Kastberg, Chen, & Team, 2015-2020) tutor. It was developed to improve additive mathematics problem-solving of 2nd /3rd grade students with difficulties in mathematics (Xin, 2012).

In the literature review, specific students' eye movement patterns have been detected corresponding to the performance of students. The authors demonstrated that the patterns were related to specific students' problem-solving strategies (Tai, Loehr, & Brigham, 2006; Moutsios-Rentzos & Stamatis, 2015). Based on the literature review, in this dissertation study, I assumed that the eye movement data of students could reveal student problem-solving patterns and indicate their strategies. I included eye movement data in my visualization system is based on the assumption. Here I conducted a brief pilot study to validate the assumption.

There are two modules (module A and module B) in COMPS-A$^{©}$ (Xin, Kastberg, Chen, & Team, 2015-2020) tutor. Module A's teaching focus is number concepts, counting strategies, and small number addition and subtraction. Module B's teaching focus is mathematical word problem-solving.

I invited three third-grade students to take tasks from module A and module B, respectively, to check whether these students' eye movements can reflect their visual attention distribution and patterns in their problem solving process. Figure 3.1, Figure 3.2, and Figure 3.3

are screenshoots from Module A of COMPS-A© computer program (Xin, Kastberg, Chen, & Team, 2015-2020). The tasks from module A ask students to solve two digits addition or subtraction word problems. The task in Figure 3.1 is a typical task from module A. In the task, students were asked: "Before you have 83 blocks. After you have forty-three blocks, how many blocks were taken away?" Students have two chances to answer the question. In the first try, forty-three blocks are shown on the "After" side. On the "Before" side, there is only number 83 without blocks. If the student fails, he/she will have another chance to solve the problem. But this time, a "playground" (Figure 3.1) is popped-out on the "Before" side, and there is a "ten" and a "one" button on the screen. Students can click the "ten" button to put ten blocks into the "playground" and click the "one" button to put one block into the playground. He/she will be asked to put 83 blocks into the playground. The playground is designed to help students get a concrete concept of 83 and compare 83 blocks to 43 blocks. The students who can use numbers and algebra to solve the problem on the first try will not see the "playground" and count blocks. But students who are not able to calculate using algebra should count blocks in the "playground". When students bring blocks into the "playground", they need to know how many blocks they already brought and how many more blocks they need to bring.

Solving the problem in Figure 3.1 is a long process from reading to answer. But, in the pilot study, the goal is merely to inspect whether students' counting strategies can be reflected in their eye movement. Thus, I only need to observe students' eye movements in the counting period. Students' eye movement data were captured using Tobii Pro X3-120 and stored in Tobii Pro bundled software – Tobii Pro Studio, which enables users to observe participants' eye movement, divide eye movement recordings into segments, visualize the results, define the area of interest (AOI), even export the eye movement dataset as excel files.

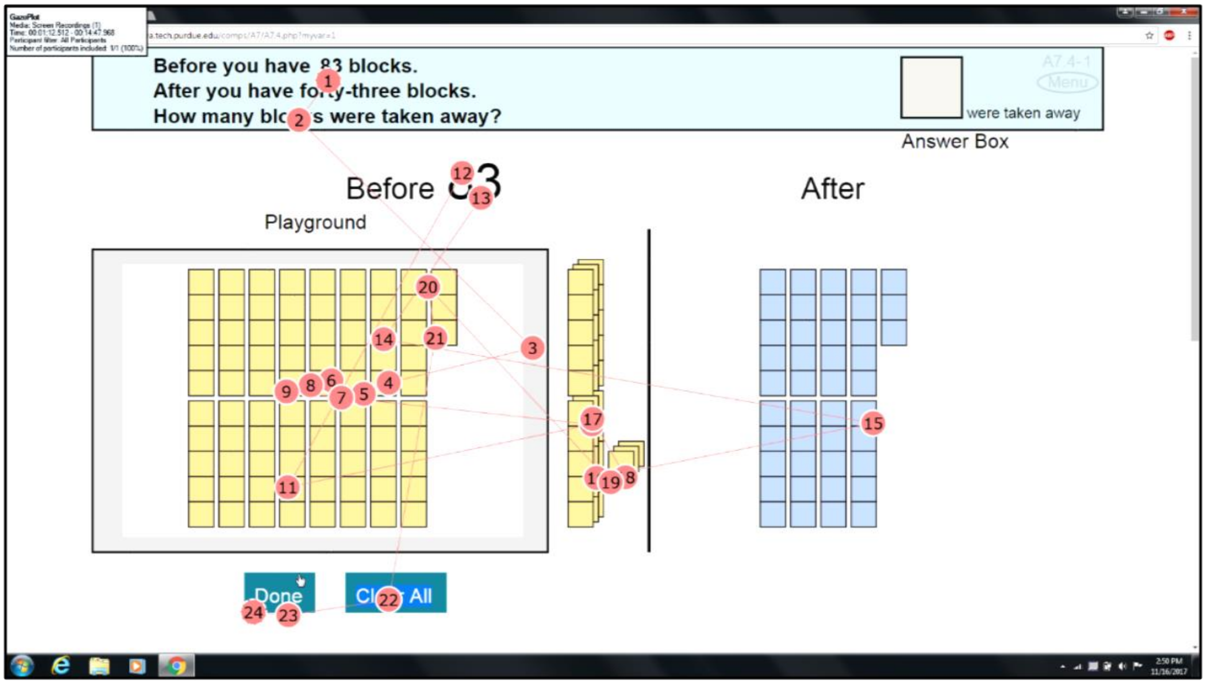

Figure 3.1 A task from COMPS-A© module A (Xin, Kastberg, Chen, & Team, 2015-2020)

As Tobii Pro Studio enables users to divide problem-solving processes into segments, aggregate and visualize students' eye movement within the segments, I directly employed the segment visualization function of Tobii Pro Studio and exported students' eye fixation gaze plots in their counting periods (Figure 3.1, Figure 3.2 and Figure 3.3.). Not all students had counting periods in their problem-solving processes, as I mentioned, some students directly answered the questions using algebra. For the students who counted blocks, there are counting periods.

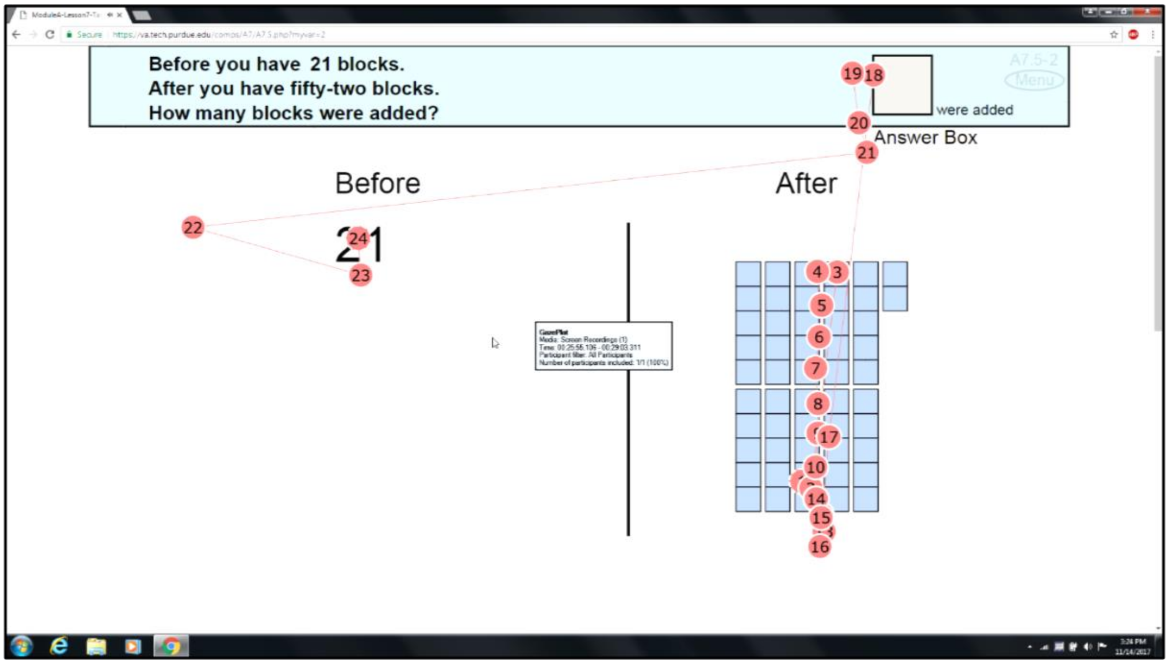

Figure 3.2 Count by One (COMPS-A© Xin, Kastberg, Chen, & Team, 2015-2020)

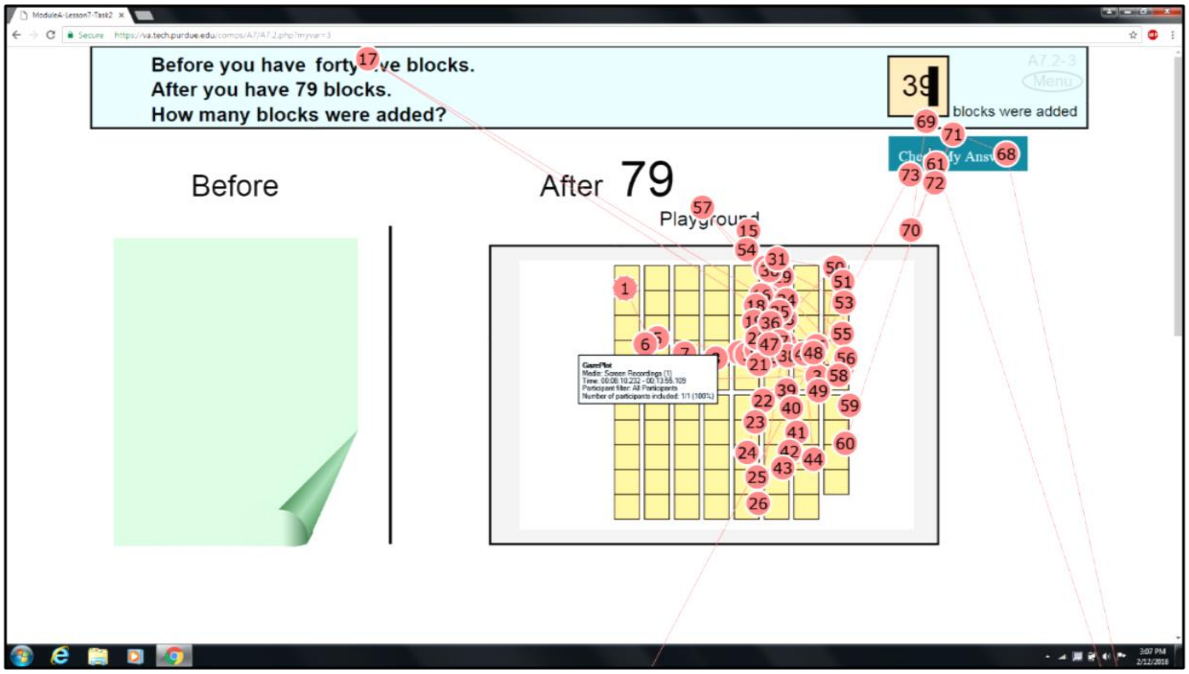

Figure 3.3 Count by Ten and One (COMPS-A© Xin, Kastberg, Chen, & Team, 2015-2020)

Figure 3.1, Figure 3.2, and Figure 3.3 illustrate the different counting approaches adopted by students. The dots in figures represent students' eye fixations, and the numbers in the dots represent the sequence of fixations. Figure 3.1 shows that the student first looked at the problem content. Then he looked at the "playground" to count the blocks. From fixation 4 to fixation 9, each fixation is located on one column of blocks, which is a group of ten blocks. Then the student's visual attention shifted to the other areas of the screen and randomly jumped back to the "playground". The strategy that the student applied to solve the problem may still be vague, but it is clear that the student counted the blocks by ten. While in Figure 3.2 the student eye fixations are located on individual blocks in sequence (from fixation 3 to fixation 16), which indicates the student counted the blocks one by one. Also, observing the eye fixation sequences in Figure 3.3, it appears the student counted blocks by ten and one.

I also explored students' visual attention distribution when they read the arithmetic problem, to check whether their eye fixation patterns are consistent with the findings in the literature. According to Hegarty et al. (1995), low-performance students focus more on keywords in the arithmetic word problem while high-performance students also focus on background and comparison words. These patterns were also found in the pilot study's eye movement data. Figure 3.4 shows that the student read every word of the problem while the student in Figure 3.5 only looked at keywords.

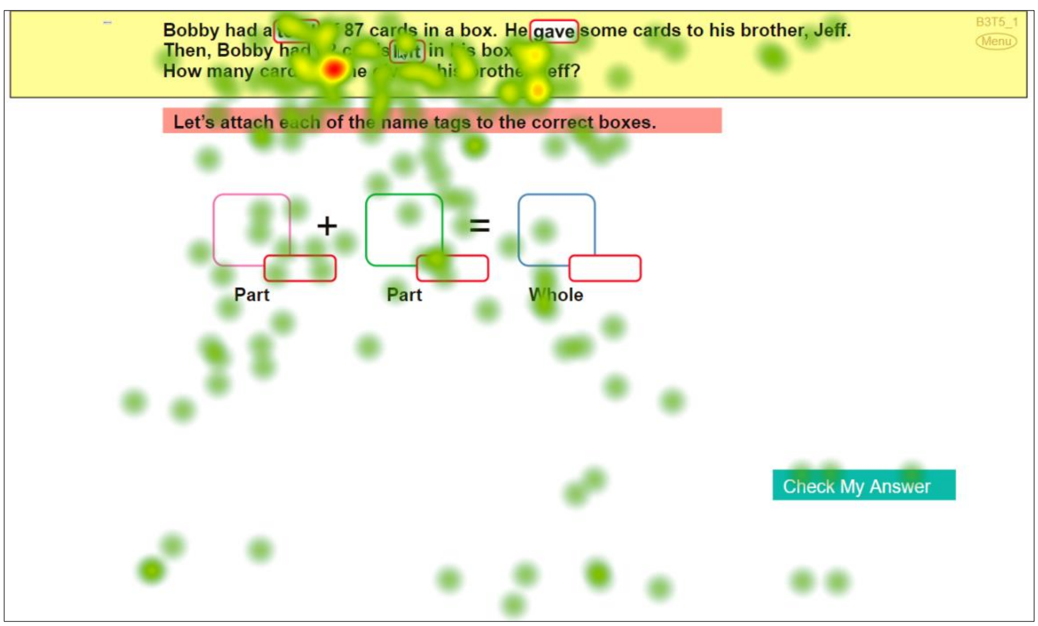

Figure 3.4 Read most part of the problem (COMPS-A© Xin, Kastberg, Chen, & Team, 2015-2020)

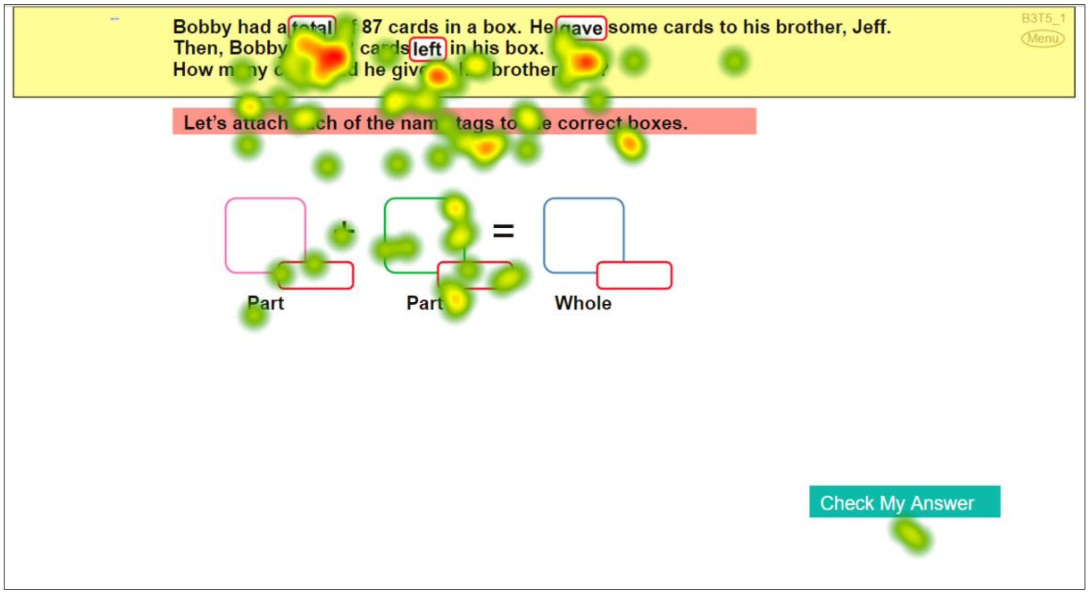

Figure 3.5 Read keywords (COMPS-A© Xin, Kastberg, Chen, & Team, 2015-2020)

The pilot eye movement study showed the eye movement's function in reflecting students' visual attention and problem-solving patterns. The eye movement patterns associated with student problem-solving approaches that found in the previous studies have also been detected in the pilot study. The pilot eye movement study provided the empirical foundation of eye movement visualization.

3.5 Students' Problem-Solving Data Collection

From the theoretical framework of problem-solving strategies, we can see that students' visual attention, operation, and performance are critical data that can reflect their problem-solving processes. So, I first collected this data in the CAL programs.

3.5.1 Mathematical Problems from COMPS-A© program

The experiment problems come from Comps-A© (Xin, Kastberg, Chen, & Team, 2015-2020) module B. There are two types of problem-solving questions: part-part-whole and comparison in module B. For each problem type, one demo, one practice question, and two test questions are chosen. So, participants took two demos, two practice problems, and four test problems in total (table 3.1). The criteria to choose the test questions are: 1. Students can't get the correct answer if they drag name tags /number tags to the equation in sequence (e.g., Drag the first tag to the first box, drag the second tag to the second box). 2. Students can't get the correct answer by using the keyword (e.g., using plus when seeing "more"). These test question choosing criteria may, to some extent, improve the probability of finding problematic behavior/problem-solving strategies/struggles.

Figure 3.6 is an example of the experiment's math problems. A question will first be read to students. Then, instruction will be provided to students. For example, in the task illustrated by Figure 3.6, the system will read aloud the instruction as: "This problem has three parts: the total number of animals, the number of chickens, and the number of pigs." After this instruction, students will be asked to drag the name tags or number tags to the part-part-whole equation/comparison equation. After they are done, they need to click the "check my answer" button to get immediate feedback (correct/wrong). If they are incorrect, they will get another chance to solve the problem. If they are wrong again, the feedback demo will show them the correct answer.

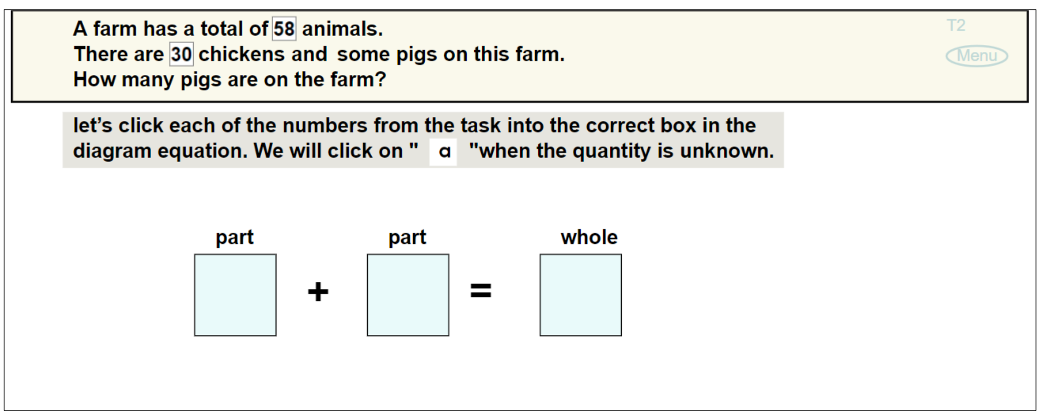

Figure 3.6 An illustration of the math problem-solving tasks (COMPS-A© Xin, Kastberg, Chen, & Team, 2015-2020)

Table 3.1 Experiment Problems (Xin, 2012)

ID	Question Content
Demo 1 (Part-Part-Whole)	Travis ordered 68 baseball cards from a magazine for himself. Then ordered some more for his brother. In all, he ordered 97 baseball cards. How many did he order for his brother?
Demo 2 (Comparison)	Wendy has 40 marbles. Taylor has 93 marbles. How many more marbles does Taylor have than Wendy?
Practice 1 (Part-Part-Whole)	David delivers pizzas in his spare time. Today, he delivers 21 pizzas in the morning. Then, he delivers some pizzas in the afternoon. In all, he delivers 37 pizzas. How many pizzas does David deliver in the afternoon?
Practice 2 (Comparison)	Lucas has 30 stamps. Lucas has 44 fewer stamps than Ben. How many stamps does Ben have?
Test 1 (Part-Part-Whole)	Bobby had a total of 87 cards in a box. He gave some cards to his brother, Jeff. Then, Bobby had 62 cards left in his box. How many cards did he give to his brother Jeff?
Test 2 (Part-Part-Whole)	A farm has a total of 58 animals. There are 30 chickens and some pigs on this farm. How many pigs are on the farm?
Test 3 (Comparison)	Patrick has 89 sports cards. Patrick has 42 more sports cards than Joy. How many sports cards does Joy have?
Test 4 (Comparison)	Lauren has 14 pencils. She has 26 fewer pencils than Brenna. How many pencils does Brenna have?

3.5.2 Data Collected

Student performance data, including their answer and time spent on each problem, was recorded. Additionally, student interaction data, including eye movement data and mouse movement data, was collected.

Eye movement data were captured using Tobii Pro X3-120 and stored in Tobii Pro bundled software – Tobii Pro. The eye movement data, such as fixation, x- and y-coordinates (in pixels), timestamps, and fixation duration (in milliseconds), were registered. Each student's eye movement data were exported from Tobii Studio and imported into a MySQL database. The areas that students expected to focus on were defined as AOIs, and the coordinates of AOIs were recorded in the database too.

Student log data was registered in a SQL database in real-time. The information registered in the database included user ID, task content, student action, feedback, segment, and timestamp (Figure 3.7). For example, the prompt start and the prompt end define a reading segment. When a student drags a name tag to the equation box, the database registered "Drag tag_x into part_box_x".

←T→				id	userid	task	prompt	action	feedback	segment	timestamp
☐	Edit	Copy	Delete	22901	34	B3.5_1				prompt start	2017-12-04 15:02:20
☐	Edit	Copy	Delete	22902	34	B3.5_1	Bobby had a total of 87 cards in a box. ...				2017-12-04 15:02:20
☐	Edit	Copy	Delete	22903	34	B3.5_1	Let us attach each of the name tags to the correct...				2017-12-04 15:02:21
☐	Edit	Copy	Delete	22918	34	B3.5_1				prompt end	2017-12-04 15:03:03
☐	Edit	Copy	Delete	22919	34	B3.5_1				operation start	2017-12-04 15:03:03
☐	Edit	Copy	Delete	22923	34	B3.5_1		Drag tag_3 into part_box_3			2017-12-04 15:03:06
☐	Edit	Copy	Delete	22940	34	B3.5_1		Drag tag_2 into part_box_2			2017-12-04 15:03:46
☐	Edit	Copy	Delete	22941	34	B3.5_1		Drag tag_1 into part_box_1			2017-12-04 15:03:50
☐	Edit	Copy	Delete	22942	34	B3.5_1				operation end	2017-12-04 15:03:51
☐	Edit	Copy	Delete	22943	34	B3.5_1				feedback start	2017-12-04 15:03:51
☐	Edit	Copy	Delete	22944	34	B3.5_1			correct		2017-12-04 15:03:51
☐	Edit	Copy	Delete	22960	34	B3.5_1				feedback end	2017-12-04 15:04:21
☐	Edit	Copy	Delete	22961	34	B3.5_1		click next button			2017-12-04 15:04:21

Figure 3.7 Log Database

Mouse and keyboard data (mouse over, mouse click, keypress) were recorded in another table of the database. The table recorded ID, user ID, action, task, value, timestamp, x, and y (Figure 3.8). The action recorded user action such as mouseover, click. The value column recorded which element students clicked or which button students pressed. The x and y columns recorded mouse's coordinates on the screen. The press button operation had no coordinate, and so the database put 'null' for the x- and y-coordinates.

←T→				ID	userid	task	action	value	timestamp	x	y
	Edit	Copy	Delete	255367	35	B3.5_1	mouseover	null	2017-12-12 15:03:17	1025	447
	Edit	Copy	Delete	255368	35	B3.5_1	mouseover	null	2017-12-12 15:03:17	1060	450
	Edit	Copy	Delete	255369	35	B3.5_1	click	svg	2017-12-12 15:03:18	1177	429
	Edit	Copy	Delete	255370	35	B3.5_1	mouseover	null	2017-12-12 15:03:18	1110	172
	Edit	Copy	Delete	255371	35	B3.5_1	mouseover	null	2017-12-12 15:03:18	1103	122
	Edit	Copy	Delete	255372	35	B3.5_1	mouseover	null	2017-12-12 15:03:19	1077	31

Figure 3.8 Mouse and Keyboard Database

Students' log data, performance data, and eye movement data were recorded in real-time as students used the computer program.

3.5.3 Data Preparation

Learning data collected by the computer program were not in a form that could be directly analyzed and illustrated. Before visual analytics were applied, data preparation required a multistage, contextualized data processing. The data processing included five stages - cleaning, transformation, segment identification, user identification, and data integration. Data were cleaned, removing irrelevant data, and determining missing data. A student might redo a task that he/she already completed, which caused data duplicate. Or a student might skip a specific task as he/she thought he/she completed yesterday, which caused data missing. After cleaning the data, it was in the desired format. For example, the timestamp format of eye movement was transformed from month/date/year to month-date-year to make it consistent with the format of mouse movement timestamp. Visualization users need to know what happened in a specific period. For example, users may like to know when feedback popped-out and whether a student looked at the feedback content. So, learning data segmentation was needed. I identified problem-solving segments using log data, which marked the start and end time points of segments, such as the start timestamp of operation, the timestamp of submitting the answer. For student eye movement data, to ensure the data quality, recordings data with very low "Gaze samples" (below 30 percent) were excluded. "Gaze samples" here is a metric defined within Tobii studio. It calculated by dividing correctly identified gaze data samples to the number of gaze capture attempts. For example, 100% means one or both eyes are detected throughout the recording. 50% means one or both eyes were detected half of the recording. Low "Gaze Samples" may insdicate low eye tracking data quality. Moving gaze out of the screen can also lead to low “Gaze Samples”. I set a 20% threshold for “Gaze Samples” is because I found in the data collection procedure that students with off-task behavior tend to have lower than 30 percent “Gaze Samples”. I also checked whether there were eye fixations on the area of interests and excluded recordings with no fixations on areas of interest. Students’ eye movement data, mouse movement data, and performance data were stored in different tables. Different data types were joined through user ID and timestamp. After problem-solving data was prepared, I started to develop the visualization system. In the developing process, some data were transformed and processed again.

3.6 Visualization Development

In the visual learning data analytics literature, the research focus ranged from student behavior, student performance, assessment to curriculum development, and domain knowledge development (Peña-Ayala, 2014). This dissertation’s visualization was designed to help teachers and educational researchers understand students’ problem-solving processes and identifying students’ problem-solving strategies and difficulties. Visualization should present students’ problem-solving behavior patterns, which is related to the investigation of students’ metacognition and strategy use (Kinnebrew, Loretz, & Biswas, 2013). Based on the demand of users, I developed the visualization system.

3.6.1 Visualization Design

When designing the visualization, one must consider the data that needs to be presented (data dimensions and variables) (Keim, 2002), demands of target users, and tasks supposed to be completed (Chen et al., 2016). In this visualization design, there were four stages. The first stage was using traditional eye movement and mouse movement visualization methods – heatmaps and gaze plots to present student problem-solving processes. The second stage was employing basic shapes such as line, bar, and circle to present students’ eye movement and mouse movement data. The third stage was based on the graphic design developed in the second stage, employing interactions to build the visualization system. The fourth stage was prototype evaluation and design iteration.

3.6.1.1 First Stage: Heatmap and Gaze plot

Traditional visual attention or eye-tracking data illustration method is using heatmaps or gaze plots (Kabugo, Muyinda, Masagazi, Mugagga, & Mulumba, 2016; Tai et al., 2006). Considering the target users of the visualization are teachers and educational researchers who have little experience in data visualization. Heatmaps, which are simple and easy to understand, were employed to present students' eye movement data. Heatmaps couldn't present the data quantity clearly, so bar charts were employed alongside heatmaps. Student performance data, such as feedback received, were presented with heatmaps (Figure 3.9). Student eye movement trajectory was also represented using gaze plots (Figure 3.10). The line shows the eye movement path. The red dots are fixations, and the radius of dots represents the duration of fixations. Furthermore, mouse movement was illustrated using the mouse movement plot (Figure 3.11). The gray dots represent the positions that were moussed over. The red dots are mouse click points. Users can select which visualization they want to observe. Users can also select multiple tasks or users to compare visual attention or mouse movement difference among students (Figure 3.12).

But later, these visualizations were found to be hard to comprehend. The prototype evaluation showed that, although there were interactions, it was hard work for users to build connections among these heatmaps and plots. Different data sources should be combined to build a scenario pattern. But individual maps and plots differentiate data sources instead of combining them. To combine different types of data, the first step is to put these data into the same visualization graph, physically decrease the distance between the data types. Considering the limited space of the visualization interface, I didn't map data completely, as was done in the gaze plot. Data was aggregated and presented using simple graphs such as lines and dots. That's the second stage of the visualization design.

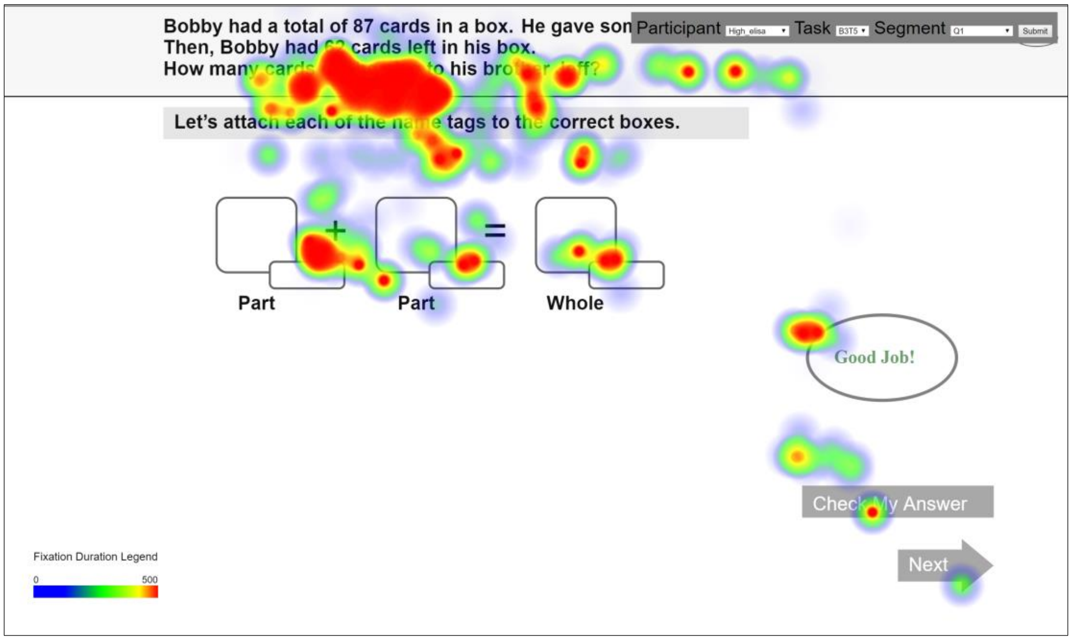

Figure 3.9 Heatmap of Eye Tracking Data

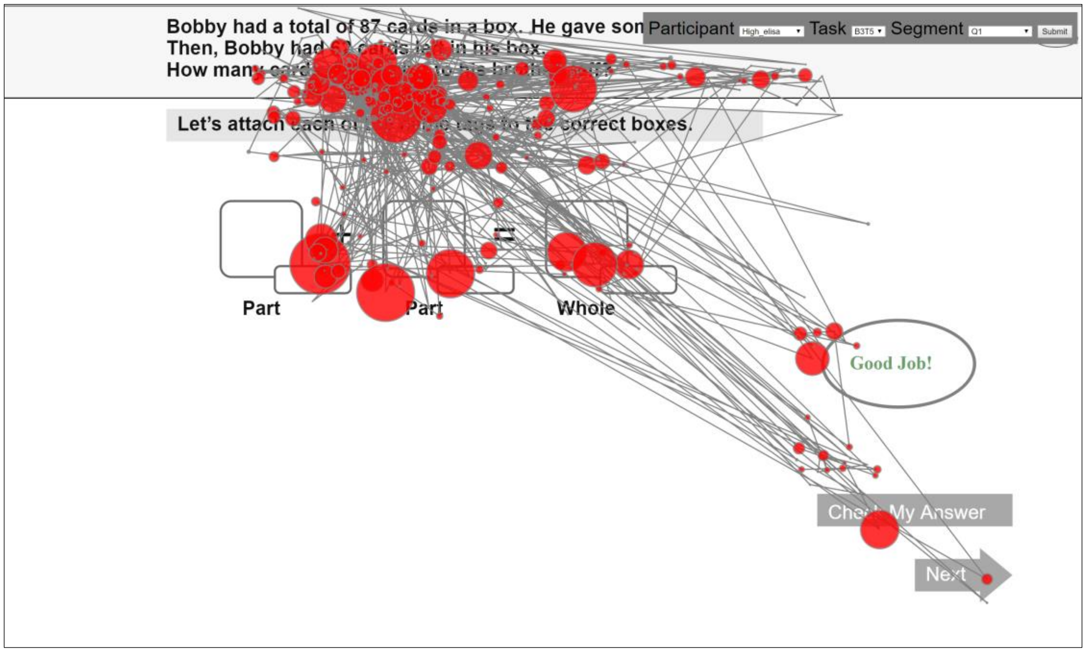

Figure 3.10 Gaze Plot

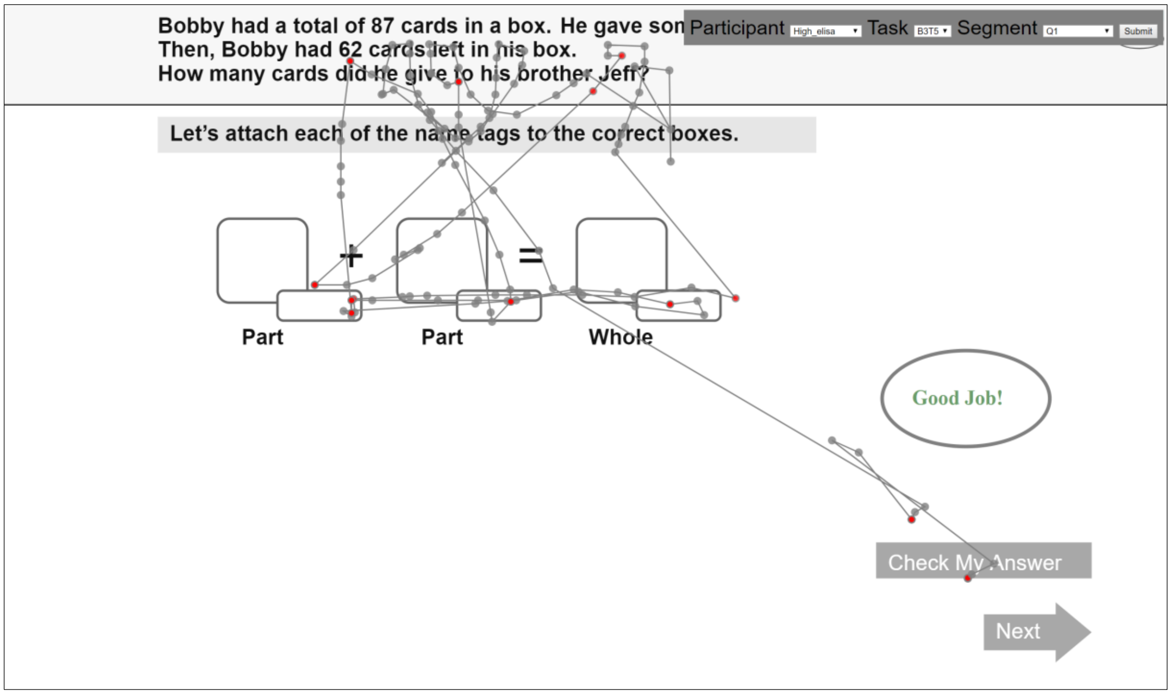

Figure 3.11 Mouse Movement Plot

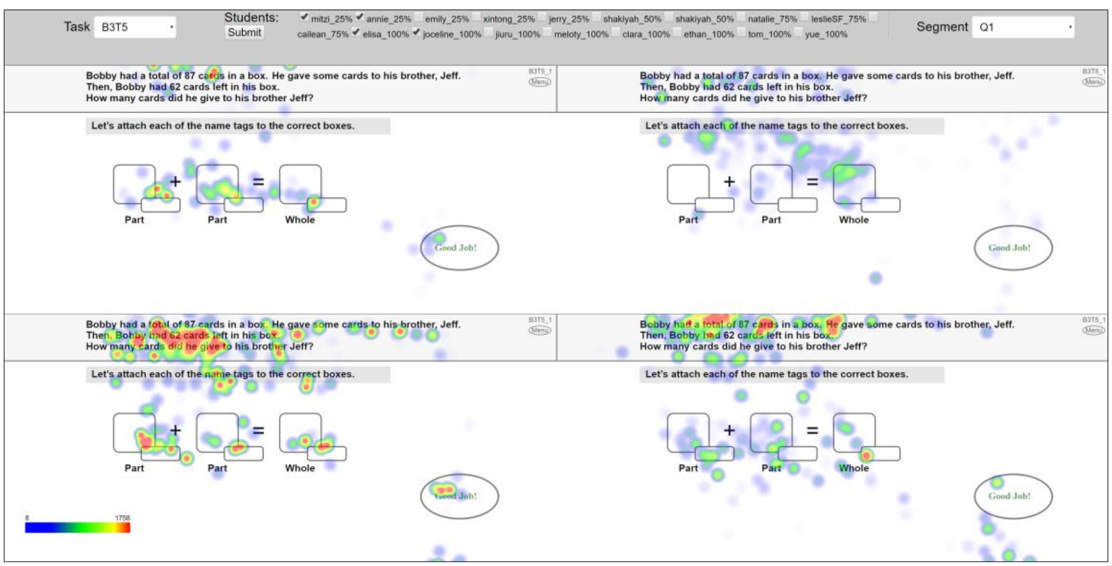

Figure 3.12 Interaction of the Heatmap Visualization

3.6.1.2 Second Stage: Employing Abstract Graphs

As discussed in the literature review, eye fixation duration, fixation count, and regression are important metrics that can track people’s visual attention and estimate people's cognitive processes (Lai et al., 2013; S. C. Chen et al., 2014; Susac, Bubic, Kaponja, Planinic, & Palmovic, 2014; Miller, 2015). Because of this, the visualization presents not only the duration, but also the number and sequence of fixations. In Figure 3.13, each word of the problem content is considered to be an area of interest, and word fixations are presented. Fixation duration is presented using the bars, while the fixations are illustrated using line segments. The line segments are connected with thick lines representing the sequence of fixations. One problem of this visualization is that, after visualizing eye movement data, there is no space for other data visualization, such as mouse movement and performance data, not to mention the unavailability of the problem-solving context.

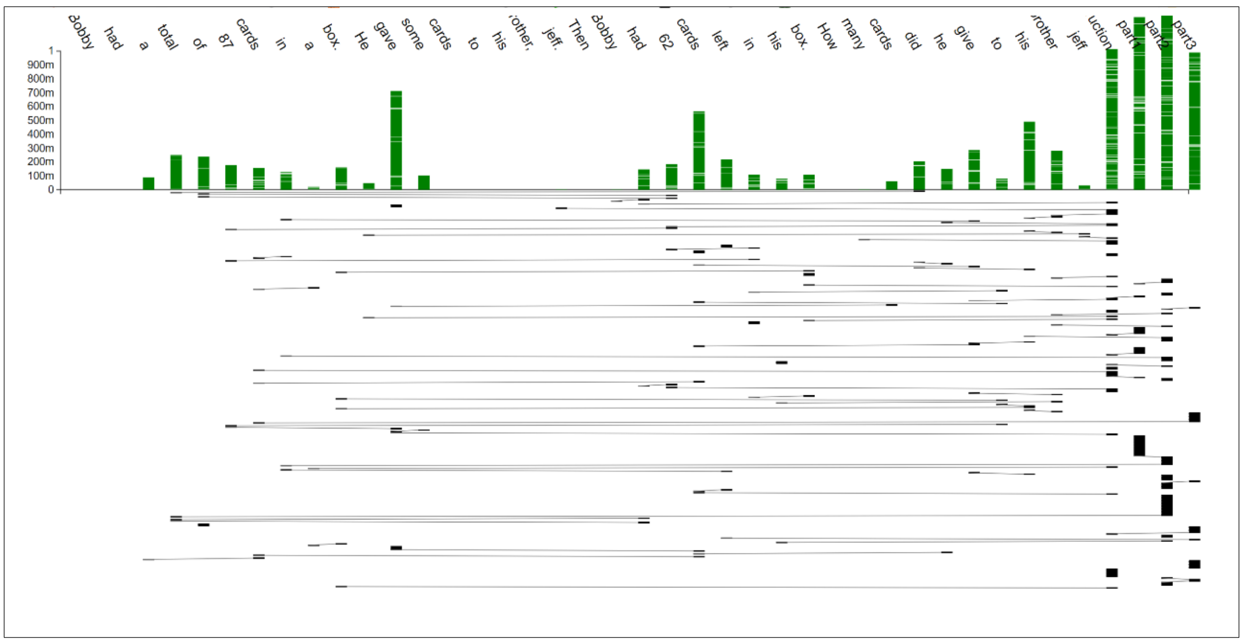

Figure 3.13 Fixation Visualization

It was found that, before presenting students' problem-solving data, it is important for users to understand the problem-solving context. Teachers and educational researchers need to know what the problem is that the students were asked to solve. After understanding the problem-solving context, then teachers and educational researchers can further understand students' problem-solving processes. So adjustments were made to the problem interface, illustrating problem-solving data within the context of the problem. In Figure 3.14, fixations were presented using bar charts. Each grid in the bar chart represents one fixation while the color depth represents the duration of the fixation. The deeper the color, the longer the fixation duration. In Figure 3.14, it can be observed that the student read every word in the problem and looked at the second part box of the diagram for a long time, ultimately failing the problem. Although in Figure 3.14, only eye movement data were presented, the problem context is clear and more data presentations could be added to the visualization. So, visualizing problem-solving data within the problem context is the correct direction. Then later, attempts were made to illustrate fixations using line segments and ellipses. Fixation sequences were also presented by ordering fixation lines. For example, in Figure 3.15, A rectangle above each word represents the total number of fixations located within the problem content, if there is one fixation locsated on the text, there will be a line segment in the rectangle. The fixation line segments are ordered according to temporal sequence in the rectangle. From bottom to top, the rectangle represents the beginning to the end of the problem-solving process. In Figure 3.16, the fixations are presented in the same way, but the rectangles were eliminated. Furthermore, the students' mouse operation is plotted in Figure 3.16.

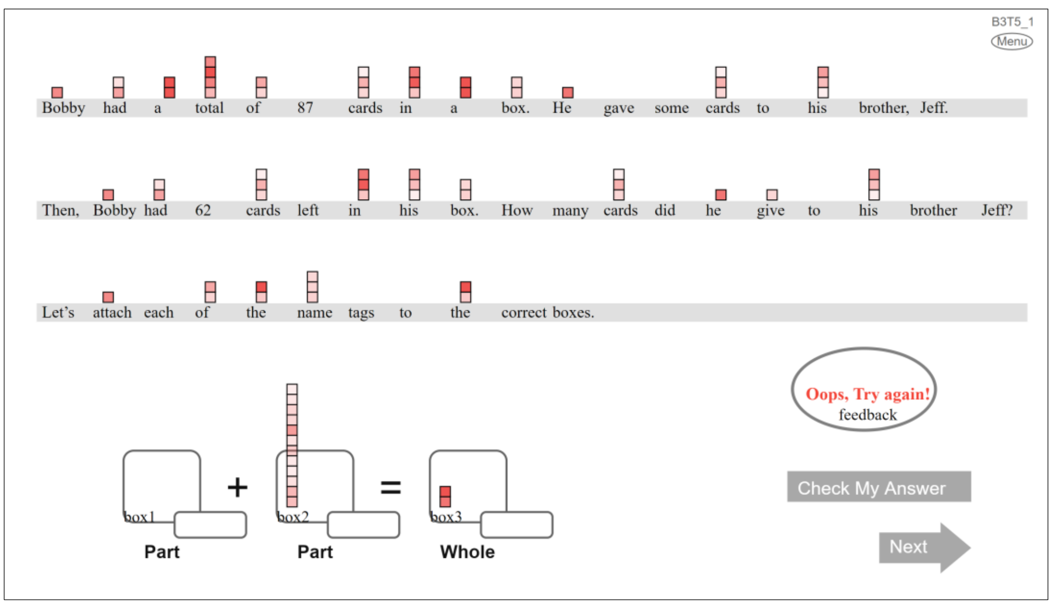

Figure 3.14 Visualizing eye movement data in problem csontext

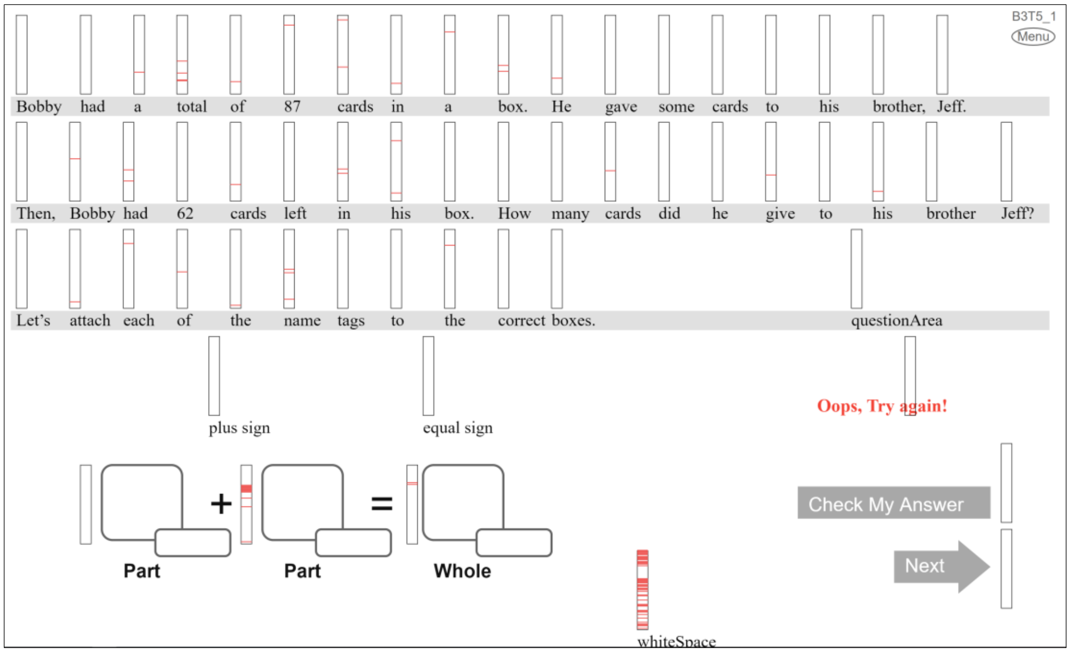

Figure 3.15 Visualizing fixations using line graph and presenting fixation sequences

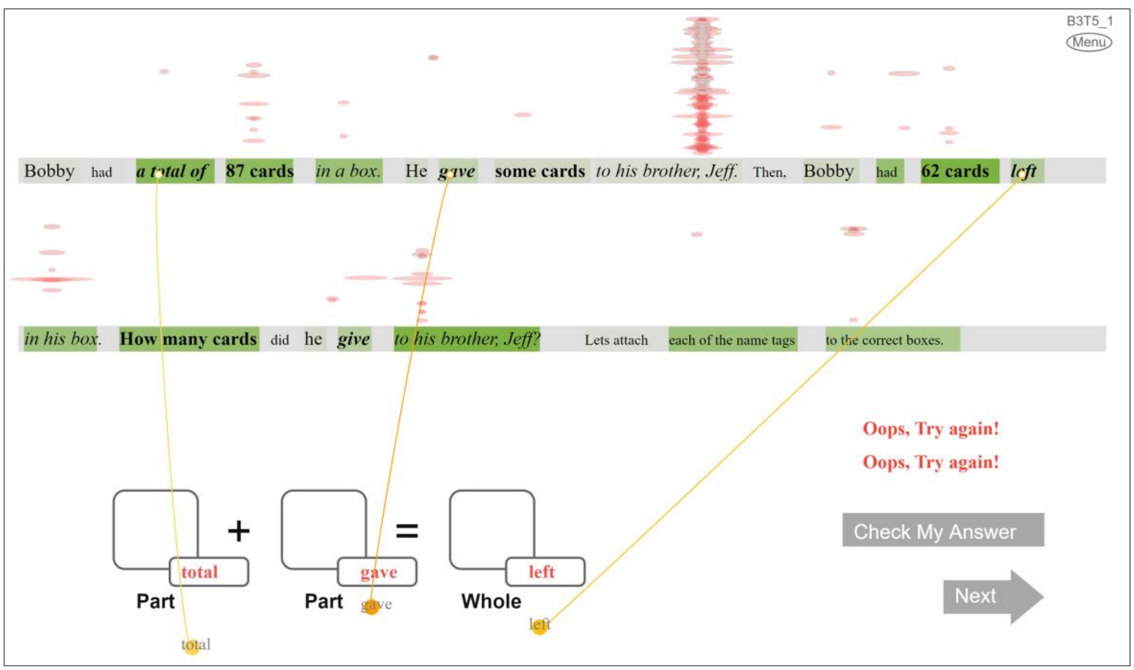

Figure 3.16 Visualizing mouse movement data

Figure 3.17 Performance Data Visualization Version 1 VS. Version 2

Students' performance data, including correctness and time spent on-task, were presented using bar charts. Two different versions of performance bar charts were designed (Figure 3.17). In the visualization, the color of the bar represents student correctness (red=incorrect, green=correct). The length of the bars represents the time spent on-task. For each problem, there were two steps. The first step was dragging name tags into the diagram equation (B3.5_1 in Figure 3.17). The second step was dragging number tags into the diagram equation (B3.5_2 in Figure 3.17). In each step, students could try twice. In Figure 3.17, some bars have two colors. It means the student was incorrect in the first attempt (red) and correct in the second (green). The second version of the performance bar chart was employed because it aligns the starting point of bars and separates the first and the second step bar, which makes the bar chart easy to compare and understand.

3.6.1.3 Third Stage: Visualization Prototype

In the third stage, performance bar charts, fixation visualization, and mouse movement visualization were integrated into a visualization system (Figure 3.18). Users could click on a bar in the performance bar chart (left side) to observe the corresponding students' eye movement, mouse operation patterns in the problem, and access insights within the data.

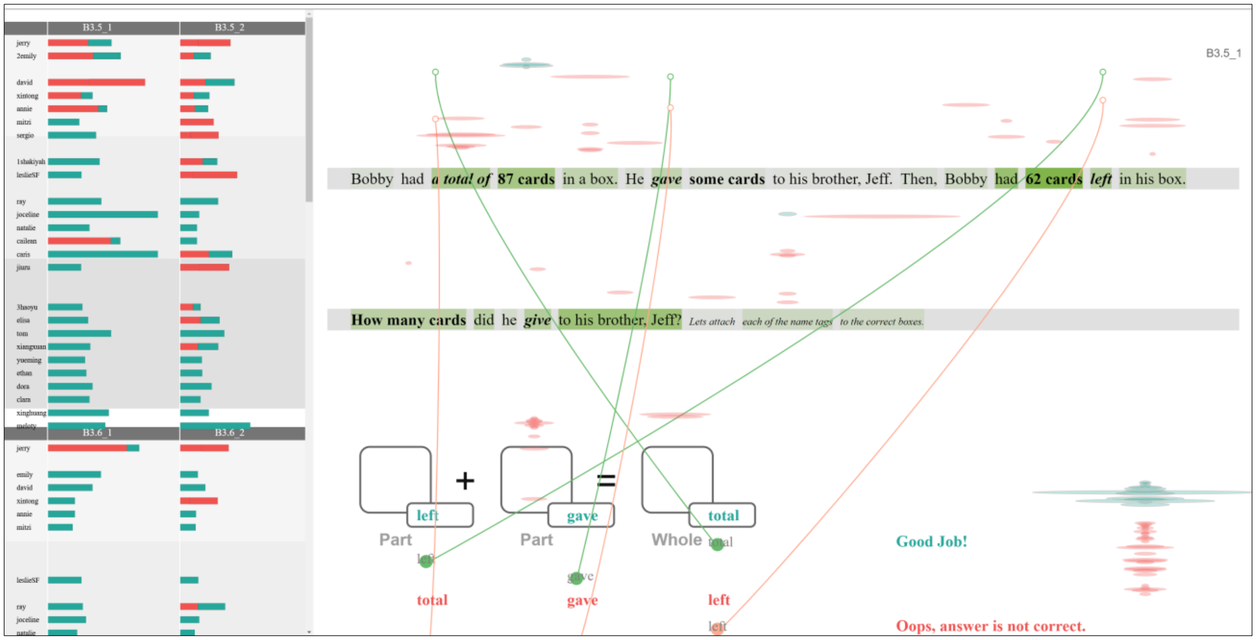

Figure 3.18 Visualization Prototype

3.6.1.4 Forth Stage: Prototype Evaluation and Design Iteration

Two educational researchers from the COMPS-A© (Xin, Kastberg, Chen, & Team, 2015-2020) project group were invited to review the visualization prototype. They were first- or second-year Ph.D. students in the special education department. They are familiar with the project background, the computer program, and the learning data. At the same time, they are the target users of the visualization who need to analyze the learning data of students. Because of the above reasons, they were asked to help evaluate the prototype and gave their suggestions and expectations of the visualization system. In addition to the educational researchers, one visualization designer who was a Ph.D. student in the Human Computer Interaction and visualization design direction was invited to evaluate the interface design of the visualization.

The evaluation lasted about 30 minutes. In the first 15 minutes, an introduction including study background, study purpose, the data depicted by the visualization, and the design of the visualization was given to the evaluators. Then the evaluators were asked to read a student problem-solving process using the visualization prototype. They were asked to speak out their thinking and give comments on the visualization. It took about 15 minutes. They went through the visualization functions and provided suggestions for the visualization system improvement, including:

1. Putting fixation lines under, instead of above, the text and letting users read fixation lines from top to bottom.

2. Consistent problem text font size and only using bold font on keywords. The prototype problem text had too many different font sizes, which could cause confusion.

3. For each problem, there were two steps. The two steps should be observed together to understand the student's problem-solving as a whole process. But in the prototype, the two steps couldn't be presented at once.

4. There is no clear data value presented in the visualization. Users need to judge the value by their subjective perception.

Based on feedback from the pilot evaluation, the visualization prototype was revised. The final visualization is introduced in the next section.

3.6.2 Final Design

The final version of the visualization consists of four parts – students' performance overview, eye movement plot, mouse operation plot, and the metrics comparison chart (Figure 3.19). The top part is the overview of students' performance. Students' IDs and the tasks they completed are presented in this part. The bar length represents the time they spend on each task, and the color of bars present their correctness (green-correct, red-incorrect). The bars and students' IDs are clickable. Users can click on bars to see the corresponding task's visualization. For example, in Figure 3.19, student P5's task B3.5 is presented. The visualization of the student's eye movement, mouse operation, and metrics comparison radar chart are popped-out.

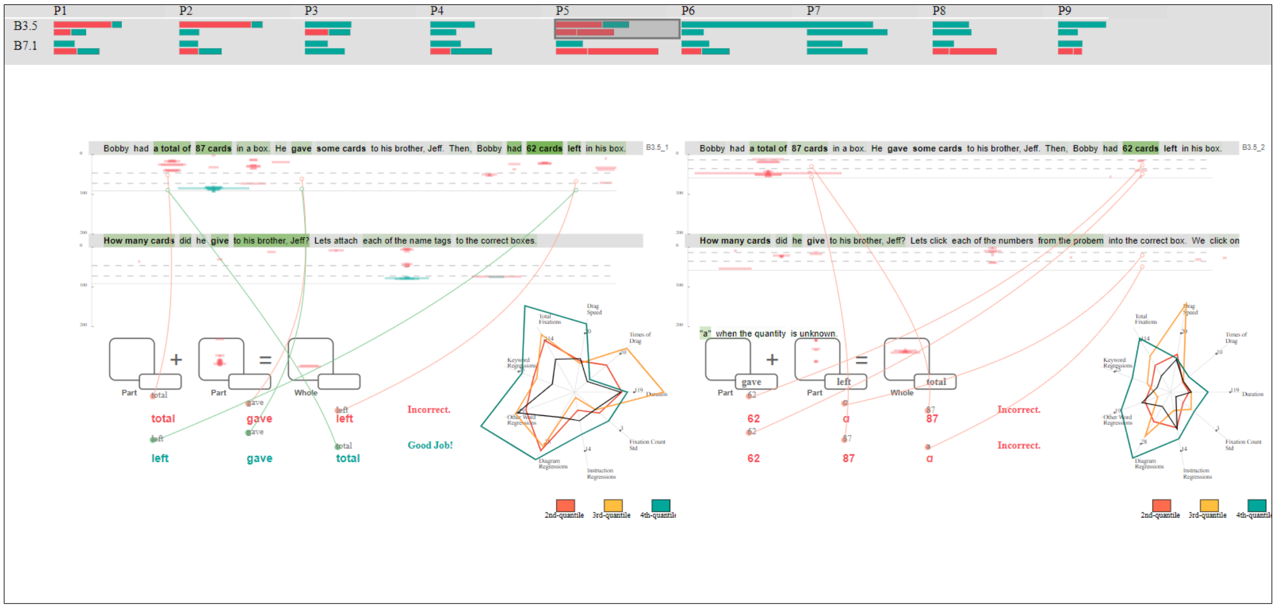

Figure 3.19 Final Version of Visualization

The problem context is presented in the visualization. The problem content is divided into "meaning units" and categorized into three groups (Keyword, Number, and Other). The text units that are recognized as Keywords and the Numbers are bolded. The lines below the text units are students' fixations. The length of a line represents the duration of fixations. The y-coordinates of lines illustrate the order of fixations. The order increases from top to bottom (Figure 3.20). Students' visual attention to different elements of a problem is presented through the ordered fixation lines. Different segments of problem-solving are marked using dashed lines. For example, when students solve the problem, there is a time before which students can only read the problem. After that point, students are allowed to operate their mouse. The operation start time is marked using a dashed line. When a student tried twice, there will be two dashed lines (Figure 3.20). The solid line indicates the end of the problem-solving.

Students were asked to drag name tags/number tags to build the diagram equation. The drag operations are depicted using curved lines. The start points of curved lines located at the position of tags. The y-coordinates of starting points are associated with fixation orders, which indicates that the drag operation and a fixation happened around the same time. For example, in Figure 3.20, on the left, a red-colored mouse dragline shows that, during the first time try, the student dragged the "total" tag immediately after the beginning of the operation. The endpoints of curved lines represent where the tags were placed in the diagram equation. The curvature of lines depicts the speed of drag operations. The more extreme the curvature, the slower the drag movement. Students' drag history is depicted by presenting the dragged tags from top to bottom (Figure 3.21). Besides that, students' correctness is presented using blue and red colors. The fixation lines, drag curves, and feedbacks are colored in correspondence with the student performance correctness.

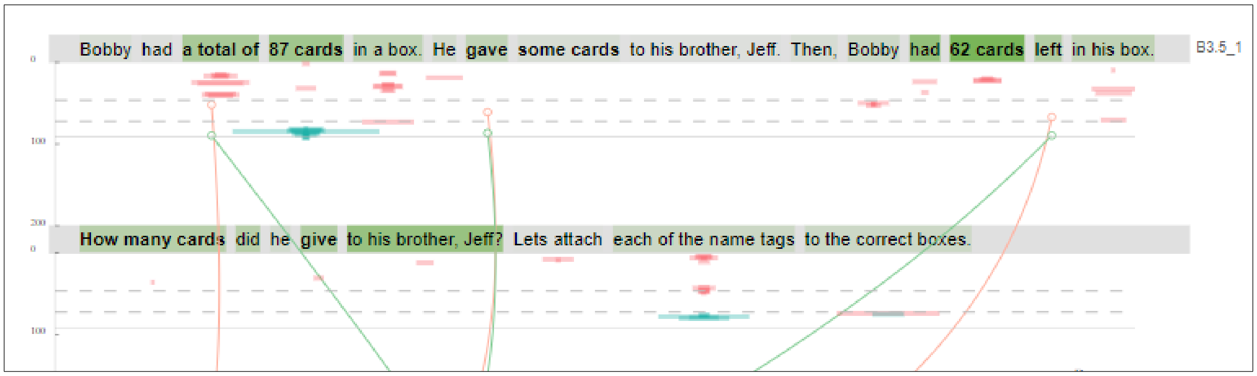

Figure 3.20 Students' fixations on problem content

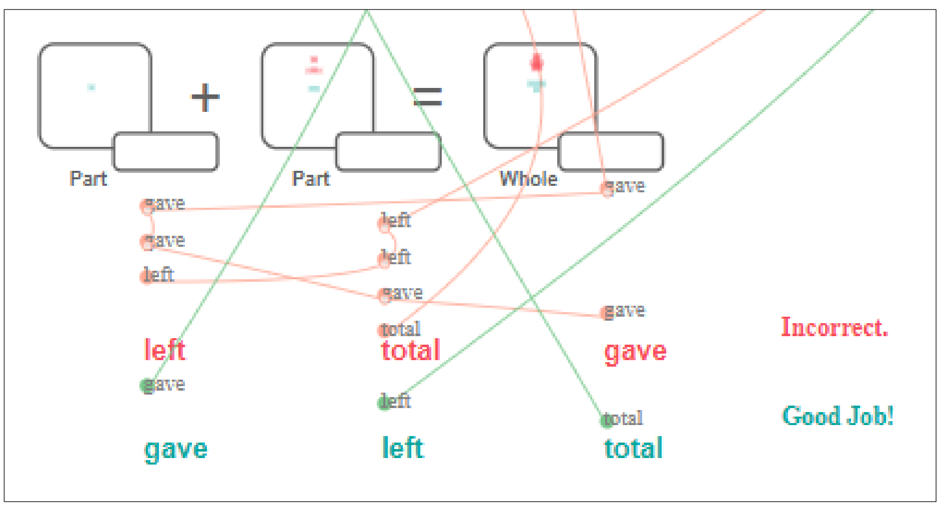

Figure 3.21 Student's drag history (from top to bottom)

Users can also compare individual student performance with the other students' performance, and the group averages by employing radar charts (Figure 3.22). The presented metrics include time spend on the task (duration), number of tries, number of drags, drag speed, total fixation count, keyword regressions, other word regressions, diagram regressions, and fixation distribution standard deviation. Fixation standard deviation is a metric developed to describe fixation distribution patterns in the problem content. Every meaning unit of the problem content is recognized as an AOI. The fixation numbers within each AOI were calculated. Then the standard deviation of fixation counts was calculated. If a student's fixations are aggregated in a few AOIs and the other AOI fixation counts are zero or small number. The differences among the fixation counts are big. Then the fixation standard deviation is big. Otherwise, if fixations are more uniformly distributed, the fixation standard deviation is small.

The target student metric points in the radar chart are connected by solid black lines, while other students' metrics are connected by colored lines. The color of the line is representing the group that the student belongs to. According to the students' correctness rates in six tasks, students were grouped into the low-performance group (red; 25% - 50%), median-performance group (yellow; 50%-75%), and high-performance group (green; 75%-100%). Each performance group's metric averages were marked with colored lines. Through the comparison radar chart, users can compare the target student's problem-solving metrics with the other students, and the average metrics of groups. For example, in Figure 3.22, we can see the target student's 'times of drag' metric is bigger than all three group averages indicate that the student dragged tags more than the average drag times of all three groups.

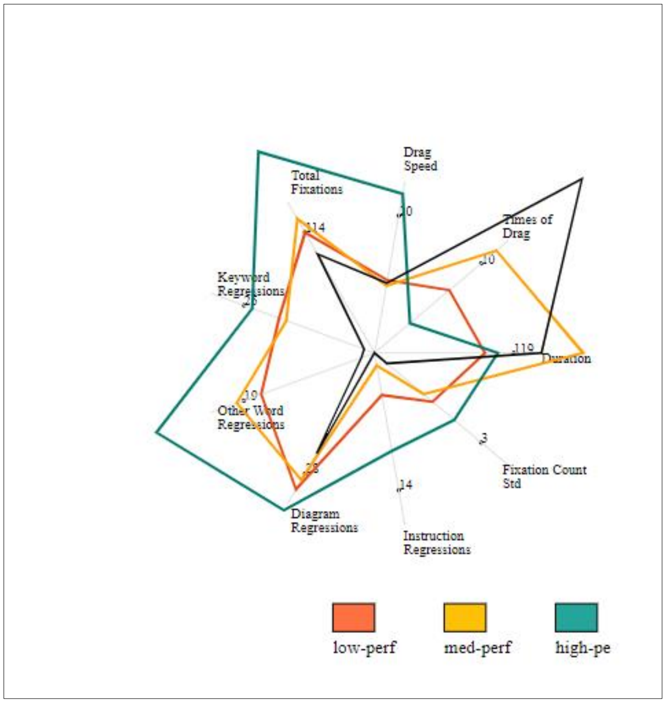

Figure 3.22 Metrics Comparison Chart

Additionally, the heatmaps of students' visual attention were also provided in the visualization system. Evaluators can use the heatmap visualization if needed.

3.7 Evaluation Experiment Design

The purpose of the evaluation is to examine whether the visualization system can help educational researchers comprehend students' problem-solving processes, including quickly and accurately identifying students' problem-solving strategies and difficulties. In the evaluation experiment, the visualization system was compared to the video analysis method, which is the traditional student problem-solving analysis method, to answer the following three evaluation questions:

1. **Compared to videos, can the visualization system make educational researchers spend less time identifying students' problem-solving strategies and their problem-solving difficulties?**

To answer this question, the time evaluators spent on each diagnosis task was recorded, calculated, and compared across visualization and videos. Evaluators might be familiar with the video analysis method and unfamiliar with the visualization system, which could lead to more time spent on the visualization system. In this case, if evaluators spent less time using the visualization, the efficiency of the visualization system could be proved to some extent. Besides that, evaluators' time spent on the tasks was analyzed to see if more evaluators took less time as they completed more tasks.

2. **Compared to videos, can the visualizations system help educational researchers get more reliable diagnoses on students' problem-solving strategies, difficulties, and performance levels?**

To answer this question, evaluators diagnosed students' problem-solving strategies, difficulties, and their performance levels (see problem-solving diagnosis tasks in the following section). If more than half of evaluators diagnosed a student as using a specific strategy or having a specific kind of difficulty, the diagnosis was considered 'reliable'. If no more than half of the evaluators agreed on the diagnosis, a discussion among evaluators was held to reach agreement on the problem-solving strategy or diagnosis. After getting reliable results on the problem-solving diagnosis tasks, evaluators' diagnosis results using different analysis methods (visualization/video) were analyzed to see whether the visualization system can help evaluators get more reliable diagnosis results.

3. **What aspects of the visualization system/video result in higher task load to educational researchers?**

To answer this question, students' task loads were measured using the NASA Task Load Index (NASA TLX), and a semi-structured interview was held.

3.7.1 Evaluator Recruitment

I recruited seven math educators as participants of the evaluation through sending emails to the graduate students in the special education (mathematics direction) and the students in the class of EDCI637 – Teaching Mathematics. These evaluators are not only math educators but also target users for the visualization system. The evaluator recruitment criteria were: 1) Having math teaching experience. 2) Being familiar with the computer-assisted math learning program. 3) Having an educational background (at least a master's or Ph.D. degree in Education).

3.8 Evaluation Experiment Conduction

As mentioned in section 3.5, 20 elementary school students were asked to take two demos, two practice tasks, and four test tasks, including two part-part-whole tasks and two comparison tasks. The elementary school students were divided into high performance group, medium performance group, and low performance group according to their performance in the test tasks (correctness percentage: 100% - 75%; 75% - 50%; 50% - 25%). The math educators were asked to analyze the student problem-solving process, which took about four to six minutes (according to my experience and the result of the prototype evaluation study). If each evaluator analyzes all four test tasks of 20 students, it will take six to eight hours in addition to introduction, interview, and survey time. Thus, due to the time constraints, I randomly selected one part-part-whole task and one comparison task from the four test tasks, and I randomly selected three students from each of the three performance groups. Each evaluator analyzed 18 tasks from the nine students. More specifically, the nine students' problem-solving processes on problem B3.5 and B7.1 were illustrated to evaluators using the visualization system or video.

Evaluators were asked to diagnose students' problem-solving strategies, and difficulties they had when solving problems. The evaluation experiment was conducted in the Purdue University Heavilon Hall computer lab. The experiment was done in two sessions over two days. The reason that the experiment was conducted over two days is that according toestimates, evaluators need about 108 minutes (2 tasks * 9 students * 6 minutes = 108 minutes) to complete diagnosis tasks. Accounting for evaluation introduction, survey, and interview time, the experiment time can reach as much as 180 minutes. Evaluators may become tired and frustrated completing the evaluation experiment all at once, which would seriously influence the evaluation results. Thus, the experiment was divided into two sessions over two days. Each session lasted about one and a half hours. The first session had three sub-sessions: Introduction and training, diagnosis tasks, and interview. In the second session, there was no introduction and training, students directly worked on diagnosis tasks, then completed the survey and interview.

3.8.1 Introduction and Visualization Training

Below are the introduction and training procedures in the first session:

1. At the beginning of the first session, the experiment was introduced to evaluators, and IRB consent form was given to the evaluators.

2. After evaluators signed the IRB consent form, their demographic information, including education background, experience in computer-assisted math learning, and years of working with elementary school students, were collected.

3. The theories of mathematic problem-solving strategies and difficulties that students may meet were introduced to the evaluators.

4. Then evaluators received visualization training. In the training session, the types of data presented by the visualization and the meaning of different visualization charts were introduced.

5. To help control for participants' visualization familiarity levels, two or more student problem-solving tasks were given as training tasks. One student problem-solving task was used as an example to introduce the visualization and present how to read the student's eye movement, drag operations, drag histories, and the radar chart. It was illustrated how to combine such student data together to understand the student problem-solving process. Then the evaluators were asked to analyze another student problem-solving task with the visualization to show they understand the visualization correctly and are able to use the visualization smoothly The participant had six minutes to read the visualization, in order to tell the student eye movement, drag operations, and problem-solving process. If he/she completed the task within six minutes, he/she could start the evaluation. If not, he/she was asked to perform more training tasks until he/she was familiar with the visualization enough to read it correctly within six minutes. The participant who can read the visualization correctly within six minutes is considered well-trained is based on two reasons. First, the video recording of one student problem-solving task is about three minutes. For participants who learn the visualization the first time, it is reasonable to give more than three minutes to read the visualization. Second, to make sure all participants are familiar with the visualization to some extent to complete the diagnosis tasks, there should be an upper-bound time limit. From the prototype evaluation, I found it took evaluators five to six minutes to go through the visualization without significant pause and frustration. Thereby, I set six minutes as the time criteria for qualified visualization training.

6. The video analysis method was introduced to the evaluators. They were informed that they could control the video to speed up, slow down, and playback as needed.

Demographic information collection and experiment introduction took about 20 minutes, while training evaluators to use the visualization took approximately 40 minutes.

3.8.2 Student Problem-Solving Diagnosis

Evaluators diagnosed 18 student problem-solving tasks in two days. In each day, evaluators were supposed to complete nine of the tasks, so the tasks were divided into two groups (Group 1 and Group 2). There were two types of problem-solving tasks – B3.5 (part-part-whole) and B7.1 (comparison), to ensure the balance of the task distribution, each group contained at least four tasks of each problem type. Other than the above criterion, the tasks were randomly divided into two groups.

Half of the evaluators were randomly selected to take Group 1 tasks on the first day and take Group 2 tasks on the second day. Another half of the participants took task groups in reverse order. Evaluators analyzed the problem-solving tasks using either visualization method or video method. For each problem-solving task, an evaluator only analyzed the task once using one of the two analysis methods. The analysis method assigning criteria are: 1. For each evaluator, he/she analyzes half of the problem-solving tasks using the video method and analyzes the remaining problem-solving tasks using the visualization method. 2. For each problem-solving task, it is analyzed by half of the evaluators using the video method and half of the evaluators using the visualization method. The task distribution criteria are also listed in Table 3.2.

Table 3.2 Task Distribution Criterions

Criterion	Group 1	Group 2
1	Nine tasks randomly ordered	Another nine tasks randomly ordered
2	Half of the participants took Group 1 first. Another half of the participants took Group 2 first.	
3	In each group, participants took half of the tasks using the visualization analysis method and took the other tasks using the video analysis method.	
4	For each task, it took by half of the participants using video and took by the other participants using the visualization analysis method.	

Table 3.3 is an example of task distribution. The diagnosis processes were videotaped. Time spent on each task and answers to each diagnosis question were recorded.

Table 3.3 Task Distribution Example

	Day1								
E1	S5-B3.5 Vis	S2-B3.5 Vid	S3-B3.5 Vis	S6-B7.1 Vid	S1-B7.1 Vid	S8-B3.5 Vis	S5-B7.1 Vis	S4-B3.5 Vid	S9-B7.1 Vis
	Day1								
E2	S4-B7.1 Vid	S8-B7.1 Vis	S3-B7.1 Vid	S2-B7.1 Vis	S7-B7.1 Vis	S7-B3.5 Vid	S1-B3.5 Vis	S9-B3.5 Vid	S6-B3.5 Vid
	Day2								
E1	S9-B3.5 Vid	S1-B3.5 Vis	S3-B7.1 Vid	S6-B3.5 Vid	S2-B7.1 Vis	S7-B7.1 Vis	S4-B7.1 Vid	S7-B3.5 Vis	S8-B7.1 Vid
	Day2								
E2	S1-B7.1 Vis	S5-B7.1 Vis	S9-B7.1 Vid	S5-B3.5 Vid	S3-B3.5 Vis	S4-B3.5 Vid	S6-B7.1 Vis	S2-B3.5 Vis	S8-B3.5 Vid

S#-B#.# is task index. For example, S5-B3.5 is Student5's task B3.5; E# represents the evaluator index. For example, E1 is Evaluator 1. Vid and Vis are diagnosis methods. Vid is Visualization, and Vis is Video. These indices describe a specific evaluator analyzing a specific task using a specific method. For example, Evaluator 1 analyzed S5-B3.5 on the first day using the visualization system.

They were asked to answer the below six diagnosis questions for each problem-solving task.

I. What was the pattern of the student's visual attention when solving this problem?

1) Viewing most part of the problem

2) Focusing only on keywords and/or numbers

3) Paying little attention to the problem

4) Other – Please describe

5) Not sure

II. What was the pattern of the student's drag-operation when solving the problem?

1) Drag only once for each tag

2) Drag multiple times for each tag

3) Miss a tag

III. How was the student's performance?

1) Correct on the 1st try

2) Incorrect on the 1st try; Correct on the 2nd try

3) Incorrect on the 1st try; Incorrect on the 2nd try

IV. What was the problem-solving strategy the student used when solving the problem?

1) Model-based problem-solving strategy

2) Direct translation strategy (keyword strategy)

3) Linear drag with little attention to the question

4) Guess and Check

5) Other - Please describe

6) Not sure

V. What were the struggles/difficulties that the student meet when solving the problem?

1) No struggles/difficulties

2) Difficulty in understanding the problem (literally)

3) Difficulty in understanding the diagram equation

4) Paid little attention to the problem

5) Careless drags/clicks

6) Other – Please describe

7) Not sure

VI. To which performance group will you classify the student?

1) Low performance – (25~50%)

2) Med performance group– (50~75%)

3) High-performance group – (75-100%)

After evaluators completed the diagnosis tasks, they went to the third sub-session. They were asked to take the NASA-TLX Survey (Figure 3.23) to measure their task load using different methods (visualization/video) to complete the diagnosis tasks. Based on evaluators' answers to NASA-TLX, a semi-structured interview was conducted to explore evaluators' diagnosis process and the reasoning behind their choices. The interview questions are listed in Table 3.4 and the interview protocol is in the Appendix A. The interview took about 20 minutes. The interview process was video recorded.

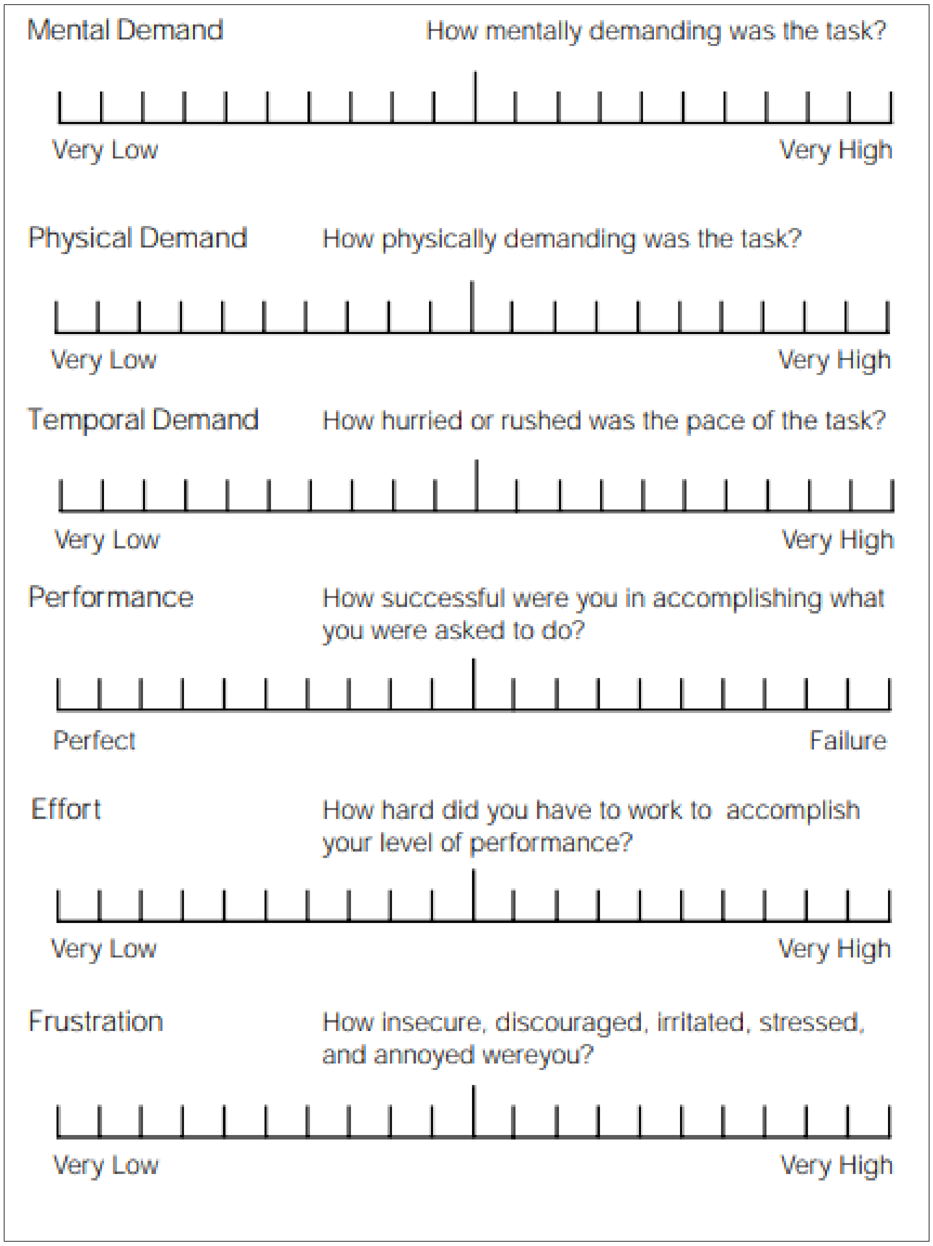

Figure 3.23 NASA-TLX (Visualization /Video)

Table 3.4 Semi-Structured Interview Questions

#	Semi-Structured Interview Questions
1	Based on your response to NASA-TLX, you have a higher/lower rating on visualization/video, why? Explain your rationale.
2	Could you describe your diagnose process? (I chose three tasks for the evaluator).

If the evaluators couldn't describe the reasoning of their NASA-TLX choices (Interview Question 1), they were asked to give an example from the tasks they did to make them feel more/less frustrated/hard/rushed. Then they went through their diagnosis process with the investigator to identify the possible elements that increased/decreased their task loads. In interview question 2, every evaluator was asked the same tasks' diagnosis tasks.

In the second session of the evaluation, evaluators were asked to do diagnosis tasks and the NASA TLX survey and interviews, which were same to the tasks, survey, and interview they completed in the first session. In the second session, they were asked one more interview question:

If the required training time of the visualization system is about 30 minutes, are you willing to use this system in your actual student problem-solving diagnosis for your current project/current class? If so, why it is worthwhile? If not, why not?

In the interview session, evaluators were also asked about their preference between the two methods.

3.8.3 Evaluation Data Collected

The data collected in the evaluation experiment, and the questions they supposed to answer are listed below (Table 3.5).

Table 3.5 Evaluation Research Questions & the Data Collected

Evaluation Research Questions	Data Collected and Analyzed
Q1. Compared to videos, can the visualization system make educational researchers spend less time identifying students' problem-solving strategies and their problem-solving difficulties?	Evaluators' time spent on each problem-solving task (quantitative data);
Q2: Compared to videos, can the visualizations system help educational researchers get more reliable diagnoses on students' problem-solving strategies, difficulties, and performance levels?	Evaluators' diagnose results on problem-solving diagnosis tasks (quantitative data);
Q3: What method the educational researchers prefer to use understanding students' problem-solving processes, and what aspects of the visualization system/video result in higher task load to educational researchers?	Evaluators' answers to: NASA-TLX (quantitative data); Semi-structured interview questions (qualitative data);

3.9 Summary

This chapter introduced the dissertation research approach, the theoretical background of problem-solving strategies, the visualization development process, and the design of the evaluation experiment. It also described in detail the data collected in the evaluation experiment and the research questions for the study. In the next chapter, the evaluation data are analyzed, and the analysis results are introduced.

CHAPTER 4. DATA ANALYSIS AND RESULTS

This chapter introduces the data collected from the evaluation experiment and results from the data analysis. The data collected were evaluators' demographic information, diagnosis results, the time spent on diagnosis tasks, and interview data. In this chapter, the data and their analysis results will be illustrated using charts, tables, and statistical models.

4.1 Evaluators and Their Demographic Information

Seven evaluators were recruited to complete the diagnosis tasks using either the developed visualization system or using the video method. The diagnosis tasks included identifying students' visual attention, drag patterns, performance, problem-solving strategies, difficulties, and their performance levels. Evaluators' performance in the diagnosis tasks were recorded and compared to conclude which analysis method (visualization/video) can more quickly and effectively help evaluators complete the diagnosis tasks.

Evaluators' demographic information, including current education, experience in computer-assisted learning, years of working with students, and years of working with elementary school students, are presented in table 4.1. All seven evaluators are Ph.D. students from Math education or Special education (the direction of mathematics problem solving involving students with math learning disabilities/difficulties). Six are senior Ph.D. students. Another is a first-year Ph.D. student. All have experience in computer assisted learning and have worked with students teaching math (mean = 5.5 years). Most evaluators have worked with elementary school students (mean = two years). Some of them have experience in video analysis. Among these evaluators, the first-year Ph.D. student (E3) had less experience in math education, and later, after analyzing the evaluator's diagnosis performance and interview data, I found the evaluator had less experience in math education, couldn't apply her educational knowledge to the diagnosis tasks or give convincing rationales for her diagnosis. The word such as "I feel," "impression," and "think" constantly appear in her diagnosis statements without data support. So, the evaluator was excluded. After excluding the evaluator, all six remaining evaluators are senior Ph.D. students with more than four years' experience in teaching student math. The detail information is illustrated in the table below.

Table 4.1 Evaluators' Demographic Information

Evaluator	Education	Experience in CAI	Years of working with elementary school students	Years of working with students	Hours of video analysis
E1	Fourth Year Ph.D. Student	Yes	2	7	>100 hrs.
E2	Fourth Year Ph.D. Student	Yes	4	6	>100 hrs.
E4	Third Year Ph.D. Student	Yes	0	5	0
E5	Ph.D. Candidate	Yes	5	8	0
E6	Ph.D. Candidate	Yes	2	4	>10 hrs.
E7	Ph.D. Candidate	Yes	0	8	0

4.2 Diagnosis Task Data Analysis and Results

Evaluators were asked to analyze student problem-solving tasks and answer diagnosis questions. For each student problem-solving task, an evaluator was required to answer six diagnosis questions, including the student's visual attention pattern, mouse drag pattern, performance (correctness), problem-solving strategies, difficulties, and student's performance group. In total, each evaluator took 18 student problem-solving tasks from nine students using either visualization or video analysis method. Evaluators' diagnosis time was also recorded. Evaluators took tasks within two days in random order.

For the diagnosis tasks, the independent variables were methods (visualization/video), student, task, evaluator, order (random), and round (day1/day2). The dependent variables were diagnosis time, correctness rate, or consensus rate.

Here the correctness rate and consensus rate are discussed in detail. For diagnosis questions such as drag patterns (how many times the student dragged the tags in the task), performance (whether the student completed the problem correctly), and performance group (which performance group the student belongs to), there are definite answers. Thus an evaluator's diagnosis can objectively be correct or wrong, and the evaluator's correctness rate is 100% or 0%. However, for diagnosis questions such as visual attention, problem-solving strategies, and difficulties, there is no objectively correct answer.

The students who took the problem-solving tasks are 2nd /3rd -grade elementary school students. They are too small to tell their strategies and difficulties clearly. In the experiment, if more than half of evaluators chose an option in their diagnosis, the option was labeled as the consensual answer. As evaluators were able to choose more than one problem-solving strategies or difficulties that student used/faced, there were some cases for which more than one consensual answer existed. Also, in some cases, no consensual answer has been achieved. In the latter case, the consensual answer was determined at a meeting between the investigator and the evaluators. After consensual answers to each question were determined, the consensus rate for each evaluator was calculated. The consensus rate was calculated by dividing the number of consensual answers chosen by the evaluator to the total number of consensual answers. For example, if more than half of evaluators diagnosed a student as using both a model-based problem-solving strategy and a keyword strategy, while an individual evaluator diagnosed the student only used model-based problem-solving strategy, that evaluator's diagnose consensus rate is 50%. There is another case where evaluators choose extra options. But there is no reason to determine extra options are wrong as problem-solving strategy and difficulty diagnoses are, to some extent, subjective. It is not reasonable to determine an evaluator's diagnoses as wrong. So, the choice of extra options will not influence evaluators' diagnosis consensus rate. For example, if there are two consensual options, an evaluator chooses one more option besides the two consensual options thenhe evaluator's consensus rate is still 100%. Evaluators' diagnosis consensus rate will only be influenced by missing consensual options, not by choosing extra options.

Figure 4.1 below illustrates the evaluators' diagnosis results of problem-solving strategies. The x-axis is students, and the y-axis is diagnosis options (1- Model-based problem-solving strategy, 2- Direct translation strategy/keyword strategy, 3- Linear drag with little attention, 4- Guess & Check, 5- Other strategies, 6-Not sure). Each sector represents one evaluator, and the colored sector edge represents the method the evaluator used to diagnose. Consensual options are marked by the green dots in the center of the sectors.

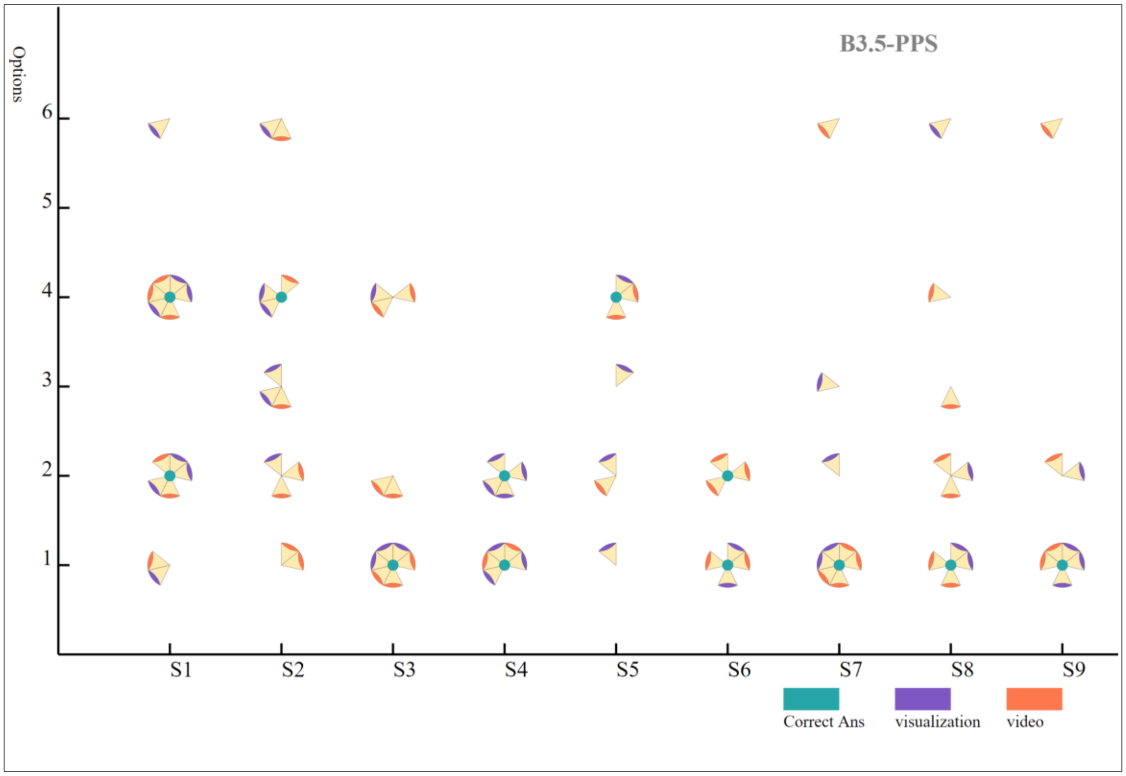

Figure 4.1 Problem-Solving Strategy Diagnosis Results

I want to investigate whether there are significant differences between visualization and video methods in diagnosis time and diagnosis correctness rate/consensus rate. So, the treatment of interest is the analysis method (visualization/video), the other independent variables, including student, task, evaluator, and round, are considered as blocks. Mixed general linear regression was chosen to analyze the diagnosis data. Student, evaluator, and order are random factors. The interactions among independent variables are also considered in the analysis process.

4.2.1 Visual Attention Diagnosis

Figure 4.2 presents a summary of the model. It shows that the analysis method is a significant indicator of visual attention diagnosis consensus rate (p-value=0.026). The mean visual attention diagnosis consensus rate of using the visualization analysis method is 0.73. The mean visual attention diagnosis consensus rate of using the video analysis method is 0.58. Visual attention diagnosis task completed using the visualization method had a significantly higher consensus rate than video, which indicates visualization is significantly better than video in helping evaluators get consistent diagnoses on students' visual attention patterns. The other factors, including task taking sequence and round, had no significant effect on consensus rate, and there is no significant difference between the first round consensus rate and the second round consensus rate.

The GLM Procedure

Dependent Variable: Rate

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	19	3.90583160	0.20557008	0.90	0.5837
Error	190	43.40845411	0.22846555		
Corrected Total	209	47.31428571			

R-Square	Coeff Var	Root MSE	Rate Mean
0.082551	72.73619	0.477981	0.657143

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Evaluator	5	1.26608310	0.25321662	1.11	0.3574
Student	8	1.02151673	0.12768959	0.56	0.8106
Task	3	0.35924691	0.11974897	0.52	0.6662
Method	1	1.15731985	1.15731985	5.07	0.0256
Sequence	1	0.06710055	0.06710055	0.29	0.5885
Round	1	0.03456447	0.03456447	0.15	0.6977

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Evaluator	5	1.22985407	0.24597081	1.08	0.3747
Student	8	1.07380545	0.13422568	0.59	0.7875
Task	3	0.38256641	0.12752214	0.56	0.6433
Method	1	1.20215269	1.20215269	5.26	0.0229
Sequence	1	0.06951428	0.06951428	0.30	0.5819
Round	1	0.03456447	0.03456447	0.15	0.6977

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

Method	Rate LSMEAN	H0:LSMean1=LSMean2 Pr > |t|
vid	0.58092377	0.0229
vis	0.73461978	

Figure 4.2 Visual Attention Diagnosis Data Analysis Results

4.2.2 Drag Pattern Diagnosis

Figure 4.3 shows the mixed general linear regression model output of drag pattern diagnosis data. It shows that analysis method is not a significant predictor of correctness rate (P-value = 0.336). Instead, task and the interaction between task and student (student*task) are significant predictors of drag pattern diagnosis correctness rate. Further, there is no significant difference in correctness between using visualization and using video in the drag pattern diagnosis. The correctness rate mean of using visualization is 0.85, while the correctness rate mean of using video is 0.89.

The GLM Procedure

Dependent Variable: Rate

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	43	7.85745299	0.18273146	1.94	0.0017
Error	166	15.67111844	0.09440433		
Corrected Total	209	23.52857143			

R-Square	Coeff Var	Root MSE	Rate Mean
0.333954	35.25853	0.307253	0.871429

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Evaluator	5	0.63804855	0.12760971	1.35	0.2452
Student	8	1.28849108	0.16106139	1.71	0.1003
Task	3	1.89636405	0.63212135	6.70	0.0003
Method	1	0.08598358	0.08598358	0.91	0.3413
Sequence	1	0.00610809	0.00610809	0.06	0.7995
Round	1	0.24611285	0.24611285	2.61	0.1083
Task*Student	24	3.69634480	0.15401437	1.63	0.0398

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Evaluator	5	0.65213110	0.13042622	1.38	0.2337
Student	8	1.39585744	0.17448218	1.85	0.0715
Task	3	1.84788383	0.61596128	6.52	$\boxed{0.0003}$
Method	1	0.08804390	0.08804390	0.93	0.3356
Sequence	1	0.01501250	0.01501250	0.16	0.6906
Round	1	0.01769834	0.01769834	0.19	0.6656
Task*Student	24	3.69634480	0.15401437	1.63	$\boxed{0.0398}$

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

Method	Rate LSMEAN	H0:LSMean1=LSMean2 Pr > |t|
vid	0.84627698	0.3356
vis	0.88932128	

Figure 4.3 Drag Pattern Diagnosis Data Analysis Results

4.2.3 Performance Diagnosis

Figure 4.4 shows the mixed general linear regression model output of the performance diagnosis data. In the mixed general linear model, student, round, the interaction between round and method (method*round), and the interaction between student and task (student*task) are significant predictors of performance diagnosis correctness rate. While method is not a significant predictor. The performance diagnosis correctness rates between using visualization method and using video methods are not significantly different (p-value = 0.16). For the video method, the mean of performance diagnosis correctness rate is 0.93, while using visualization, the correctness rate mean is 0.97. When considering the diagnosis for video in the first round, the mean of performance correctness rate is 0.86, and for the second round the mean increased to 0.997. This increase is significant (P-value=0.01). For visualization in the first round, the mean of performance correctness rate is 0.968, while in the second round, the mean is 0.975. The increase is not significant.

The GLM Procedure

Dependent Variable: Rate

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	44	4.02610159	0.09150231	2.36	<.0001
Error	165	6.39770793	0.03877399		
Corrected Total	209	10.42380952			

R-Square	Coeff Var	Root MSE	Rate Mean
0.386241	20.77956	0.196911	0.947619

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Evaluator	5	0.25898109	0.05179622	1.34	0.2516
Student	8	0.96212529	0.12026566	3.10	0.0027
Task	3	0.27772266	0.09257422	2.39	0.0709
Method	1	0.09975182	0.09975182	2.57	0.1106
Sequence	1	0.00082223	0.00082223	0.02	0.8844
Round	1	0.27786798	0.27786798	7.17	0.0082
Task*Student	24	1.98218107	0.08259088	2.13	0.0030
Method*Round	1	0.16664945	0.16664945	4.30	0.0397

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Evaluator	5	0.31897805	0.06379561	1.65	0.1508
Student	8	0.89131610	0.11141451	2.87	0.0051
Task	3	0.28995254	0.09665085	2.49	0.0619
Method	1	0.07916172	0.07916172	2.04	0.1549
Sequence	1	0.00861592	0.00861592	0.22	0.6380
Round	1	0.21495904	0.21495904	5.54	0.0197
Task*Student	24	2.07677596	0.08653233	2.23	0.0017
Method*Round	1	0.16664945	0.16664945	4.30	0.0397

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

Method	Rate LSMEAN	H0:LSMean1=LSMean2 Pr > |t|
vid	0.93082538	0.1549
vis	0.97165648	

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

Method	Round	Rate LSMEAN	LSMEAN Number
vid	1	0.86497752	1
vid	2	0.99667324	2
vis	1	0.96849287	3
vis	2	0.97482008	4

Least Squares Means for effect Method*Round
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: Rate

i/j	1	2	3	4
1		0.0098	0.0600	0.0278
2	0.0098		0.9123	0.9546
3	0.0600	0.9123		0.9988
4	0.0278	0.9546	0.9988	

Figure 4.4 Performance Diagnosis Data Analysis Results

4.2.4 Problem-Solving Strategy Diagnosis

In Figure 4.5, the mixed general linear regression analysis results show that there are no significant differences between video and visualization methods in diagnosing students' problem-solving strategies (p-value from a two-tailed t-test = 0.0816). Using the visualization method, the problem-solving strategy diagnosis consensus rate mean was 0.81, while using the video method, the problem-solving strategy diagnosis consensus rate mean was 0.67.The visualization method seemed to help evaluators get a higher consensus rate than the video method in diagnosing students' problem-solving strategies (p-value =0.0408).

The GLM Procedure

Dependent Variable: Rate

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	17	3.02341862	0.17784815	1.19	0.2913
Error	87	13.02420042	0.14970345		
Corrected Total	104	16.04761905			

R-Square	Coeff Var	Root MSE	Rate Mean
0.188403	52.42078	0.386915	0.738095

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Evaluator	5	1.39320728	0.27864146	1.86	0.1094
Student	8	0.77315026	0.09664378	0.65	0.7373
Task	1	0.08358183	0.08358183	0.56	0.4570
Method	1	0.44673063	0.44673063	2.98	0.0876
Sequence	1	0.30234067	0.30234067	2.02	0.1589
Round	1	0.02440795	0.02440795	0.16	0.6874

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Evaluator	5	1.47243915	0.29448783	1.97	0.0915
Student	8	0.68458690	0.08557336	0.57	0.7985
Task	1	0.11143329	0.11143329	0.74	0.3906
Method	1	0.46475996	0.46475996	3.10	$\boxed{0.0816}$
Sequence	1	0.29782037	0.29782037	1.99	0.1620
Round	1	0.02440795	0.02440795	0.16	0.6874

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

Method	Rate LSMEAN	H0:LSMean1= LSMean2 Pr > |t|
vid	0.67171963	0.0816
vis	0.80686841	

Figure 4.5 Problem-Solving Strategy Diagnosis Data Analysis

4.2.5 Difficulty Diagnosis

Figure 4.6 presents the mixed general linear regression model output of difficulty diagnosis data. It shows that the mean of difficulty diagnoses consensus rate using video was 0.62, while the mean of difficulty diagnoses consensus rate using visualization was 0.76. Difficulty diagnosis tasks using the visual analysis method have significant higher consensus rates compared to the video analysis method (p-value from a one-tailed t-test = 0.037).

The GLM Procedure

Dependent Variable: Rate

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	17	2.64809383	0.15577023	0.93	0.5402
Error	87	14.54238236	0.16715382		
Corrected Total	104	17.19047619			

R-Square	Coeff Var	Root MSE	Rate Mean
0.154044	59.21196	0.408844	0.690476

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Evaluator	5	0.83304155	0.16660831	1.00	0.4246
Student	8	0.69759920	0.08719990	0.52	0.8372
Task	1	0.22554955	0.22554955	1.35	0.2486
Method	1	0.48056524	0.48056524	2.87	0.0935
Sequence	1	0.06052857	0.06052857	0.36	0.5489
Round	1	0.35080973	0.35080973	2.10	0.1510

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Evaluator	5	0.84225234	0.16845047	1.01	0.4181
Student	8	0.77633837	0.09704230	0.58	0.7913
Task	1	0.20129211	0.20129211	1.20	0.2755
Method	1	0.54866990	0.54866990	3.28	0.0735
Sequence	1	0.06808051	0.06808051	0.41	0.5250
Round	1	0.35080973	0.35080973	2.10	0.1510

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

Method	Rate LSMEAN	H0:LSMean1=LSMean2 Pr > |t|
vid	0.61756735	0.0735
vis	0.76441038	

Figure 4.6 Difficulties Diagnosis Data Analysis Results

4.2.6 Performance Group Diagnosis

Figure 4.7 presents the mixed general linear regression model output of students' performance group diagnosis data. It shows student and the interaction between student and task (student*task) as being significant predictors of performance diagnose consensus rate. Although method was did not achieve significance as a predictor, the visualization method was trending higher, with a higher consensus rate mean (mean=0.60) than the video method (mean = 0.54).

The GLM Procedure

Dependent Variable: Rate

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	25	12.52848400	0.50113936	2.97	0.0001
Error	79	13.31913505	0.16859665		
Corrected Total	104	25.84761905			

R-Square	Coeff Var	Root MSE	Rate Mean
0.484706	73.07381	0.410605	0.561905

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Evaluator	5	0.94565826	0.18913165	1.12	0.3557
Student	8	6.29522267	0.78690283	4.67	0.0001
Task	1	0.01951079	0.01951079	0.12	0.7346
Method	1	0.44273750	0.44273750	2.63	0.1091
Sequence	1	0.00078276	0.00078276	0.00	0.9458
Round	1	0.15903365	0.15903365	0.94	0.3344
Task*Student	8	4.66553837	0.58319230	3.46	0.0018

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Evaluator	5	0.89281320	0.17856264	1.06	0.3894
Student	8	6.18800964	0.77350121	4.59	0.0001
Task	1	0.00576456	0.00576456	0.03	0.8538
Method	1	0.08638773	0.08638773	0.51	0.4762
Sequence	1	0.08288391	0.08288391	0.49	0.4853
Round	1	0.09151834	0.09151834	0.54	0.4634
Task*Student	8	4.66553837	0.58319230	3.46	0.0018

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

Method	Rate LSMEAN	H0:LSMean1=LSMean2 Pr > |t|
vid	0.53522408	0.4762
vis	0.59552265	

Figure 4.7 Performance Group Diagnosis Data Analysis Resultss

4.2.7 Diagnosis Task Data Analysis Summary

Table 4.2 summarizes the mean of diagnosis correctness rate or consensus rate in different tasks using either visualization analysis method or video analysis method. It also shows that there are significant differences between the two analysis methods. The data analysis results show that visualization method can significantly help researchers to get more consensual diagnoses of the visual attention patterns. For problem-solving strategy and difficulty diagnosis tasks, analyzing using the visualization method can achieve higher consensus rates than using the video method.

Additionally, it can be observed that the diagnosis correctness rates of drag patterns and performance patterns are high (>85%). While for the other diagnosis tasks, the consensus rates are low (<81%). That may be because drag patterns and performance patterns are clear and objective, while visual attention, problem-solving strategies, difficulties, and performance group diagnosis are evaluators' subjective judgments based on students' learning behavior and evaluators' professional knowledge. In the experiment, evaluators only analyzed one or two tasks by each student to diagnose the student's performance group. The limited task may lead to bias. Evaluators might need more tasks from the same student to make reliable diagnoses about the student's performance group. That could be the reason why students' performance group diagnosis correctness was so low (smaller than 60%).

Table 4.2 The Means of Diagnosis Correctness Percentage

	Mean of Consensus Rate (visualization)	Mean of Consensus Rate (video)	P-Value	Significance
Visual Attention Pattern	0.73	0.58	0.023	Yes
Drag Pattern	0.89	0.85	0.336	No
Performance Pattern	0.97	0.93	0.155	No
Problem-Solving Strategies	0.81	0.67	0.082	No
Difficulties	0.76	0.62	0.0735	No
Performance Group	0.60	0.54	0.476	No

4.3 Diagnosis Time Analysis and Results

The time evaluators spent to complete the diagnosis tasks was recorded and analyzed. The figure below shows individual evaluators' diagnosis time distribution. The blue line represents diagnosis time using the visualization method, and the orange line represents diagnosis time using the video method. It can be observed that the diagnosis time was decreased as more tasks were taken for both visualization and video methods. Also, diagnosis using the video method tend to spend more time than using the visualization method.

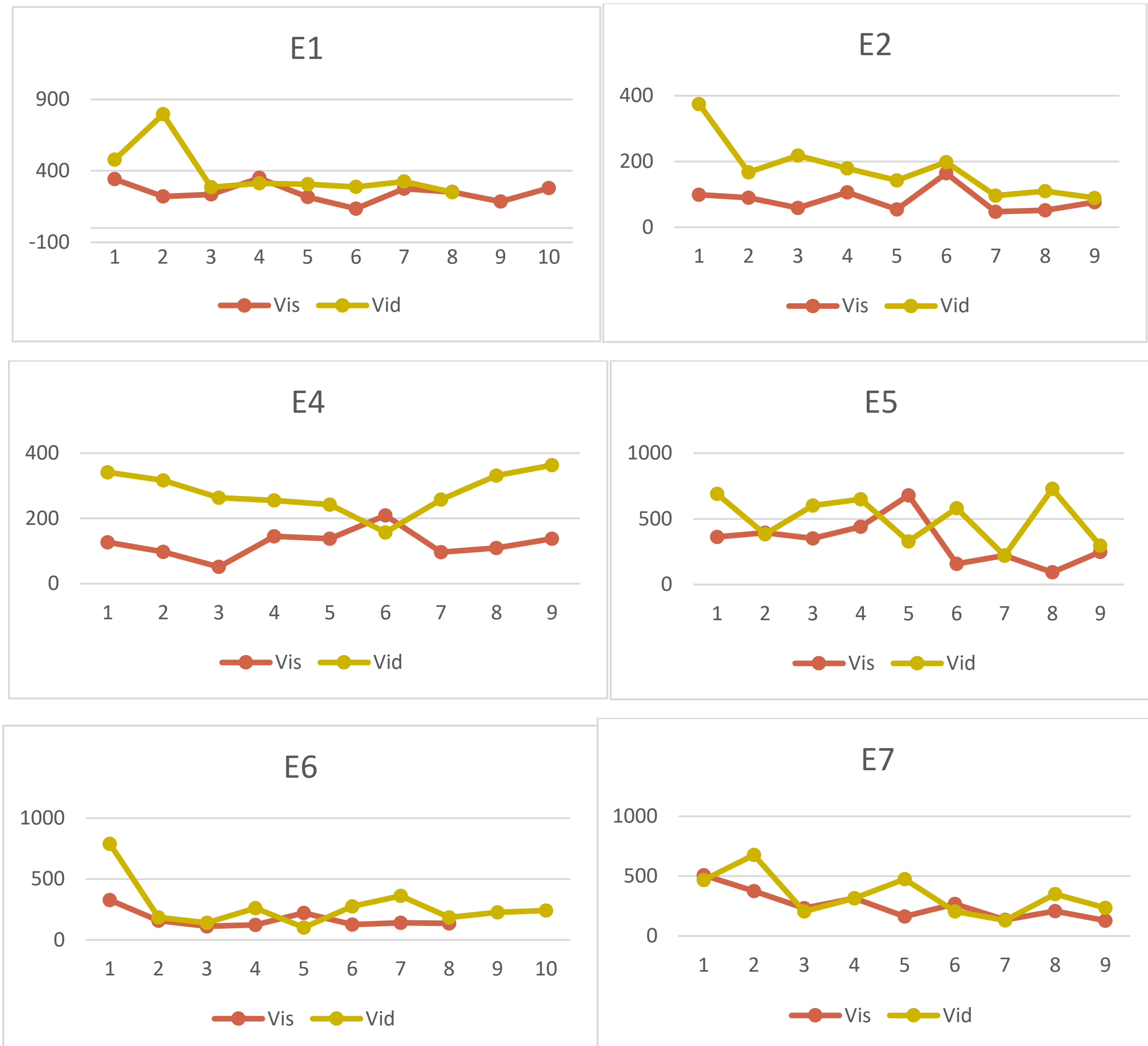

Figure 4.8 Individual Evaluators' Diagnosis Time Distribution

The general mixed effects linear regression analysis results (Figure 4.9) show that independent variables: evaluator, student, method, order, and round are all significant predictors of diagnosis time. There is significant difference (P-value < 0.0001) between diagnosis time using visualization method (mean=198.54 seconds) and using video method (mean = 308.64 seconds). Using the visualization method to complete the diagnosis tasks can significantly decrease the diagnosis time comparing with using the video method.

After doing more tasks, evaluators' diagnosis time decreased greatly for both visualization analysis method (Round1=227.09 seconds, Round2=169.99 seconds, p-value=0.1578) and video

(Round1=339.29, Round2=277.98, p-value=0.11) analysis method. In both rounds, the diagnosis time of using visualization method was significantly lower than using video method (Round1 difference = 112.19 seconds, P-value<0.001; Round2 difference =107.99 seconds, p-value<0.0001).

The GLM Procedure

Dependent Variable: Duration

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	26	2137470.057	82210.387	8.36	<.0001
Error	98	963258.640	9829.170		
Corrected Total	124	3100728.697			

R-Square	Coeff Var	Root MSE	Duration Mean
0.689344	38.92179	99.14217	254.7215

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Evaluator	6	937506.3387	156251.0564	15.90	<.0001
Student	8	387537.4138	48442.1767	4.93	<.0001
Task	1	497.8089	497.8089	0.05	0.8224
Method	1	351915.5200	351915.5200	35.80	<.0001
Order	8	351939.3483	43992.4185	4.48	0.0001
Round	1	107969.1667	107969.1667	10.98	0.0013
Method*Round	1	104.4608	104.4608	0.01	0.9181

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Evaluator	6	907533.3043	151255.5507	15.39	<.0001
Student	8	287897.0031	35987.1254	3.66	0.0009
Task	1	2846.7325	2846.7325	0.29	0.5917
Method	1	340293.5029	340293.5029	34.62	<.0001
Order	8	348701.4026	43587.6753	4.43	0.0001
Round	1	107972.7502	107972.7502	10.98	0.0013
Method*Round	1	104.4608	104.4608	0.01	0.9181

Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

Method	Duration LSMEAN	H0:LSMean1=LSMean2 Pr > |t|
0	308.635911	<.0001
1	198.542437	

0-Video, 1-Visualization

Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

Round	Duration LSMEAN	H0:LSMean1=LSMean2 Pr > |t|
1	283.190534	0.0013
2	223.987814	

1-Round1, 2-Round2

Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

Method	Round	Duration LSMEAN	LSMEAN Number
0	1	339.287598	1
0	2	277.984223	2
1	1	227.093470	3
1	2	169.991405	4

Least Squares Means for effect Method*Round
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: Duration

i/j	1	2	3	4
1		0.1143	0.0008	<.0001
2	0.1143		0.2209	0.0007
3	0.0008	0.2209		0.1578
4	<.0001	0.0007	0.1578	

Figure 4.9 Diagnosis Time Analysis Results

4.4 Task Load Report Data Analysis and Results

The NASA-TLX survey was conducted to measure evaluators' task load when completing the diagnosis tasks. Evaluators' answers were recorded, and the rationales were obtained via interviews. The NASA-TLX ratings were analyzed using paired t-test.

Overall, there was a significant difference between visualization and video methods (Figure 4.10). Diagnosis tasks using visual analysis method have a significant lower task load than using the video analysis method (Visualization Mean=8.12, Video Mean = 11.68, p-value <0.0001).

The TTEST Procedure

Difference: vis - vid

N	Mean	Std Dev	Std Err	Minimum	Maximum
72	-3.6111	5.8876	0.6939	-16.0000	10.0000

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
-3.6111	-4.9946	-2.2276	5.8876	5.0583	7.0446

DF	t Value	Pr > |t|
71	-5.20	<.0001

Figure 4.10 Overall Task Load Rating Analysis Results

Table 4.3 lists the evaluators' self-reported task load rating analysis results. For every task load measurement, the video method has a higher rating than the visualization method. Particularly, the video analysis method cased significantly higher temporal demand (p-value = 0.01) and physical demand (P-value<0.01) compared to the visualization method. In the interviews, evaluators stated that although they can play the video back, forth, speed up, and speed down, they still needed to follow the video to get information, which made them feel rushed. Additionally, the activities of controlling videos, tracking student eye and mouse movement in the videos also caused high physical demand (mean =9.25). There was little physical demand caused by visualization (mean = 4.08). If there was any, it may have come from evaluators' physically presenting the evaluation experiment. And the temporal demand for using visualization method (mean = 4) came from internal pressure to complete the tasks quickly. This kind of internal pressure also existed in video analysis. It caused even more temporal demand (mean = 10.08) in the video as evaluators couldn't take full control of the video.

The mental demands were very high for both visualization (mean = 13.17) and video (mean = 14) as the mental demands may have come from the complexity of data, the difficulties of interpreting different data sources together, and the challenges of making correct diagnoses. At the same time, there were some other opinions on the two methods. Some evaluators pointed out that combining and interpreting data while watching the video caused even more mental demand. Some evaluators thought the video could naturally combine data and present what's going on. For the visualization analysis method, evaluators agreed that it was challenging to combine all the information, although they had all data present in the visualization, and there was no time limit. For diagnosis performance, evaluators felt they didn't make very successful diagnoses in either visualization (mean = 8.42), or video (mean = 10.92) analysis tasks as evaluators had concerns, whether they interpreted information correctly, and whether they made correct diagnoses. Additionally, when using the video analysis method, evaluators needed to worry about whether they tracked and remembered students' eye movement and mouse movement correctly. This unconfident feeling may be the main reason for low self-reported performance successful rating.

Table 4.3 Task Load Measurement Rating Analysis Results

Task Load (1-Very Low,21-very High)	**Visualization Rating Mean**	**Video Rating Mean**	**P-Value**	**Significance**
Mental	13.17	14	0.52	No
Physical	4.08	9.25	0.009	Yes
Temporal	4	10.08	0.01	Yes
Performance	8.42	10.92	0.06	No
Effort	10.67	13.42	0.14	No
Frustration	8.17	12.5	0.053	No

Due to the difficulty of the diagnosis tasks, evaluators needed to spend a lot of effort completing the tasks in both visualization (effort rating mean = 10.67) and video (effort rating mean = 13.42) tasks. For frustration, the video analysis method (mean =12.5) brought more frustration than the visualization analysis method (mean = 8.17). In the interviews, many evaluators expressed their insecure feeling about data capturing, data interpreting, and insight identification. Some evaluators stated they felt pressure to make the correct diagnoses. For the visualization analysis method, one advantage is that it provided multiple data sources that can validate each other. The evaluator stated that "If I got similar diagnoses from performance, fixations, and other data, they kind of encourage me." For the video analysis method, one big advantage is that evaluators can get more information, especially the real mouse drag trajectory, which may sometimes reflect students' hesitation. Additionally, video naturally presents different sources of data at the same time. While using visualization, evaluators needed to combine different data sources by themselves. At last, all evaluators agreed that their task load decreased greatly after taking more tasks, especially for the visualization method.

Table 4.4 shows the mental demand, effort, and frustration decreased in the second round. The performance rate increased. As evaluators became more familiar with visualization in the second round, the mental demand rating of visualization decreased significantly. As the mental demand of visualization decreased significantly, the effort of using visualization to complete the diagnosis tasks also decreased greatly. Using the video analysis method, frustration decreased significantly. Some evaluators stated that they became more familiar with diagnosis tasks in the second round and they knew more about what information to focus on to answer later diagnosis questions.

In other measurements, including physical demand and temporal demand, ratings increased. That may be because evaluators became familiar with the diagnosis tasks and wanted to complete the tasks more quickly, which in turn increased the demands of physical and temporal. The task load measurements and reasons are summarized in Table 4.5.

Table 4.4 Task Load Change in Two Rounds

	Visualization-Round1	Visualization-Round2	P-Value	Video – Round1	Video – Round2	P-Value
Task Load	8.72	7.44	0.09	12	11.39	0.44
Mental (1-Very Low, 21-very High)	15.5	10.83	0.05	14.67	13.33	0.55
Physical (1-Very Low, 21-very High)	3.5	4.67	0.63	8.33	10.17	0.58
Temporal (1-Very Low, 21-very High)	3.67	4.33	0.72	9.17	11	0.39
Performance (1-Perfect, 21-Failure)	8.5	8.33	0.82	11.5	10.33	0.40
Effort (1-Very Low, 21-very High)	12	9.33	0.06	14.33	12.5	0.19
Frustration (1-Very Low, 21-very High)	9.17	7.17	0.40	14	11	0.03

Table 4.5 Summarization of Evaluators' Task Load

Mental Demand (Visualization Mean =12.21)	**Mental Demand (Video Mean = 13.64)**
Remember diagnosis instruction, diagnosis tasks, and options	
Interpret graphics	Follow the video
Interpret all the data in the visualization	Remember the student's Problem-solving process, including eye movement, mouse operation, and performance
Combine all data to get insights	Combine all information presented to make judgments in a short period
Physical Demand (Visualization Mean = 4.43)	**Physical Demand (Video Mean = 9.57)**
Sit there to complete diagnosis tasks	Track students' eye movement makes eye tired
	Stop the video, playback and forth many times
Temporal Demand (Visualization Mean = 4.64)	**Temporal Demand (Video Mean = 10.21)**
NO rush	I don't want to slow down too much as I am not patient enough. I want to quick and correct. So, I am in a rush.
	I need to catch so much information at the same time, making me feel rushed.
	When I familiar with the diagnosis tasks and familiar with the problem-solving process, it's less rushed.

Table 4.5 continued

Performance (Visualization Mean = 8.71)	**Performance (Video Mean = 11.21)**
The diagnosis task itself is hard.	
Unexperienced in this kind of tasks	
Unconfident in interpreting data correctly	
I just saw one or two tasks of one person. If I could see more tasks, it may be better.	
	Video is hard to track
	I need to diagnose base on my memory of students' problem-solving process, which I can't ensure is correct.
Effort (Visualization Mean = 10.64)	**Effort (Video Mean = 13.29)**
All the reasons listed above	
Frustration (Visualization Mean = 8.07)	**Frustration (Video Mean = 12.14)**
Insecure: not sure if my diagnoses are correct.	
	Insecure: Data is moving, not sure whether I remember everything correctly.
	Stressed: Refresh memory many times.

4.5 Diagnosis Process Data Analysis and Results

In In the interview, evaluators were asked to describe their diagnosis process, including what elements they looked at to make their diagnoses and how they got their conclusions. I transcribed the interview recording (Appendix D) and coded the transcript using the top-down approach. The purpose of the interview is to figure out how evaluators explored the student problem-solving processes to achieve their judgments on student problem-solving strategies and difficulties. Hence, I had a small set of codes. The codes were based on the theoretical framework of arithmetic problem-solving, illustrated in Chapter 3 regarding specific student problem-solving activities may indicate a specific problem-solving approach. I color-coded the interview transcript into elements, element attributes, target areas, insights, and conclusions (Table 4.6). Figure 4.11 is an example of color-coded interview transcripts.

P9-B3.5-video-interview transcript

Me: Could you describe your diagnosis process of task P9-B3.5, why you choose the problem-solving strategy that the student used is model-based problem solving strategy, and the student may have difficulty in the diagram equation.

E3: The students' fixations on diagram equation are very long like here, and here, And the student went back to look at it again and again, so I think the student used model-based problem solving strategy.

Me: So you looked at the diagram equation and thought the student tried to understand the diagram equation

E3: Yeah, from the fixations here (point at the fixations located at question area), the duration and regression here, so he try to understand the problem first and then he try to understand the question again. He try to understand both of them, the problem and equation, so I think it's model based.

Me: OK.

Me: So, you choose the student had difficulty in understanding the diagram equation.

E3: Yeah, but it's a little difficult to decide. The fixations (on the question area) are long, but compare to diagram equation, the fixations on diagram equations are relative longer, so I decided the student may have difficulty in diagram equation.

Me: But the student answered correct.

E3: Yeah. For the answer, he may not have the problem, but for the understanding of equation and problem, he may have the difficulty.

Figure 4.11 Coded Interview Transcript

Table 4.6 Codes and the Summary of the Transcript in Figure 4.11

Task: P9-B3.5 Analysis method: Video					
Codes	Elements	Element Attributes	Target Areas	Insights	Conclusion
Problem-Solving Strategy	Fixation	Long	Diagram equation		Model-based Problem-Solving Strategy
	Regression	Again and again	Diagram equation		
	Fixation	Duration; Regression	Question area	The student tried to understand question first, then tried to understand equation.	
Difficulties	Fixation	Long	Question area		Difficulty in Diagram Equation
	Fixation	Longer	Diagram equation		

After analyzing all interview transcripts, the elements and corresponding attributes employed by evaluators were summarized in Table 4.7. In Table 4.8, elements and attributes that related to diagnosis results are listed. The elements mentioned by every evaluator in their diagnosis process statements are bolded. Table 4.8 suggests that fixation distribution, performance correctness, and the number of mouse drags are critical elements for problem-solving strategy's diagnosis. Evaluators employed these elements alongside other student learning information to diagnose what problem-solving strategies that students used in a task. For example, an evaluator told me: "The student read the whole problem (*fixation-distribution*), not only focus on keywords. He correctly answered the questions (*performance-correctness*). So, I think the student used model-based problem-solving strategies". Another evaluator described his diagnosis process: "The student put 87 in the correct place at first, but when he got feedback that he was wrong *(performance-wrong)*, he directly changed (*drag-count*) the position of 87 without thinking (*time between feedback and operation-short*). He didn't think why he was wrong. So, I think the student used guess and check strategy".

Table 4.7 Students' Problem-Solving Metrics and Attributes

Element	Fixations	Regression	Time on task	The time between reading/feed-back and operation	Performance	Drag
Attribute	Count Duration Distribution	Count Distribution	Duration	Duration	Correctness	Count Speed Switch

Table 4.8 The elements that evaluators used to diagnose students' problem-solving strategies

Problem-Solving Strategies	**Fixations**	**Regression**	**Time on task**	**The time between reading/ feedback and operation**	**Performance**	**Drag**
Guess & Check	Distribution Duration		Duration	**Duration**	**Correctness**	**Count** Switch Speed
Model-Based Problem Solving Strategies	**Distribution** Count		Duration		**Correctness**	Speed Count
Keyword Strategies	**Distribution** Duration Count	Count Distribution	Duration		**Correctness**	

To diagnose a student's difficulties in solving mathematical problems, evaluators analyzed more details. Besides looking at the elements mentioned above, they tried to use the students' perspectives to understand students' thinking processes. For example, an evaluator described a student's problem-solving process as follows: "The student was wrong on the first try (*performance-correctness*), but she kept putting 14 in the first box. She knew 14 should go to Lauren. She switched "a" and 26 (*switch*). She didn't look at problem keywords in this process (*fixation -distribution*) but kept looking at the diagram equation (*fixation - distribution*), she may have difficulty with the diagram equation". The evaluator concluded that the student used a guess and check strategy to solve the problem as she had difficulty in understanding the diagram equation and couldn't arrive at the correct answer.

Student problem-solving strategy and difficulty diagnosis is a hard task. But with student problem-solving data (eye movement, mouse movement, performance, etc.), researchers may have a bridge to understand student cognition processes and identify their strategies or difficulties. Both the visualization system and video can present problem-solving processes for evaluators. The elements that evaluators used to complete diagnosis are similar in both analysis methods.

4.6 Preference Interview Data Analysis

Interview Question 4: If the required training time of the visualization system is about 30 minutes, are you willing to use this system in your actual student problem-solving diagnosis for your current project/current class? If so, why is it worthwhile? If not, why not?

All evaluators agreed that the training time was worthwhile. Less than one hour of training time was totally fine since visualization could save them more time later.

As the answer to this question was so consistent, they were further asked which method they preferred between visualization and video? Why?

Among the six evaluators, two of them prefer visualization, one of them prefer video, and the other three evaluators would like to combine these two methods or use the two methods in different contexts to take the advantage of the two methods. So here, I summarized the advantages and disadvantages of visualization.

The advantages of visualization:

1. Visualizations are graphs, and users can control the time spent on each graph.

2. The visualization presents data statically, so users don't need to memorize anything.

3. The visualization can preserve the quantity of data, for example, the number of fixations, the speed of drag. While looking at the video, users need to judge the quantity of data by themselves.

4. The visualization presents extra summarized information that can't be identified on video, such as the average metrics of performance groups.

The disadvantages of visualization:

1. Compared to video, analyzing visualization is more complex, especially for inexperienced users. While the video is easier to understand as everything is presented as is.

2. The video presents data as is, which can include some background information. For example, students moving the mouse back and forth can reflect their hesitation. The data presented by the visualization system has been sorted, filtered, and aggregated, and some detailed information was filtered out.

Both these two methods have their advantages and disadvantages, that's the reason why many of the evaluators would like to use these two methods together. They stated: "If I have a lot of students/tasks, I will use visualization. If I have time, I will look at the videos, as video gives more context. After looking at the video, I would like to go back to visualization to validate my judgments/findings."

4.7 Conclusion

Evaluation data analysis shows that with the visualization system, users can complete the diagnosis tasks significantly quicker than using the video analysis method. For the diagnosis consensus rate, evaluators can get an equivalent or even higher level of consensus using the visualization system. The self-reported task load measurements revealed that visualization requires a smaller task load.

Especially for physical and temporal demand, using visualization requires significantly fewer demands than using video. But, it should be noted that the mental demand for diagnosis tasks is high no matter which analysis method is used. Evaluators expressed their preference for both analysis methods, and some would like to use them together or choose one of them in the specific scenario. Finally, all evaluators agreed that, considering the advantages of visualization, less than one hour of training time is worthwhile.

CHAPTER 5. CONCLUSION

This chapter first answers the three evaluation questions based on the evaluation data analysis results from the previous chapter. This section illustrates the advantages and disadvantages of the visualization system in helping people understand students' problem-solving processes and provides insights into the problem-solving data elements. This chapter subsequently summarizes the visualization development framework, which can also be applied to the other visualizations.

5.1 Answers to Evaluation Questions

There are three evaluation research questions. This section states the conclusion to the evaluation questions based on the evaluation data analysis results and interview data.

Evaluation Question 1: Cosmpared with videos, can the visualization system allow educational researchers to spend less time identifying students' problem-solving strategies and their problem-solving difficulties?

Yes, compared with video, the proposed visualization system can make educational researchers spend less time to identify students' problem-solving strategies and their problem-solving difficulties.

Evaluators spent almost double the time using the video analysis method compared to the visualization method (308.64 seconds VS. 198.54 seconds). The analysis results show that the diagnosis time using the visualization system was significantly lower than using the video method (P-value <0.0001). After doing more tasks, the diagnosis time of using both methods decreased greatly, which can be reflected in differences between the round 1 and round 2 average diagnosis time. For both visualization and video methods, the round 2 diagnosis time decreased by about 110 seconds compared to round 1 diagnosis time. But still, the visualization method led to significantly lower diagnosis time than the video. That's because, after a short training to become familiar with the different charts in the visualization system, users can directly pick the information they need from the visualization system. While using the video analysis method, users need to follow the video to get information. Although they can speed up and speed down, still, they need to spend time following the video. Besides following videos, playback is another time-consuming operation. Video users are passive information receivers in that they can't decide when they will get certain information. When something pops out, they may miss it without preparation. Then, users need to do a playback. But the starting point of playback is vague. Users need to locate the timepoint based on an estimation. That process is time-consuming and frustrating, especially when users play a video back many times. Often users lose patience and give up. Just as we observed in the evaluation, some evaluators tried to avoid playback at the expense of missing data while subject to high mental demands.

Evaluation Question 2: Compared to videos, can the visualization system help educational researchers get more reliable diagnoses on the problem-solving diagnosis tasks?

Yes, the visualization system can help educational researchers get equivalent or even more reliable diagnoses on the problem-solving diagnosis tasks.

The evaluation was designed to answer this question by distributing a set of diagnosis questions to evaluators, including visual attention pattern, drag pattern, performance, problem-solving strategies, difficulties, and performance group. For diagnosis tasks: identify students' mouse drag pattern, identify students' problem-solving performance (whether they are correct or incorrect), and determine which performance group the students belong to. In these tasks, evaluators got higher correctness rates; their diagnoses are more reliable. In the other diagnosis tasks such as problem-solving strategies and difficulties, there are hardly any 'correct' answers, especially in situations with no opportunity to talk with the student face-to-face. Thus, consensus rate was used in the evaluation data analysis to indicate whether most of the evaluators agreed with the answer. If more than half of the evaluators agreed with the answer, the answer was recognized as the "correct" answer. Thus, the consensus rate was used as an indicator of diagnosis reliability.

In these diagnosis problems, diagnosis tasks using the visualization method got higher correctness rates/consensus rates relative to video analysis. In visual attention, problem-solving strategies, and difficulty diagnoses, the consensus rates using the visualization analysis method were fifteen percent higher than using the video analysis method (73% ~ 81% vs. 58% ~ 67%). In the drag pattern and performance pattern diagnoses, the visualization method got the same level of consensus rates as the video analysis method (89% ~97% vs. 85% ~ 93%).

The reasons why evaluators got lower correctness/consensus rates using video analysis method could be due to the pressures of time, unreliable of memory, and individual differences.

Time pressure: the video method required a relatively long time and many playbacks to analyze, which gave evaluators some time pressure. Although in the experiment study, there was no time limit, they were told to take as long as needed. Many evaluators expressed a sense of urgency. They want to complete the tasks quickly. If they felt it took too long, they became impatient. Time pressure might influence their performance and cause a lower consensus rate.

Limits of memory: visualization presented all information through graphics, and evaluators searched for information to diagnose and verified their diagnoses anytime they wanted. But video presented information temporarily. Evaluators needed to remember what they saw, and then diagnose based on their memory, which can be shown in their interview transcription that they often use words: fell, impression, remember. People's working memory has limits (Schweppe & Rummer, 2014; Lee Swanson, 2011). Evaluators might remember something inaccurately, especially in long videos. When information is too much for people to memorize, people will primarily select the information they deem important. The information selection process focuses people's attention on key information, but at the same time, it also leads to the neglect of other information. This kind of limitation may not happen in the visualization analysis method, as the visualization system does not require users to maintain information in working memory.

Individual differences: as mentioned in the above sections, the video analysis method caused time pressure and required a large memory load. People's response to stress and cognitive abilities are different. That directly influenced the amount of information that evaluators accessed. Also, the above section pointed out that with limited memory ability, people will attend to the information they think is important. Then for different evaluators, they might attend to different information. For example, drag operations directly related to students' problem-solving, so every evaluator paid a lot of attention to it. But except that, some evaluators followed students' eye trajectory closely while some other evaluators paid attention to the time students spend on the problem content and tasks. The information type and mount differences might cause lower consensus rates.

Besides, visualization gave more summarized information for evaluators. Also, it should be noticed that the video can also provide information not available in the visualization system. In the visualization system, summarized information such as the number of fixations on keywords, number of fixations on other words, time spent on the task, drag speed were calculated and illustrated in the radar chart of the visualization. These data provided clear and direct support for evaluators' diagnosis that they didn't need to feel or estimate anymore. More than that, the visualization system presented the data measurements of high-performance, median performance, and low-performance group for evaluators as diagnosis reference. That's the reason why evaluators were more confident in their visualization diagnosis results.

There was one limitation of the visualization system. The mouse movement data were abstracted and aggregated into the start point, the endpoint, and the drag speed. The mouse trajectory was not presented in the visualization system. However, evaluators found that students dragged tags into one diagram box, then, before they dropped the tag into the box, they might change their mind and put it into another diagram box. This information was filtered out and not presented in the visualization, but it was important to determine whether the students had confusion or difficulties. That was the shortage of the visualization design, but it might not be reflected from the consensus rate. Future, this information should be added to the visualization system.

Evaluation Question 3: What aspects of the visualization system/video result in higher workload to educational researchers?

For the visualization system, interpreting data and combining heterogeneous data together to build the student problem-solving situation model is the first workload source. Based on the student problem-solving situation model to identify patterns, strategies, and difficulties is another source of high workload.

For the video, remember a student's problem-solving process and interpret it into the student problem-solving situation model in mind is the first workload source. Based on the memorized problem-solving situation model to identify patterns, strategies, and difficulties is another source of high workload of video method.

Both analysis methods had high mental workloads as understanding students' cognitive processes underling the problem-solving processes itself is a hard task. The uncertainty of student problem-solving data interpretation led to evaluators' frustration in both analysis methods. The flow characteristic of the video led to higher demands in temporal and physical aspects. The unconfident feeling in memory cased the higher performance failure feeling and higher frustration. In summary, the evaluation results showed that the visualization method could help educational researchers diagnose students' problem-solving patterns faster and more reliable compared with the video analysis method. In another word, the proposed visual analytic method can help educational researchers to understand students' problem-solving processes in the computer-assisted math learning program.

But it should be noticed that when evaluators were asked about their preference in the interview, many of them want to employ both methods. They agreed that they want to use the visualization system if there are too many tasks and they would like to save time. But when there is enough time, they would like to use the video. Or they would like to look at the video at first, then look at the visualization to search the information they need. From the interview data analysis, there are two reasons for that. The first is video can present students' problem-solving process as it is which makes evaluators can easily build their problem-solving situation mode. The second reason is that video presented all the information directly while visualization presented filtered, aggregated data. Evaluators could get more background information from the video, and they think the data before processing is more reliable. So evaluators would like to use both of them or select one based on the demand.

5.2 Problem-Solving Data in CAL

The literature review about students' problem-solving data analysis showed that students' performance data, mouse movement data, and eye movement data are important data that can reveal students' cognitive processes underlining their problem-solving behaviors (Catrysse et al., 2018). The dissertation study proved the function and the importance of these data as the visualization system visualized these data and got an equivalent and an even higher level of reliable diagnosis results. The dissertation study further identified the specific data elements that can help educational researchers to understand students' problem-solving processes. The data elements include eye fixations, eye regressions, time on task, time before operation, number of drag attempts, and performance correctness. These elements were used again and again by all the evaluators in the study no matter what analysis methods they employed. Eye fixations and regressions were used to determine students' visual attention, their focus, and whether they had confusion/difficulties. For example, if a student looks at a specific word again and again (regression), the student may think the work is important, or the word is confusing to him(Zachary Jacobson & Dodwell, 1979; Rayner & Pollatsek, 1989). Also, the eye fixations and eye regressions indicate whether the student is involved in the problem-solving process. Time on task and time before operation have similar functions. For example, if a student immediately starts to operate after they got feedback, the student either already knows the answer or the student just want the work is done. It could be determined by combining other data. The number of drag attempts can reflect students' thinking processes. If a student drags a tag everywhere, the student may don't know the answer. He/she is guessing. While problem-solving correctness is the most direct revealer of students' problem-solving skills. Evaluators used these data to determine students' problem-solving strategies and identify their difficulties. For example, an evaluator described his diagnose process. The student wronged in the first try, directly operate after he got feedback, then wronged again. It seems the student has no thinking time before the operation. Besides that, the student didn't look at the problem content a lot. He kept looking at the diagram equation. So I think the student may use guess and check strategy. He should pay more attention to the problem content, and he may have confusion/difficulty in the diagram equation.

This study proved students' problem-solving data, including eye movement, mouse movement, and performance data, collected by the computer-assisted learning program could help educational researchers understand students' problem-solving and further benefit the teaching and researching of teachers and educational researchers.

5.3 Problem-Solving Data Visualization

The dissertation study developed a visual analytic method for teachers and educational researchers. The development method is: first, review the literature to identify possible problem-solving behavior patterns associated with problem-solving strategies and difficulties. Second, based on the possible behaviors, determine the data should be employed in the problem-solving process understanding. Third, based on the data type and analysis purpose, design and develop visualization to present data, illustrate patterns within the data. Forth, invited target users to review the visualization prototype. Forth, design iteration. In the dissertation study, after review literature, students' eye movement, mouse movement, performance are selected, and the association between different problem-solving patterns and students' problem-solving patterns were summarized. The dissertation visualization study proved and further developed the association. In the visualization design and development process, context-based, simple geometry icons were employed to illustrate the problem-solving data instead of using the traditional heatmap and gaze plots. Actually, Heat maps were also provided to evaluators in the study, but it was barely used in the diagnosis process, as it only presented accumulated data with no temporal information. And the information presented by the heatmap already clearly illustrated in the visualization system and even more detailed. The iteration design and prototype evaluation study adjusted the design direction and make sure it was designed to fulfill the user demands.

The development process ensured the usability of the visualization system and provided a new visual analytic method for teachers and educational researchers to identify students' problem-solving strategies and difficulties more quickly and reliable compared with the video method.

CHAPTER 6. DISCUSSION

This chapter will discuss the benefits of the visualization systems for educators, teachers, and parents. This chapter will also discuss the implications of the present study for education and data visualization based on the findings in previous chapters. The limitations of the dissertation study are discussed in this chapter and the approaches could be applied to eliminate the limitations are introduced. Future research directions are also illustrated in this chapter.

6.1 The Value of the Visualization System for Educational Researchers

Here educational researchers are target users of the visualization systems. As researchers, they have the demand to study student problem-solving behaviors and cognition processes. The evaluation conclusions in the previous chapter have shown that, compared to the video analysis method, the visualization system can better help researchers to make assumptions, identify patterns, and diagnose student problem-solving strategies in terms of diagnosis efficiency and consistency aspects.

The problem-solving data visualization method proposed in this dissertation affords a new method for educators to comprehend student eye movement, mouse movement, and performance. More than that, the visualization system gives statistical metrics (mean, standard deviation, etc.) and group performance to help researchers have a relatively comprehensive understanding of the student knowledge, skills, and performance. The proposed visualization method can improve educators' analysis efficiency by providing rich data and presenting data statically. As all data are presented statically on the screen, educators have control over time and analyzing pace. With more time and all of the data on the screen, the risk of missing information is decreased compared to the video analysis method.

In addition, the visualization method also provides a new communication method for educators to facilitate their discussion, illustration, and presentation. With static visualization, researchers can present their analyzing processes, assumptions, and conclusions at the same time using graphics to support their statements. The rich data and statistical metrics delivered in the visualization will help to strengthen the persuasion of researchers' arguments.

Finally, researchers may use screenshots of the visualization in their publication to support their writing. Using figures in publications is common. However, most of the time, authors only depict one kind of data in the figure, such as eye movement (Yoon & Narayanan, 2004), mouse movement (Zushi, Miyazaki, & Norizuki, 2012), etc. Few of them present multiple data sources in one figure, although they may reach their conclusions by analyzing multiple data sources. This may relate to the writing styles of authors. If authors discuss data sources separately in their writing, it is natural for them to present different data sources respectively. But sometimes, authors present multiple data types separately is due to the complexity of data. Authors do not know how to present all data in one figure, although the data themselves are naturally associated. Here, the visualization system proposed in this study presents all data sources within the problem-solving context and depict the relations among the data sources. Authors may use the visualization in their paper to ease the understanding of their writing and enhance the convincing of their paper.

6.2 The Value of the Visualization System for Teachers

The target users of this study are not teachers due to two reasons. First, the design of the visualization system is relatively complex which necessitates training and practice. Teachers often have a tight schedule of teaching, writing, grading, and taking care of kids. They may not have the time to use the visualization system. I have introduced the visualization to an elementary school teacher at an educational conference (AERA 2019). She agreed with the value of the visualization system for assisting teaching but also mentioned that the complexity of the visualization system would take teachers a lot of time and energy to comprehend. Secondly, the visualization was designed to help researchers make assumptions, detect patterns, and identify strategies. Aiming at the design purpose, the visualization system provides rich information such as problem content, eye fixation sequences, mouse movement history. But for teachers, the conclusions about student difficulties and problem-solving strategies are much more important than the comprehension of student problem-solving processes. They need simple and conclusive judgment based on which they can provide personalized instruction or adjust their curriculum. For the above reasons, teachers are not the target users of the visualization system.

But this does not mean the visualization system has no benefits to teachers. Teachers can still take advantage of the visualization system to learnstudent problem-solving strategies, comprehend student problem-solving processes, and understand student learning difficulties. The evaluators in the dissertation study indicated that the visualization system could save them a lot of time, and they preferred the visualization system over video if they have more than five students. The feedback from the evaluators shows that the visualization system may be a good choice if teachers want to observe student problem-solving process, since most elementary school teachers have more than five students.

Other than analyzing student problem-solving processes, teachers can also use some specific functions in the visualization system to quickly figure out student performance and status. For example, inspecting the student performance bar chart, teachers can quickly learn how long the student spends on the problem and his/her correctness. Teachers can also learn the performance of all students in the class. Besides the performance bar chart, the heatmap over the problem content shows the attention distribution of a student and the attention allocation of high-performance students. Teachers can instantly recognize the student's reading focus and the attention differences between one student and the high-performance group by looking at the heatmap.

To further benefit teachers, more work is needed to improve the visualization system. One direction is enabling the visualization system to provide judgments on the student problem-solving patterns, difficulties, and strategies for teachers. For example, for a specific student, the system can list student problem-solving characters and give judgments on possible strategies and difficulties:

Characters:

1. The student answered the question incorrectly;
2. The student fixated on keywords for 10 seconds while looking at other words for 3 seconds;
3. The student regressed on the word "fewer" 8 times about 3 seconds;

Judgments:

1. The student may have difficulty understanding word "fewer".
2. The student may have little understanding of the problem context.
3. The student may use keyword strategy.

The interview data from the evaluation experiment in the dissertation study shows that evaluators use many metrics such as fixation duration and number, mouse click number, time on task to identify student problem-solving difficulties and strategies. Currently, the judgments from evaluators are subjective. But in the future, we can summarize their judgment model and quantify judgment metrics. Then, the model and metrics can be employed to enable the system to provide student characters, strategies, and difficulties for teachers.

Finally, the visualization system can work together with the online educational programs to be used on student performance reports. Many educational programs on the market provide performance reports but these are limited to basic information such as visit frequency, correctness, and tasks completed. Employing the visualization system onperformance report in the educational program can provide more valuable information for teachers and parents.

6.3 Implications for Education

The visual analytic method could be applied to other areas of education. In this study, student problem-solving data collected in the CAL programs were analyzed and visualized. Besides mathematics problem-solving programs, there are many other computer programs on the market designed to teach students math, reading, science, and so on. These programs can provide plenty of student learning data. The visual analytic method developed above provides a fast and reliable method for people to take advantage of the student learning data and understand students' learning processes.

The visual analytic method can be applied to many education areas such as, but not limited to, reading, coding, problem-solving, etc.

Besides the visual analytic method, the data elements identified in this study can contribute to data mining in education contexts. This dissertation study identified key elements that help to identify students' problem-solving strategies and difficulties, including fixation numbers, distribution, regression numbers, time on task, the time before operation, etc. These elements can be considered as independent variables to build prediction models of students' performance. Using visual analytics to identify the important data elements and then employ these elements to build the learning data mining model is another direction of implication in education.

6.4 Implications for Data Visualization

The developed visualization system presented student eye and mouse movement, and performance data. The student-problem-solving data in the dissertation study is typical spatial-temporal data. The data contains timestamps presenting the time attributes of the problem-solving process. The data has coordinates presenting the spatial attribute of students' mouse and eye movement on the computer screen. The visualization develop approach used in this study employed a top-down approach: identify data patterns that associate with specific insights, and then, based on the identified patterns to aggregate, analyze, and illustrate data using context-based visualization. The top-down approach is the main spatial-temporal data visualization direction (Bailey-Kellogg, Ramakrishnan, & Marathe, 2006; Chiu & Russell, 2011; Jänicke, Heine, & Scheuermann, 2013). But the student problem-solving data has a special property in the logical structure of the problem, or in other words, the cognitive process of students. Often spatial-temporal data has no complex logical structure. For example, GPS data is spatial-temporal data. It contains the information that at a specific time point, someone is at a specific position. It may present people's movement patterns. It may reveal people's movement outliers. But there is no complex cognitive process behind the data. Then there is no necessity for the visualization to present cognitive processes. But student problem-solving data visualization has this requirement. It requires visualization to present the context information and the problem structure clearly to help users understand the cognitive processes of students. Thus, it's more complex than typical spatial-temporal data. The visualization approach in this study proposed context-based visualization, which presents the situation that students faced to the visualization users and let users see the processes of the student. This context-based visualization can also be applied to the visualization of other spatial-temporal data involving others such as electroencephalograph (EEG) data, CAL data, interaction data, and so on.

6.5 Limitations

There are some limitations in the study that should be considered in future studies. The limitations concern eye movement data quality, sample size, and the variety of evaluators.

Eye movement data has been used in many papers. Eye movement data quality is a topic discussed more and more often in the literature (Nyström, Andersson, Holmqvist, & van de Weijer, 2013; Blignaut & Wium, 2014; Dalrymple et al., 2018). Sustaining eye movement data quality is important as it directly relates to the reliability of conclusions. To ensure eye movement data quality, in the dissertation study, I made participants sit within the advised operational distance, with consistent illumination, and removed participants with no fixations on the Area of Interest. But I found more measures could be adopted to sustain eye movement data quality, including using a black/white background, keeping student eyes centered in the detecting box, and recalibrating when students move away from the box.

Another limitation of the dissertation study is the small sample size of evaluators and tasks. Due to time and financial constraints, only six evaluators were recruited to analyze two tasks from each of the nine students. If more evaluators could be recruited, we would have more subjects complete the diagnosis tasks using the visualization and video analysis methods. Thus, we would obtain more robust results for diagnosis efficiency and consistency. If evaluators can analyze many tasks from the same student, evaluators can get a more comprehensive understanding of the student's characteristics, problem-solving skills, and cognition processes. In that way, evaluators can give more reliable diagnoses on the student problem-solving patterns, difficulties, and strategies. Finally, increasing the number of students will increase problem-solving data size from which evaluators can identify representative student patterns, characteristics, and difficulties.

The third limitation is that all evaluators were educational researchers. Visualization designers and elementary school teachers should also be considered as potential evaluators of the visualization system. A visualization system is not only about function but also about graphic design. Good design will ease the use of the system, decrease users' cognitive load, and avoid confusion. Visualization designers do not need to do diagnosis tasks but examine the design of visualizations. Teachers should be another group of evaluators as they are potential users of the visualization system. They interact with elementary school students every day, so they may deeper understanding of student' behavior. They teach elementary school students, and understanding students' difficulities is very important. Feedback from teachers is beneficial to the future improvement of the visualization system. However, due to limited time and financial resources, only the direct target users – educational researchers were involved in the evaluation study.

6.6 Future Work

Future work directions include decreasing the limitations of the study, improving the visualization system, and further exploring the potential of the visual analytic method in education.

Decrease the limitations of the study. In section 6.5, the limitations of the dissertation study were discussed. In future studies, many measures can be adopted to decrease the limitations. For a specific eye tracker, researchers should check the corresponding eye-tracking equipment user manual to implementthe ideal conditions for the eye movement data collection and sustain data quality. The conditions include many variables such as gaze angle, illumination, background color, eye placement in box, etc. (Tobii Technology, 2011). Another limitation of the dissertation study was the small sample size. I have discussed the advantages of increasing sample size in the previous section. However, the fact should be noted that even for the relatively small sample size, each evaluator spent roughly three hours on the evaluation. Due to the complexity of the study and evaluation experiment design, increasing one variable size will multiplyevaluation time. For example, if each evaluator needs to analyze four tasks for each student instead of two tasks, the evaluation time for each evaluator will increase from three hours to six hours. Therefore, for future studies, it is critical for researchers to know the study's purpose and expected conclusions so that researchers can reference the goal to determine the sample size of variables.

Improve the visualization system. In the evaluation study, compared with the video analysis method, the visualization system showed its superiority in diagnosis speed while maintaining diagnosis reliability. The self-report task load survey showed that the visualization system requires lower levels of workload. But in the interview, evaluators stated that video analysis is needed as it shows more information. Thus, there are two research tasks in future work to improve the visualization system.

First, it may be useful to present students' mouse movement data completely. Current mouse movement data visualization presented some key elements such as mouse drag start points, endpoints, and drag speed. But other important elements such as drag attempts were filtered out. Retaining these data and presenting them efficiently is one potential next step for the visualization development.

Second, it would be ideal to decrease users' mental demands when combining different types of data. In the current visualization system, different data types such as mouse movement and eye movement were presented separately, although they share the same timeline and same visualization space. Users need to associate these by themselves. If the visualization can directly present these data together, users' mental demands can be greatly decreased. It is a challenge because of the complexity of the data, but it is also an opportunity to improve the efficiency of the visualization.

Explore the potential of the visual analytic method in education. In the dissertation study, the visual analytic method is applied to present arithmetic word problem-solving processes. The visual analytic method can be extensively applied to other online learning data. As long as the online educational programs ask students to read and interact with the program and record the data, the visual analytic method has the potential to help researchers and teachers to comprehend the learning process and identify student learning patterns.

As a matter of fact, the educational program employed in the dissertation study – COMPS-A© (Xin, Kastberg, Chen, & Team, 2015-2020) covers not only arithmetic word problem-solving but also counting, which was peripherally touched on in the pilot study in Chapter 3. Applying the visual analytic method to more online learning areas will extend its applicable range and may be beneficial to educational research, teaching, and learning.

These are some potential directions for the future work of visualization system development. I will start from the dissertation study improving the visualization system, perfecting the study and evaluation experiment design, and applying the visual analytic method to the counting strategies in the COMPS-A© (Xin, Kastberg, Chen, & Team, 2015-2020).

REFERENCES

Ahn, J., Gubbels, M., Yip, J., Bonsignore, E., & Clegg, T. (2013). Using social media and learning analytics to understand how children engage in scientific inquiry. *Proceedings of the 12th International Conference on Interaction Design and Children - IDC '13*, 427–430. https://doi.org/10.1145/2485760.2485805

Alesi, S. M., & Trollip, S. R. (2001). *Multimedia for Learning: Methods and Development*. Pearson.

Ali, L., Hatala, M., Gašević, D., & Jovanović, J. (2012). A qualitative evaluation of evolution of a learning analytics tool. *Computers & Education*, *58*(1), 470–489. https://doi.org/10.1016/J.COMPEDU.2011.08.030

Alina Zapalska, Dallas Brozik, D. R. (2012). Development of Active Learning With Simulations and Games. *US-China Education Review*, *2*, 164–169.

Antonenko, P. D., Toy, S., & Niederhauser, D. S. (2012). Using cluster analysis for data mining in educational technology research. *Educational Technology Research and Development*, *60*(3), 383–398. https://doi.org/10.1007/s11423-012-9235-8

Azcarraga, J., & Suarez, M. T. (2013). Recognizing Student Emotions using Brainwaves and Mouse Behavior Data. *International Journal of Distance Education Technologies*, Vol. 11, pp. 1–15. https://doi.org/10.4018/jdet.2013040101

Bailey-Kellogg, C., Ramakrishnan, N., & Marathe, M. V. (2006). Spatial data mining to support pandemic preparedness. *ACM SIGKDD Explorations Newsletter*, *8*(1), 80–82. https://doi.org/10.1145/1147234.1147246

Bakhshinategh, B., Zaiane, O. R., ElAtia, S., & Ipperciel, D. (2018). Educational data mining applications and tasks: A survey of the last 10 years. *Education and Information Technologies*, *23*(1), 537–553. https://doi.org/10.1007/s10639-017-9616-z

Bartolotti, J., & Marian, V. (2012). Language Learning and Control in Monolinguals and Bilinguals. *Cognitive Science*, *36*(6), 1129–1147. https://doi.org/10.1111/j.1551-6709.2012.01243.x

Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. *Washington, DC: SRI International*, 1–57. https://doi.org/10.2991/icaiees-13.2013.22

Blignaut, P., & Wium, D. (2014). Eye-tracking data quality as affected by ethnicity and experimental design. *Behavior Research Methods*, *46*(1), 67–80. https://doi.org/10.3758/s13428-013-0343-0

Bolden, D., Barmby, P., Raine, S., & Gardner, M. (2015). How young children view mathematical representations: a study using eye-tracking technology. *Educational Research*, *57*(1), 59–79. https://doi.org/10.1080/00131881.2014.983718

Burch, M., Kull, A., & Weiskopf, D. (2013). AOI Rivers for Visualizing Dynamic Eye Gaze Frequencies. *Computer Graphics Forum*, *32*(3pt3), 281–290. https://doi.org/10.1111/cgf.12115

Catrysse, L., Gijbels, D., Donche, V., De Maeyer, S., Lesterhuis, M., & Van den Bossche, P. (2018). How are learning strategies reflected in the eyes? Combining results from self-reports and eye-tracking. *British Journal of Educational Psychology*, *88*(1), 118–137. https://doi.org/10.1111/bjep.12181

Cetintas, S., Si, L. S. L., Xin, Y. P. X. Y. P., & Hord, C. (2010). Automatic Detection of Off-Task Behaviors in Intelligent Tutoring Systems with Machine Learning Techniques. *IEEE Transactions on Learning Technologies*, *3*(3), 228–236. https://doi.org/10.1109/TLT.2009.44

Cetintas, S., Si, L., Xin, Y. P., & Hord, C. (2009). Predicting Correctness of Problem Solving from Low-level Log Data in Intelligent Tutoring Systems. *Proceedings of the 2nd International Conference on Educational Data Mining*, 230–239.

Chen, M. C., Anderson, J. R., & Sohn, M. H. (2001). What can a mouse cursor tell us more? *CHI '01 Extended Abstracts on Human Factors in Computing Systems - CHI '01*, 281. https://doi.org/10.1145/634067.634234

Chen, Y., Chen, Q., Zhao, M., Boyer, S., Veeramachaneni, K., & Qu, H. (2017). DropoutSeer: Visualizing learning patterns in Massive Open Online Courses for dropout reasoning and prediction. *2016 IEEE Conference on Visual Analytics Science and Technology, VAST 2016 - Proceedings*, 111–120. https://doi.org/10.1109/VAST.2016.7883517

Chiu, C. Y., & Russell, A. D. (2011). Design of a construction management data visualization environment: A top-down approach. *Automation in Construction*, *20*(4), 399–417. https://doi.org/10.1016/j.autcon.2010.11.010

Christmann, E. P., & Badgett, J. L. (2003). A Meta-Analytic Comparison of the Effects of Computer-Assisted Instruction on Elementary Students' Academic Achievement. *Information Technology in Childhood Education Annual*, *2003*(1), 91–104.

Copeland, L. (2016). Eye Tracking to Support eLearning (Dotoral thesis, Australian National University). https://doi.org/10.25911/5D7789D9C492C

Dalrymple, K. A., Manner, M. D., Harmelink, K. A., Teska, E. P., & Elison, J. T. (2018). An examination of recording accuracy and precision from eye tracking data from toddlerhood to adulthood. *Frontiers in Psychology*, *9*(MAY). https://doi.org/10.3389/fpsyg.2018.00803

De Corte, E., & Verschaffel, L. (1986). Eye-movement data as access to solution processes of elementary addition and subtraction problems. *67th Annual Meeting of the American Educational Research Association (San Francisco, CA, April 16-20, 1986)*.

De Witte, K., Haelermans, C., & Rogge, N. (2015). The effectiveness of a computer-assisted math learning program. *Journal of Computer Assisted Learning*, *31*(4), 314–329. https://doi.org/10.1111/jcal.12090

Doleck, T., Jarrell, A., Poitras, E. G., Chaouachi, M., & Lajoie, S. P. (2016). Examining diagnosis paths: A process mining approach. *Proceedings - 2016 2nd International Conference on Computational Intelligence and Communication Technology, CICT 2016*, 663–667. https://doi.org/10.1109/CICT.2016.137

Dyckhoff, A. L., Zielke, D., Bültmann, M., Chatti, M. A., & Schroeder, U. (2012). Design and implementation of a learning analytics toolkit for teachers. *Educational Technology and Society*, *15*(3), 58–76. https://doi.org/10.1177/0002764213479367

Fletcher-Flinn, C. M., & Gravatt, B. (1995). The Efficacy of Computer Assisted Instruction (CAI): A Meta-Analysis. *Journal of Educational Computing Research*, *12*(3), 219–241. https://doi.org/10.2190/51D4-F6L3-JQHU-9M31

Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, *42*(1), 226–241. https://doi.org/10.3758/BRM.42.1.226

Gleason, M., Carnine, D., & Boriero, D. (1990). Improving CAI Effectiveness with Attention to Instructional Design in Teaching Story Problems to Mildly Handicapped Students. *Journal of Special Education Technology*, *10*(3), 129–136. https://doi.org/10.1177/016264349001000301

Handal, B., & Herrington, A. (2003). Re-Examining Categories of Computer-Based Learning in Mathematics Education. *Contemporary Issues in Technology and Teacher Education*, *3*, 275–287.

Hannafin, M. J., Hall, C., Land, S., Hill, J., Hannafin, M. J., Hall, C., … Hill, J. (2018). *Environments : Open-Ended Learning in Assumptions , Methods , and Implications*. *34*(8), 48–55.

Hegarty, M., Mayer, R. E., & Green, C. E. (1992). *Comprehension of Arithmetic Word Problems - Evidence From Students' Eye Fixations*. *84*(1), 76–84.

Hegarty, M., Mayer, R. E., & Monk, C. A. (1995). Comprehension of Arithmetic Word Problems: A Comparison of Successful and Unsuccessful Problem Solvers. *Journal of Educational Psychology*, *87*(1), 18–32. https://doi.org/10.1037/0022-0663.87.1.18

Hehman, E., Stolier, R. M., & Freeman, J. B. (2015). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes & Intergroup Relations*, *18*(3), 384–401. https://doi.org/10.1177/1368430214538325

Hershkovitz, A., & Nachmias, R. (2012). Log-based assessment of motivation in online learning. In *Handbook of Educational Data Mining* (pp. 287–298). https://doi.org/10.1201/b10274-23

Hessels, R. S., & Hooge, I. T. C. (2019). Eye tracking in developmental cognitive neuroscience – The good, the bad and the ugly. *Developmental Cognitive Neuroscience*, *40*. https://doi.org/10.1016/j.dcn.2019.100710

J H Boonen, A., & Jolles, J. (2015). Second Grade Elementary School Students? Differing Performance on Combine, Change and Compare Word Problems. *International Journal of School and Cognitive Psychology*, *2*(2), 2–7. https://doi.org/10.4172/2469-9837.1000122

Jacovina, M. E., Snow, E. L., Allen, L. K., Roscoe, R. D., Weston, J. L., Dai, J., & McNamara, D. S. (2015). How to Visualize Success : Presenting Complex Data in a Writing Strategy Tutor. *Proceedings of the 8th International Conference on Educational Data Mining*, 594–595.

Jänicke, S., Heine, C., & Scheuermann, G. (2013). GeoTemCo: Comparative Visualization of Geospatial-Temporal Data with Clutter Removal Based on Dynamic Delaunay Triangulations. *Communications in Computer and Information Science*, *359 CCIS*, 160–175. https://doi.org/10.1007/978-3-642-38241-3_11

Jugo, I., Kovačić, B., & Slavuj, V. (2015). Integrating a Web-based ITS with DM tools for Providing Learning Path Optimization and Visual Analytics. *Proceedings of the 8th International Conference on Educational Data Mining*, 574–575.

Kabugo, D., Muyinda, P. B., Masagazi, F. M., Mugagga, A. M., & Mulumba, M. B. (2016). Tracking students' eye-movements when reading learning objects on mobile phones: A discourse analysis of luganda language teacher- trainees' reflective observations. *Journal of Learning for Development*, *3*(1), 51–65.

Kaczorowski, T., & Raimondi, S. (2014). eWorkbooks for Mathematics: Mapping the Independent Learning Experiences of Elementary Students with Learning Disabilities. *Journal of Learning Analytics*, *1*(3), 179–182.

Karatekin, C. (2007). Eye tracking studies of normative and atypical development. *Developmental Review*, *27*(3), 283–348. https://doi.org/10.1016/j.dr.2007.06.006

Koedinger, K. R., Baker, R. S. J., Cunningham, K., & Skogsholm, A. (2010). Handbook of Educational Data Mining. In *Handbook of Educational Data Mining* (Vol. 20103384). https://doi.org/10.1006/jsvi.1993.1340

Koedinger, K. R., D'Mello, S., Mclaughlin, E. A., Pardos, Z. A., & Rosé, C. P. (2015). Data mining and education. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(4), 333–353. https://doi.org/10.1002/wcs.1350

Koedinger, K. R., McLaughlin, E. A., & Heffernan, N. T. (2010). A Quasi-Experimental Evaluation of An On-Line Formative Assessment and Tutoring System. *Journal of Educational Computing Research*, *43*(4), 489–510. https://doi.org/10.2190/EC.43.4.d

Lee, F. J., & Anderson, J. R. (2001). Does Learning a Complex Task Have to Be Complex?: A Study in Learning Decomposition. *Cognitive Psychology*, *42*(3), 267–316. https://doi.org/10.1006/cogp.2000.0747

Lee Swanson, H. (2011). Working Memory, Attention, and Mathematical Problem Solving: A Longitudinal Study of Elementary School Children. *Journal of Educational Psychology*, *103*(4), 821–837. https://doi.org/10.1037/a0025114

Mazza, R., & Dimitrova, V. (2004). Visualising student tracking data to support instructors in web-based distance education. *Proceedings of the 13th International World Wide Web Conference on Alternate Track Papers & Posters - WWW Alt. '04*, 154. https://doi.org/10.1145/1013367.1013393

Mc Alister, M., Dunn, J., & Quinn, L. (2005). Student teachers' attitudes to and use of computers to teach mathematics in the primary classroom. *Technology, Pedagogy and Education*, *14*(1), 77–105. https://doi.org/10.1080/14759390500200194

Moeller, K., Neuburger, S., Kaufmann, L., Landerl, K., & Nuerk, H. C. (2009). Basic number processing deficits in developmental dyscalculia: Evidence from eye tracking. *Cognitive Development*, *24*(4), 371–386. https://doi.org/10.1016/j.cogdev.2009.09.007

Moos, D. C., & Azevedo, R. (2008). Self-regulated learning with hypermedia: The role of prior domain knowledge. *Contemporary Educational Psychology*, *33*(2), 270–298. https://doi.org/10.1016/j.cedpsych.2007.03.001

Moutsios-Rentzos, A., & Stamatis, P. J. (2015). One-step "change" and "compare" word problems: Focusing on eye-movements. *Electronic Journal of Research in Educational Psychology*, *13*(3), 503–528. https://doi.org/10.14204/ejrep.37.14133

Murphy, R. F., Penuel, W. R., Means, B., Korbak, C., Whaley, A., Allen, J. E., & Allen. (2002). E-DESK: A Review of Recent Evidence On the Effectiveness of Discrete Educational Software Prepared. *SRI Project*, *11063*(April). https://doi.org/10.1080/01944360903490923

Navalpakkam, V., & Churchill, E. (2012). Mouse tracking: measuring and predicting users' experience of web-based content. *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems - CHI '12*, 2963. https://doi.org/10.1145/2207676.2208705

Niehorster, D. C., Cornelissen, T. H. W., Holmqvist, K., Hooge, I. T. C., & Hessels, R. S. (2018). What to expect from your remote eye-tracker when participants are unrestrained. *Behavior Research Methods*, *50*(1), 213–227. https://doi.org/10.3758/s13428-017-0863-0

Nigh, J., Pytash, K. E., Ferdig, R. E., & Merchant, W. (2015). Investigating the Potential of MOOCs in K-12 Teaching and Learning Environments. *Journal of Online Learning Research*, *1*(1), 85–106.

Nyström, M., Andersson, R., Holmqvist, K., & van de Weijer, J. (2013). The influence of calibration method and eye physiology on eyetracking data quality. *Behavior Research Methods*, *45*(1), 272–288. https://doi.org/10.3758/s13428-012-0247-4

Papamitsiou, Z., & Economides, A. A. (2014). Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence. *Educational Technology & Society*, *17*(4), 49–64.

Peña-Ayala, A. (2014, March 1). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems with Applications*, Vol. 41, pp. 1432–1462. https://doi.org/10.1016/j.eswa.2013.08.042

Pilli, O., & Aksu, M. (2013). The effects of computer-assisted instruction on the achievement, attitudes and retention of fourth grade mathematics students in North Cyprus. *Computers & Education*, *62*, 62–71. https://doi.org/10.1016/J.COMPEDU.2012.10.010

Polotskaia, E., Savard, A., & Freiman, V. (2015). Duality of mathematical thinking when making sense of simple word problems: Theoretical essay. *Eurasia Journal of Mathematics, Science and Technology Education*, *11*(2), 251–261. https://doi.org/10.12973/eurasia.2015.1325a

Poole, A., & Ball, L. J. (2005). Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects. *Encyclopedia of Human-Computer Interaction*, 211–219. https://doi.org/10.4018/978-1-59140-562-7

Prantner, C. K. (2016). The evaluation of the results of an eye tracking based usability tests of the so called Instructor's Portal framework (http://tanitlap.ektf.hu/csernaiz). *6th IEEE Conference on Cognitive Infocommunications, CogInfoCom 2015 - Proceedings*, 459–465. https://doi.org/10.1109/CogInfoCom.2015.7390637

Rayne RC & Baggott GK (2004) Computer-based andcomputer-assisted tests to assess procedural andconceptual knowledge. In: Proceedings of the 8th International Computer-Assisted Assessment Conference,2004, Loughborough University. M. Ashby & R. Wilson, eds.pp. 307-309.Rayner, K., & Pollatsek, A. (1989). The psychology of reading. *Choice Reviews Online*, *26*(11), 26-6565-26–6565. https://doi.org/10.5860/choice.26-6565

Renkewitz, F., & Jahn, G. (2012). Memory indexing: A novel method for tracing memory processes in complex cognitive tasks. *Journal of Experimental Psychology: Learning Memory and Cognition*, *38*(6), 1622–1639. https://doi.org/10.1037/a0028073

Robert P. Taylor, I. (1980). The Computer in School: Tutor, Tool, Tutee. Retrieved July 14, 2018, from https://eric.ed.gov/?id=ED207670

Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, *40*(6), 601–618. https://doi.org/10.1109/TSMCC.2010.2053532

Roschelle, J., Shechtman, N., Tatar, D., Hegedus, S., Hopkins, B., Empson, S., … Gallagher, L. P. (2010). Integration of Technology, Curriculum, and Professional Development for Advancing Middle School Mathematics: Three Large-Scale Studies. *American Educational Research Journal*, *47*(4), 833–878. https://doi.org/10.3102/0002831210367426

Ryan, A. W. (1991). Meta–Analysis of Achievement Effects of Microcomputer Applications in Elementary Schools. *Educational Administration Quarterly*, *27*(2), 161–184. https://doi.org/10.1177/0013161X91027002004

Schneider, B., & Pea, R. (2014). The Effect of Mutual Gaze Perception on Students ' Verbal Coordination. *Proceedings of the 7th International Conference on Educational Data Mining (EDM)*, (Edm), 138–144. https://doi.org/10.1016/j.jpowsour.2003.11.036

Schneider, M., Heine, A., Thaler, V., Torbeyns, J., De Smedt, B., Verschaffel, L., … Stern, E. (2008). A validation of eye movements as a measure of elementary school children's developing number sense. *Cognitive Development*, *23*(3), 409–422. https://doi.org/10.1016/J.COGDEV.2008.07.002

Schweppe, J., & Rummer, R. (2014). Attention, Working Memory, and Long-Term Memory in Multimedia Learning: An Integrated Perspective Based on Process Models of Working Memory. *Educational Psychology Review*, *26*(2), 285–306. https://doi.org/10.1007/s10648-013-9242-2

Wei, S., Lei, Q., Chen, Y., Xin, Y. P., & Kastberg, S. (2018). Evaluating the effects of highlighting text animations on the attention distribution of students with math learning difficulties. ASEE Annual Conference and Exposition, Conference Proceedings, 2018-June.

Shum, S. B., & Crick, R. D. (2012). Learning dispositions and transferable competencies: Pedagogy, modelling and learning analytics. ACM International Conference Proceeding Series, 92–101. https://doi.org/10.1145/2330601.2330629

Siemens, G., & Baker, R. S. J. d. (2012). Learning analytics and educational data mining. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK '12*, 252. https://doi.org/10.1145/2330601.2330661

Smith, A. D., Mestre, J. P., & Ross, B. H. (2010). Eye-gaze patterns as students study worked-out examples in mechanics. *Physical Review Special Topics - Physics Education Research, 6*(2), 020118. https://doi.org/10.1103/PhysRevSTPER.6.020118

Stultz, S. L. (2008). The effectiveness of computer-assisted instruction for teaching mathematics to students with specific learning disability. *Dissertation Abstracts International Section A: Humanities and Social Sciences*, *2*(5-A), 1738.

Susac, A., Bubic, A., Kaponja, J., Planinic, M., & Palmovic, M. (2014). EYE MOVEMENTS REVEAL STUDENTS' STRATEGIES IN SIMPLE EQUATION SOLVING. *International Journal of Science and Mathematics Education*, *12*(3), 555–577. https://doi.org/10.1007/s10763-014-9514-4

Tai, R. H., Loehr, J. F., & Brigham, F. J. (2006). An exploration of the use of eye-gaze tracking to study problem-solving on standardized science assessments. *International Journal of Research and Method in Education*, *29*(2), 185–208. https://doi.org/10.1080/17437270600891614

Taspinar, Z., & Bulut, M. (2012). Determining of problem solving strategies used by primary 8 , grade students' in mathematics class. *Procedia - Social and Behavioral Sciences*, *46*(1998), 3385–3389. https://doi.org/10.1016/j.sbspro.2012.06.071

Tobii Technology. (2011). Accuracy and precision Test method for remote eye trackers. *Technology*.

Tsai, M.-J., Hou, H.-T., Lai, M.-L., Liu, W.-Y., & Yang, F.-Y. (2012). Visual attention for solving multiple-choice science problem: An eye-tracking analysis. *Computers & Education*, *58*(1), 375–385. https://doi.org/10.1016/j.compedu.2011.07.012

Vieira, C., Parsons, P., & Byrd, V. (2018). Visual learning analytics of educational data: A systematic literature review and research agenda. *Computers & Education*, *122*, 119–135. https://doi.org/10.1016/j.compedu.2018.03.018

Why are math word problems SO difficult for elementary school children? (n.d.). https://doi.org/10.1046/j.1365-2567.2001.01164.x

Xin, Y. P. (2012). Conceptual model-based problem solving: Teach students with learning difficulties to solve math problems. Rotterdam, Netherlands: Sense. doi:10.1007/978- 94-6209-104-7_1

Xin, Y. P., Kastberg, S., & Chen, Y. (2015). Conceptual Model-based Problem Solving (COMPS): A Response to Intervention Program for Students with LDM. National Science Foundation (NSF) funded project.

Yoon, D., & Narayanan, N. H. (2004). Mental imagery in problem solving: An eye tracking study. *Eye Tracking Research and Applications Symposium (ETRA)*, 77–83.

Zachary Jacobson, J., & Dodwell, P. C. (1979). Saccadic eye movements during reading. *Brain and Language*, *8*(3), 303–314. https://doi.org/10.1016/0093-934X(79)90058-0

Zushi, M., Miyazaki, Y., & Norizuki, K. (2012). Web application for recording learners' mouse trajectories and retrieving their study logs for data analysis. *Knowledge Management and E-Learning*, *4*(1), 37–50. https://doi.org/10.1177/014662168000400201

APPENDIX A. INTERVIEW PROTOCOL

Introduction:

You have completed the diagnosis tasks and NASA_TLX survey. Next, I will ask a few questions about your diagnosis process, your perceptions of the task-loads of the analysis methods, and your preference for the two methods. It will take about twenty to thirty minutes. There are no correct or wrong answers. I would like you to feel comfortable with saying your judgments and perceptions. If you feel OK, I will tape-record our conversation so that I can transcribe our conversation later to get all the interview details. I assure the recording will remain confidential.

Interview questions:

Task load of analysis methods (NASA_TLX survey)

1. **According to your NASA_TLX survey, you choose the XX demand (e.g., mental demand) of video/visualization is very high/low. Would you talk about why you feel the XX demand is high?**

Probes: Would you give an example?
Which specific parts of the video/visualization require high XX demand do you think?

2. **According to your NASA_TLX surveys, the XX demand of using video (visualization) method is low/neural/high, while using visualization (video) method, the demand is high/neural/low. Would you explain why you feel that way?**

Probes: Do you think what differences between the two methods lead to the demand requirement differences between the two methods?
What analysis activities you did in the diagnosis processes make you feel the XX demand is different between the two methods?

Student problem-solving strategies and Difficulties

3. **Could you describe your diagnosis process of problem-solving task XX (task list: P1-B3.5, P4-B3.5, P4-B7.1, P5-B3.5, P5-B7.1, P8-B7.1)? Why did you determine the student used XX problem-solving strategies?**

Probe: What behaviors of the student make you determine the student applied XX strategy?

4. **What difficulties do you think the student had in task XX (task list: P1-B3.5, P4-B3.5, P4-B7.1, P5-B3.5, P5-B7.1, P8-B7.1)? How did you get the judgment on the student difficulties?**

The Judgement of student performance group

5. **Which performance group do you think the student belongs to? Would you explain how did you determine that the student belongs to the XX (high-performance/medium performance/low performance) group?**

Probe: Which data/chart you referenced to make the judgment?

Perceptions of the two analysis methods

6. **Do you think the training time of the visualization is acceptable?**

7. **Which method do you prefer? Video or visualization? Why?**

Probe: Would you describe in which condition you will use the video/visualization method?
Do you think what advantages/disadvantages the method has?

APPENDIX B. PREFERENCE INTERVIEW SUMMARY

Video pros	Video cons
The video presents more information compared with visualization, such as very detail interaction, hesitation (mouse trajectory, drag tags here and there), and so on.	We may do not have enough patience to see where the student looked at. If the student spent one hour on the problem, the teacher needs to spend one hour to look.
Video is easier for me to compare different things going on at the same time, like where they are looking, how much time they are spending, how many times they are dragging and clicking. All above information is in the visualization, but it's easier to watch the video and make the judgment.	I think even the video, I can stop and playback, but still, it's temporally, and there is a memory component of that. But here in the visualization, every data presents here. But I am still not very used to the visualization actually. After these tasks, I start to prefer the visualization but still a little bit not used to it.
I would prefer video if I have a lot of time.	

Visualization pros	Visualization cons
The visualization is just a picture, and I can control the time to spend on each image. So It could be a quick way for me to tell what happened and change my instruction/curriculum.	I will not use the visualization to evaluate my students because it's too speculative, open to interpretation. For example, trying to decide what method students used to solve the problem based on what they looked at and how much they clicked. For me, I am not sure that's a good way to do that for sure.

If we practice on the visualization, we will get much more progress than video because I can control it. The student is so slow in the video, although I can speed up, but it not very helpful in answering some questions.	
It probably needs more than an hour of training to become good at it. But today it's easier for me to do the visualization compared with yesterday. Probably if I keep doing with the visualization, I will do better. (training time of visualization is fine)	

Training Time Worthwhile
If my student number is very large, say more than five students, I would like to use the visualization. But if less than five students, I would like to know more about them. Training time is not a big issue.
Yeah, it's worthwhile. You saw today when we watch the videos, I can't make much sense until we watch the visualization. I like both of them. I would look at the video and then go to visualization. If you just let me choose one, I will choose the visualization. I think the video goes fast, and you can't tell what's popping out. But the visualization, you can see. But when you look at the visualization data, it's hard to understand. Like in the visualization, you can't see the student drag back and forth, back and forth… this information is missing … But you can see it in the video. That's why both of them are useful.
I think it depends. Because for videos, I can get a lot of information. Although visualization, I can also get a lot of info. But the video gives more, like how they drag, the real trajectory of mouse, their confusion.

But at the same time, the visualization can present information all together. It can also present some very important information. For example, there is no way for me to really see whether the student put a lot of attention to a specific word. But visualization can directly show me. Besides, I think a teacher may need to see tens or hundreds of a student problem-solving process to really understand the student in case if there is no way to talk to the student. In that case, the video may take a lot of time, which is frustrated. I would say I prefer visualization when I need to know the students' visual attention. But I will use video if I want to know the student's hesitation and confidence. If I must choose one, I will choose the video, even it’s complicated, as it provides more information.
As I said, I felt very difficult yesterday, but I feel much better today as I never used visualization before. But practice makes me much better today.
Yes, it’s worthwhile if it’s not up to one hour, because it gives you a lot of information that you can’t attain from the video. I do think the visualization is a little confusing and a little busy. But I definitely think it worthwhile. Which method do you prefer? : I could say both. More information is better. But also, it depends on time, if I have a lot of students, tasks I will use visualization. If I have time, I will look at the video. As the video gives more context, you can see what’s happening in time. Then you can go back to visualization to validate your findings in the video.

APPENDIX C. TASK LOAD INTERVIEW SUMMARY

Mental demand (Visualization)	**Menta demand (Video)**
I need to learn how to look at the graphics, how to interpret all the data on the screen.	Follow video; think task, replay video to validate
The training part is a little bit mentally demanded.	Follow student eye movement, how long looked at the specific part.
Combine and compare data. *More training will be better*.	Well, when I looked at the video, I see all the jumping, I couldn't figure out what that means, and it's hard to follow. I just, Oh, I don't know… I am looking at the right thing? And I am thinking, am I interpreting it correctly?
Decreased compared to R1 because of training.	Decreased comparing to R1 because of training.
It's hard, but not that hard. The interpreting, like what the information should I look at, am I understand the right meaning of that. Do I understand the data correctly? That's hard. I kept looking at all the instructions and questions.	I think the mental demand of visualization comes from understanding the graphics and combine them together. For the video, it's similar, come from combine all information presented at the same time in a short time.
I got headache, I don't want to make my head. I don't know… I thought I should put lower physical demand than last time.	I am more used to the vis this time after taking more tasks, especially the radar chart. For video, I would say no big change even slightly

	more demanding, because I am familiar with tasks and try to catch more information in the video than yesterday. I think a little bit more mental demand this time. Besides that, it's easier in some tasks, but it harder in some videos because some of the videos today there are very strange eye movement data in the video. I got better, but the video itself get harder.
(both methods) I have to make so many decisions, handle some much information. It's of course mental demand.	

Physical demand (Visualization)	**Physical demand (video)**
	Track eye movement; playback many times; I need to follow the dots every time to interpret performance, So I need to look at all the places on the screen, and I need to go back to the video to find what I want to look at… Like several times…
	Video stops and playback. Locate the replay start time point;
	Follow eye movement, eye tired.

	For physical demand, I played forth and back. But I think it’s fine.
	Because I track more information today, so I clicked more this time to make sure I am in the right spot. And then also compare to visualization, I don’t need to do any physical thing except look at the graph.

Performance (Visualization)	**Performance (Video)**
Vis can be interpreted quickly;	I need to follow my feeling when I look at the video. The reason I am going back and forth is I am not sure I am correct. So, I need to look at it again to decide my answer.
Vis: unconfident in the track/interpret of data.	Video is hard to track (eye movement);
In the radar chart, you give me more specific information such as the duration, how long they fixated at a specific unit, which makes me more confident with my performance. I can’t get direct and clear data or concepts about duration in the video.	I am still not sure whether my memory is correct as I speed up sometimes. For example, they switch labels or drag labels. I can’t remember how many times they switch labels as the video is long.

I believe in your data (in visualization). I just don’t know the way I am interpreting it is as expected… I am not confident.	Yes, because I keep checking the video. For one question, I checked about three times. So I am very sure I am correct. But to make sure correct, I spend a lot of energy. But I should say in the visualization. I don’t need recheck.
Because it’s hard to diagnose what the other people think about through what they do. I am not saying I make the wrong decision, but I am not that confident. The other reason is I just saw one or two tasks of one person, if I could see more tasks, it may be better. Say, if I saw nine tasks from the same person, I maybe will more confident that I make a good decision.	I only followed students’ eye movements, but for me, I still had a hard time really consider what the student thought about.
I feel unsure do these diagnoses correctly… But for some tasks, I think I make the right diagnose as some tasks are pretty easy to defines. I guess I just a little unsure. I think the unsure come from my inexperienced. I didn’t do this kind of analysis before.	I am not confident that I am interpreting the data correctly. So, I am not felt very successful.
I am more confident this time than yesterday. More practice makes it better. I will become more and more used for visualization.	

Temporal Demand (Visualization)	**Temporal Demand (Video)**

For both methods, I didn't feel rushed.	Remember task; speed up to save time; make sure not to lose information; (I don't want to spend too much time to read the video.)
I try to complete quickly. So we have time to talk, so I probably a little bit rushed this time.	If I speed up too fast, I can't catch the details. If I slow down too much, I don't think I have the patience. I want to quick and correct. So, I am rushed.
	I didn't feel rush when I do the tasks, but I feel I should do the tasks as quickly as possible.
	Because I need to follow their eye movement and their operation at the same time, even I slow down. I still need to see, ok, where they look at, where they put tags... But for vis, it's just there, so I just need to see one by one, then it's done.
	At first, it's rushed. I just rushed to get all the information. Then after the first task, I start to use play back and forth, and it make me less rushed.

Effort (Visualization)	**Effort (Video)**
Visualization can give me clear data about student performance. But for video, I only have an overview, no data backup. I need to understand the visualization, such as the definition of radar chart metrics. But after I	Trying to interpret these data or videos data into my understand and diagnoses.

understand these metrics, they are very helpful for me to complete the diagnosis tasks.	

Frustration (Visualization)	**Frustration (Video)**
Not too bad. Be encouraged by different data sources. If I got similar results from performance data, radar charts, and fixations. They kind of encourage me.	Stressed, refresh memory many times
	The human sense is not very reliable. I could be less frustrated if I got clear data, just like what I got in the visualization.
(Both methods) Just the insecure aspect, I am not sure whether my answer is what you need. Sometimes I am wondering the reason I think the student is correct is the same as your understanding. We see the same evidence, but I am not sure whether we can get the same result. That's my insecure comes from. But except that, I feel no pressure or stress.	For the video, I feel very insure. for the video, it's moving, I know I can stop, playback and forth, it's still … but for the vis, I can look clear, compare. I can look at multiple things and compare them. But for the video, I have to keep it in my mind… sometimes I may miss or remember something wrong. I mean it's may not very reliable.
Decrease, more familiar with tasks, vis.	I feel a little insecure as I have no experience. I would say a little bit stressed but insecure is the most. Stress also comes from insecurity.

APPENDIX D. INTERVIEW TRANSCRIPTIONS - EXAMPLE

Interview Transcription 1 – Evaluator 1 (Re-researcher, E1-Evaluator 1)

Task Load Interview:

Re: According to your NASA TLX survey, the mental demand of using visualization method is very high. Would you talk about why you feel the mental demand is high?

E1: To interpret the visualization, I need to know a lot of information before I do the task. So I need to learn how to look at the graphics, how to interpret all the data on the screen. So mental demand is very high… I guess.

Re: Do you feel tired when you do the tasks?

E1: No, when I do the task, It's fine. But before I do the tasks...

Re: The training part…

E1: Yes, the training part is a little bit mentally demanded.

Re: For video, in the physical demand, your rating is very high. Could you explain it?

E1: When I look at the video, I need to follow the dots every time to interpret performance, So I need to look at all the places on the screen, and I need to go back to the video to find what I want to look at.. Like several times…

Re: So, this sounds is frustrated?

E1: Yeah, I went back five times to look at it again. It repeated.

Re: For the performance (How successful were you in accomplishing what you were asked to do?), you choose neural in video, and choose almost perfect in visualization. Could you help to explain why?

E1: for the visualization, I can be sure… I can believe I can interpret it quickly, but for the video, I feel like I may incorrect.

Re: You feel you may like incorrect?

E1: Yeah, for video, I look at dots, try to follow dots, but there is no ... research that I can see?

Re: No research…?

E1: um… I need to follow my feeling when I look at the video. The reason I am going back and forth is I am not sure I am correct. So, I need to look at it again to decide my answer.

Re: So, for video, that' s the reason why you think you probably make something wrong?

E1: Yeah.

Diagnosis Process Interview:

Student9-B3.5-video

Re: Could you describe our diagnosis process of P9-B3.5? Why did you determine the student usedmodel-based problem-solving strategies, and the student may have difficulty in the diagram equation?

E1: The students' fixations on diagram equations are very long (duration) like here, and here, And the student went back to look at it again and again, so I think the student used model-based problem-solving strategy.

Re: So, you look at the diagram equation and think he try to understand the diagram equation

E1: Yeah, from the fixations here (point at the fixations located at question area), the duration and regression here, so he tried to understand the problem first, and then he tried to understand the question again. He tried to understand both of them, the problem and equation, so I think it's model-based.

Re: OK.

Re: So, for the difficulty question, you choose the student has difficulty in understanding the diagram equation.

E1: Yeah, the student looked at the diagram for a long time and repeated.

Re: Did you look at the fixations located in the question area?

E1: Yeah, but it's a little difficult to decide. The fixations (on the question area) are long, but compare to the diagram equation, the fixations on diagram equations are relative longer, so I decided the student may have difficulty in the diagram equation.

Re: But the student is correct at last, the student looked at the equation a lot, right?

E1: Yeah. For the answer, he may not have the problem, but for the understanding of the equation and problem, he may have difficulty.

Student 8-B7.1-Video

Re: Please tell me how you used video to understand the student problem-solving process in P8-B7.1 and describe any difficulties/struggles you think the student meet.

E1: I try to track all the red dots of the student. Then from the student eye movement, I can see where the student looked at whether the student has the problem on the task. If the student eye movement focuses on keywords, I can see they focus on keywords. If the answer is incorrect, I can see the student have … have difficulty.

Re: Now, we go through the video. Then you tell me what you think.

E1: She looks at numbers a lot, so I think she uses keyword strategy. Yeah... Focus on keywords and numbers… The student didn't look at the task a lot. Before drag name tag, she just looked at keywords

Re: Any difficulty or problem of the student you can see?

E1: The student has problems. Because here, she moves number tags a lot. So, the student has some difficulty (in the question area).

Re: Then, do you have any assumptions about the student's difficulty?

E1: ……. I don't know exactly… but….

Re: So anything you can see...?

E1: It's hard to decide from the video. ……..

Student1-B3.5-Visualization

Re: For P1-B3.5, can you describe your diagnosis process?

E1: Here, the reason why I choose paid little attention….

Re: So, the way you answer the question is looking at the question, then find the answer?

E1: yeah.

Re: Could you describe your understanding process?

E1: The student didn't look through the task. He just tried to solve the problem. The name tag is dragged all over the diagram equation (changed several times), he first put gave here, then put gave here, then here and here, so changed his mind several times. It means he doesn't understand the problem at all. He just guessed the answer at first. So, he put total here, and left here, and move here. Then finally, incorrect. He changed several times and incorrect at last. He may get some hint from feedback. Then in the second try, he put total in the whole box. And then gave, left. Finally correct. So, I think he used guess and check.

Re: So, there are two cases of guess & Check. One is trying to understand the problem but has difficulty and failed then use guess and check. Another case is not paying attention, just guess and check to submit the answer. Which case do you think this student is?

E1: I think the student has difficulty because here, the duration of other behavior is really long. So he is trying to think, try to solve the problem. He wants to be correct, he is struggling.

Re: Then how about this (the second question of the task)?

E1: The student didn't pay much attention to understand the problem. He just dragged the number tags then submit. He looked at keywords and numbers. He put 62 to total, then 87… In the second try, the student looked at total for a long time compare with others. So, he figured out it is total. So, he put total (the number of total) here. Then, he has left a, so he put a here.

Re: What strategy do you think the student use?

E1: Guess and check. Because you can see the student didn't put a lot of time on the task (radar chart, duration). He didn't put a lot effort on his first try. Very short operation time. In his second try, he put more effort on here (numbers) to understand the problem.

Re: Yeah, it seems like the second try, the student at least put attention on keywords. Do you think any difficulties/struggles the student may have?

E1: I think the student may have some difficulties but not a lot. Because I can see the student didn't have too many regressions (radar chart)

Re: So, you think at first time, student use guess & check. Incorrect. Then look at keywords to solve the problem.

E1: Yeah, some difficulties I can see. Some regressions on diagram equation.

Student Performance Group Interview

Re: Which group you think the student belong to?

E1: When I decide I always look at drag efficiency, and the other regressions. But here I don't think the student use an efficient way to solve the problem. Because durations is long but other metrics such as regression and drag efficiency is not very high.

E1: For this student, the first try is always incorrect, the second try is always correct.

Re: Any comments on the student

E1: For me the reason why the student always fail in the first try. He didn't look at task, he didn't try to understand the problem. In his first, he just put tags everywhere. Then, the second chance, he tried to solve the problem correctly. So, if he can put more effort in the first try, I think he can solve the problem correctly.

Re: OK. Thanks

Interview Transcription 2 – Evaluator2 (Re-researcher, E2-Evaluator 2)

Diagnosis Process Interview

Student5-B7.1-Visualization

Re: Could you describe your diagnosis process of student 5's B7.1.

E2: The student read the whole problem, not only focus on keywords. they correctly answer the questions. So, I think the student use model-based problem-solving strategies.But in the second step, the student tried drag different labels again and again, which means the student doesn't know which tag should be put in which box. So, he changed his answer multiple times. So, I think the student use guess and check strategy.

Re: So, what elements you use to decide the student use guess and check.

E2: Because the student put the tag everywhere. He just confused.

Re: So, what difficulty the student has? How do you know that?

E2: I think the student both have difficulty in understanding task content, and diagram equation. In the second step, the student didn't read the problem, only focused on the words need to be dragged. The student first tried to use fewer to put the number in "smaller" box. But he is not correct, then he put fewer in "difference", everywhere. So, I think he doesn't know both task content and equation.

Student8-B7.1-Vido

Re: Could you describe your diagnosis process of Student8-B7.1 (video).

E2: I can see the student read most of the problem, not only focus on keywords. He dragged only once and get the correct answer. So, I think the student used model-based problem solving since the student changed a lot of tags and focus only on keywords in the second step.

Re: You choose model-based problem-solving strategy for the first step and guess and check for the second step.

E2: Yeah, because in the second step, the student didn't use model-based problem-solving strategy. The student switches their answer (number tags) to check, which is correct, which is incorrect.

Re: You also said the student use keyword strategy.

E2: Yeah, the student used fewer, 26, only focus on keyword. The student didn't confident in his answer, so he just changed and checked his answer.

Re: What difficulty do you think the student had?

E2: About difficulty, I feel like … the student had no difficulty in the label but had difficulty in keywords and the equation. The student may think fewer is smaller…. Oh… the student failed twice, which means the student didn't understand the problem. If the student corrected in the second try, he might have some sense of the problem. If a student wronged twice, either because the student didn't understand the problem, or they did not pay attention.
Re: You think the student is a low-performance student.Would you explain your determine process?
E2: The student was correct in the first step and wronged twice in the second try. The first step is easy, but the second step is hard. The student failed twice in the second step, so I think the student is low performance. I don't know… we may see the visualization. (She clicked the participant's name in the visualization to see radar chart). Although the student has high drag efficiency. But except that, the other metrics the student is even lower than the average of low performance group.

Task Load Interview
Re: Would you talk about why you feel the mental demand of video is so high?
E2: Because I need to 'think' to follow the video. During the video playing, I need to think about each question and follow the video. If I couldn't follow it, I need to reread the video. So, I have to be careful about each step. For the physical demand, I need to stop and playback, which is the physical demand of video. I also need to drag back on the different parts to where I am not sure. For temporal demand, it's so rushed. I need to remember task. I need to speed up as I don't want to spend too much time reading the video. When the video playing, I need to make sure I didn't lose information at the same time. I want to finish it as soon as possible and make sure it's correct. So, it's really rushed for me, although I don't have a time limit.
How successful… I am not sure. Although I double checked the video, I am still not sure whether my memory is correct as I speed up sometimes. For example, when they switch labels or drag labels, I can't remember how many times they switch labels as the video is long.
How hard… Yes, it's hard based on all the above demands, I work very hard.
How struggle/stressed…. I feel I need to refresh my memory many times, so it's compressed.

Re: It seems you feel the mental demand of visualization is higher than the other demand, could you explain why you feel that way?

E2: Yeah, I need to think. To combine, compare the student data. But with more training, I think the mental demand will lower. No rush, no physical demand. I am pretty sure my answer as I know what I am doing, and I am clear with my data, I have visual, data presented there for me, so I am confident. How hard…. That's OK. Not that hard as I have enough practice, it will not be too hard. Frustration? That's OK. Because if I got similar results from performance data, radar charts, and fixations. They kind of encourage me.

For the visualization, training time is worthwhile or not?
E2: I think it depends. If my students work with computer programs and If my participants' number is very large, say more than five students, I would like to use the visualization. But if less than five students, I would like to know more about them. But it's ok to use it in the classroom as in the classroom there are a lot of students. Teachers don't have enough time for each student. But if teachers at visualization, they can spend equal time with each student.
Re: What's the difference between video and visualization? What makes you decide to choose video if you have more than five students?
E2: For problem-solving difficulty, the visualization is better than video. But for different purposes, for example, I want to learn more about students the video is more helpful.
Re: So, you think the student number is important for your decision of choosing which method to use?
E2. Yeah, If I have a large number of students, I couldn't have enough time to check their videos. Especially if I need to teach them tomorrow, then I don't have time to check their videos, but as the visualization is just a picture and I can control the time spend on each image. So It could be a quick way for me to tell what happened and change my instruction/curriculum tomorrow. So, like every day, I can follow students' progress.
Re: So, what's the difference between video and visualization that make you choose video if you have enough time?
E2: The video still presents more information compared with visualization, such as very detailed interaction, hesitation, and so on.

Diagnosis Process Interview
Student5-B3.5-Video

E2: I speed up the video to 5. I can see the student focus on the keywords. The student visual attention is on keywords and pays little attention (to other parts). Yeah, it's difficult to use video to judge it. The student drag once on the second step and the student focuses on keywords. Problem-solving strategy, I choose guess and check because both tries of the student are incorrect.

Re: So, you just looked at the student performance?

E2: I just looked at drag and performance. So you can see in the second step, the student incorrectly in the first try and then immediately switches the number tags quickly without think. So I think he uses guess and check.

E2: For the difficulty, I think the student has difficulty in understanding equation, task content, and paid little attention to the task content. It seems the student doesn't know the meaning of a. He put a as part in the first try and move it to total at the second step. He doesn't know the meaning of a.

Re: So, you said the student paid little attention to the question, could you explain why you say that?

E2: After the first try, the student started the second try without thinking.

Re: The student eye movement is on the task content area…

E2: Yeah, it seems the student is reading... that's the problem of video, we may not have enough patience to see where the student looked at. If the student spent one hour to think about the problem, the teacher needs to spend one hour to look. Besides, the student watched the number 78,64 without reading the whole problem.

Re: You think the student is low performance, based on what?

E2: Based on both tries, not correct, not understand the problem, not understand the equation. So I think he is a low-performance student.

Student1-B3.5-Visualization

Re: Could you describe your diagnosis process of P1-B3.5 (visualization)

E2: OK, I think this student looked at keywords. The second step is same. Problem-solving strategy for the first step is keyword strategy. The second step is "guess and check" …uh … I don't know… it seems hedoesn't understand the equation because he switched part and whole to get the correct answer in the second step. He may not understand task content, as well. He doesn't know the meaning of total and equation. In the first try of the first step, the student just put name tags everywhere. He is not sure. He may also use guess and check.

Re: You also said the student use keyword strategy.

E2: Yeah, only keyword can't solve the problem. The student need to read the problem and understand it. If you just know left and gave, how do you know which is part which is whole. So, the student needs to understand the problem. He believes in key words too much. He just read keywords, he doesn't care what the question about. then when he was struggling with the equation, instead of reading the whole problem, he just switched the tags. He thought he doesn't need to read the problem he can solve it or he just doesn't want to read. We don't know.

Re: You think the student belongs to low performance group?

E2: Yes, I checked the radar chart. See these metrics smaller than mid-performance average, so I think he belongs to the low-performance group.

Task Load Interview

Re: According to your NASA TLX survey, you choose the temporal demand of video is very high. Would you talk about why you feel the temporal demand is high?

E2: Yeah, if I speed up too fast, I can't catch details. If I slow down too much, I don't think I have the patience. I want to quick and correct. So, I am rushed.

Re: But you are satisfied with your performance (according to the survey).

E2: Yes, because I keep checking video. For one question, I checked about three times. So I am very sure I am correct. But to make sure correct, I spend a lot of energy.

Re: Did you check several times when using visualization?

E2: I checked the student visualization and the tasks radar chart, it's the only thing I switch. Except that there is no other check/switch.

Re: It seems the task load of visualization in the second round is decreased.

E2: some of them such as mental load is decreased. Because of training.

Re: Video is much better in the second round.

E2: Yeah, it's also because of practice. Also, I am more familiar with the tasks in the second round.

Re: If you have more time practice, do you think you will improve even more?

E2: Yeah, maybe, I think so. Practice makes perfect. But I have to say, if we practice visualization, we will get much more progress than video.

Re: Why?

E2: I will be more and more fast. I can control it. The student is so slow in the video, although I can speed up, but it not very helpful in answering some questions.