2018

MS. AN (Meeting Students’ Academic Needs): A Socially Adaptive Robot Tutor For Student Engagement In Math Education

Karina Liles
University of South Carolina

Follow this and additional works at: https://scholarcommons.sc.edu/etd
Part of the Computer Engineering Commons

Recommended Citation

This Open Access Dissertation is brought to you by Scholar Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact dillarda@mailbox.sc.edu.
MS. AN (MEETING STUDENTS’ ACADEMIC NEEDS): A SOCIAL CALLY ADAPTIVE ROBOT TUTOR FOR STUDENT ENGAGEMENT IN MATH EDUCATION

by

Karina Liles

Bachelor of Science
Spelman College, 2007

Master of Education
University of South Carolina, 2012

Master of Engineering
University of South Carolina, 2015

Submitted in Partial Fulfillment of the Requirements
For the Degree of Doctor of Philosophy in
Computer Science and Engineering
College of Engineering and Computing
University of South Carolina
2018

Accepted by:

Jenay Beer, Major Professor
Jason O’Kane, Committee Member
George Roy, Committee Member
Matt Thatcher, Committee Member
Marco Valtorta, Committee Member

Cheryl L. Addy, Vice Provost and Dean of the Graduate School
DEDICATION

For my family, Dad (John), Mom (Jennifer), Jalaya, Jaye, Kaylan, Jaylon, and Jayden Liles.
ACKNOWLEDGEMENTS

This research has been a labor of love with hopes that it can one day contribute to the amazing world of education and innovation. A special thank you to the National Consortium for Graduate Degrees for Minorities in Engineering and Science, Inc. (GEM), Pacific Northwest National Lab (PNNL), and Support to Promote Advancement of Research and Creativity (SPARC) for funding support throughout this journey.

I would like to acknowledge and thank my advisor, Jenay Beer, my review committee, lab mates, undergraduates, and special friend, Rachel Stuck for the tedious support provided throughout my research process. A special thank you to Dr. Juan Gilbert for connecting me to the African American community and providing me with unyielding support and guidance and Claflin University for allowing me to work with a new generation of young leaders.

A special thank you to Blenheim Elementary Middle School for their generosity and graciousness in allowing me the space and the students to conduct my research. The 5th grade teachers who worked with me to pull students and the wonderful 5th grade students who opened themselves up to what Ms. An had to offer. I am indebted to my mother Jennifer Liles, an educator at Blenheim Elementary Middle School for making my data collection a seamless process and to Sherolyn Mumford at Blenheim Elementary Middle School for being a steward of my onsite data collection process.

Thank you to my best girlfriends and Spelman sisters Jovi Porter, Gayla Robbins-Mair, and Chanceity Robinson for your relentless support, friendship, and
encouragement. Also, thanks to Jovi for your meticulous review and editing of my final paper.

A final and heartfelt thank you to my parents, John “Johnny” and Jennifer Liles for always making me feel like I can do anything that I put my heart and mind to and for setting me up for success; and my siblings Jalaya Liles Dunn, John “Jaye” Liles, and Kaylan Liles for being my circle and always standing with me. I am thankful for this journey. Ashe
This research presents a new, socially adaptive robot tutor, Ms. An (Meeting Students’ Academic Needs). The goal of this research was to use a decision tree model to develop a socially adaptive robot tutor that predicted and responded to student emotion and performance to actively engage students in mathematics education. The novelty of this multi-disciplinary project is the combination of the fields of HRI, AI, and education. In this study we 1) implemented a decision tree model to classify student emotion and performance for use in adaptive robot tutoring—an approach not applied to educational robotics; 2) presented an intuitive interface for seamless robot operation by novice users; and 3) applied direct human teaching methods (guided practice and progress monitoring) for a robot tutor to engage students in mathematics education.

Twenty 4th and 5th grade students in rural South Carolina participated in a between subjects study with two conditions: A) with a non-adaptive robot (control group); and B) with a socially adaptive robot (adaptive group). Students engaged in two one-on-one tutoring sessions to practice multiplication per the South Carolina 4th and 5th grade mathematics state standards.

Although our decision tree models were not very predictive, the results gave answers to our current questions and clarity for future directions. Our adaptive strategies to engage students academically were effective. Further, all students enjoyed working with the robot and we did not see a difference in emotional engagement across the two groups.

This study offered insight for developing a socially adaptive robot tutor to engage
students academically and emotionally while practicing multiplication. Results from this study will inform the human-robot interaction (HRI) and artificial intelligence (AI) communities on best practices and techniques within the scope of this work.
# TABLE OF CONTENTS

DEDICATION ........................................................................................................ iii

ACKNOWLEDGEMENTS .................................................................................. iv

ABSTRACT ......................................................................................................... vi

LIST OF TABLES ............................................................................................... x

LIST OF FIGURES ............................................................................................ xi

LIST OF ABBREVIATIONS .............................................................................. xiv

CHAPTER 1: INTRODUCTION ......................................................................... 1

1.0 EDUCATION ............................................................................................... 1

1.1 HUMAN-ROBOT INTERACTION ............................................................... 7

1.2 ARTIFICIAL INTELLIGENCE ................................................................. 8

1.3 SUMMARY ............................................................................................... 10

CHAPTER 2 IMPETUS OF RESEARCH .......................................................... 12

2.0 PROBLEM ................................................................................................. 12

2.1 PAST STUDIES ....................................................................................... 12

2.2 RESEARCH GOAL .................................................................................. 13

CHAPTER 3 METHODOLOGY ........................................................................ 15

3.0 PLATFORM .............................................................................................. 15

3.1 LESSON ................................................................................................. 16

3.2 DEVELOPMENT ...................................................................................... 20
LIST OF TABLES

Table 3.1 Multiplication problem types ................................................................. 16
Table 3.2 Example robot responses ........................................................................ 20
Table 3.3 Decision tree features for robot adaptability (emotion) ............................ 24
Table 3.4 Decision tree features for robot adaptability (performance) .................... 24
Table 3.5 Adaptive robot versus non-adaptive robot .............................................. 28
Table 4.1 Confusion matrix for each emotion ....................................................... 38
Table 4.2 Confusion matrix for emotion by valence .............................................. 38
Table 4.3 Confusion Matrix for performance .......................................................... 39
Table 4.4 Confusion Matrix for performance – data from current study ............... 40
Table 4.5 Average delay in answer (seconds) ......................................................... 47
Table 4.6 RPI by item for control ........................................................................... 50
Table 4.7 RPI by item for adaptive ........................................................................ 51
Table 4.8 Statistically significant RPI items by condition and agent factors .......... 51
Table 4.9 Key findings from interview (HRI) .......................................................... 53
Table 4.10 Key findings from interview (Education) .............................................. 54
Table 4.11 Key findings from interview (Artificial Intelligence) ............................... 55
Table 5.1 Key findings and design recommendations ............................................. 62
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>5th grade mathematics test results</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>Percentage of rural and urban students by grouped score classifications</td>
<td>3</td>
</tr>
<tr>
<td>1.3</td>
<td>Characteristics of a decision tree</td>
<td>8</td>
</tr>
<tr>
<td>1.4</td>
<td>Agent system (tutor)</td>
<td>9</td>
</tr>
<tr>
<td>1.5</td>
<td>Sample decision tree</td>
<td>11</td>
</tr>
<tr>
<td>3.1</td>
<td>NAO humanoid robot</td>
<td>15</td>
</tr>
<tr>
<td>3.2</td>
<td>Example session question</td>
<td>17</td>
</tr>
<tr>
<td>3.3</td>
<td>Sample problem with common misconceptions</td>
<td>18</td>
</tr>
<tr>
<td>3.4</td>
<td>Progress monitoring interface</td>
<td>19</td>
</tr>
<tr>
<td>3.5</td>
<td>Guided practice example</td>
<td>20</td>
</tr>
<tr>
<td>3.6</td>
<td>GEW emotion classifications</td>
<td>21</td>
</tr>
<tr>
<td>3.7</td>
<td>Emotion state diagram</td>
<td>22</td>
</tr>
<tr>
<td>3.8</td>
<td>Performance state diagram</td>
<td>23</td>
</tr>
<tr>
<td>3.9</td>
<td>Emotion decision tree classified by individual emotions</td>
<td>25</td>
</tr>
<tr>
<td>3.10</td>
<td>Emotion rules classified by individual emotions</td>
<td>25</td>
</tr>
<tr>
<td>3.11</td>
<td>Performance decision tree</td>
<td>25</td>
</tr>
<tr>
<td>3.12</td>
<td>Performance rules</td>
<td>25</td>
</tr>
<tr>
<td>3.13</td>
<td>Emotion decision tree classified by valence</td>
<td>26</td>
</tr>
<tr>
<td>3.14</td>
<td>Emotion rules classified by valence</td>
<td>26</td>
</tr>
<tr>
<td>3.15</td>
<td>Interface, start page</td>
<td>27</td>
</tr>
</tbody>
</table>
Figure 3.16 Choregraphe interface .................................................................27

Figure 3.17 Python programming ......................................................................28

Figure 3.18 Ms. An predicts that the student is sad and that the student answer the upcoming question incorrectly (adaptive) .................................................................29

Figure 3.19 Ms. An predicts that the student is happy and that the student will answer the upcoming question correctly (adaptive) .................................................................29

Figure 3.20 The student is surprised and answers the upcoming question incorrectly (non-adaptive) ........................................................................................................30

Figure 3.21 The student is happy and answers the upcoming question correctly (non-adaptive) ..................................................................................................................30

Figure 3.22 Student-robot communication diagram ........................................31

Figure 3.23 Technology experience profile ....................................................32

Figure 3.24 Agent characteristics and relationship to agent factors ...............34

Figure 3.25 Study procedure ............................................................................36

Figure 4.1 Formula for normalized learning gain (pre-/post-test) ..................41

Figure 4.2 Pre-/post-test 1 average learning gains ........................................42

Figure 4.3 Pre-/post-test 1 scores .....................................................................43

Figure 4.4 Pre-/post-test 2 average learning gains ........................................44

Figure 4.5 Pre-/post-test 2 scores .....................................................................44

Figure 4.6 Session 1 to session 2 average learning gains ...............................45

Figure 4.7 Pre-/post-test 2 percent correct by answer type ............................46

Figure 4.8 Overall RPI rating ...........................................................................48

Figure 4.9 RPI by agent factor .........................................................................49

Figure 4.10 RPI by agent characteristics ..........................................................49

Figure 5.1 Images of students who self-reported happy emotion ..................57

Figure 5.2 Examples of student work using wrong answer type ....................59
Figure 5.3 Student work and errors made (No error).................................................................64

Figure 5.4 Student work and errors made (Calculation Error: Partial products addition error) .................................................................64

Figure 5.5 Student work and errors made (Procedure error: Multiplying columns) ........65
LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI</td>
<td>Artificial Intelligence</td>
</tr>
<tr>
<td>API</td>
<td>Agent Persona Inventory</td>
</tr>
<tr>
<td>ELA</td>
<td>English and Language Arts</td>
</tr>
<tr>
<td>GEW</td>
<td>Geneva Emotion Wheel</td>
</tr>
<tr>
<td>HCI</td>
<td>Human-Computer Interaction</td>
</tr>
<tr>
<td>HRI</td>
<td>Human-Robot Interaction</td>
</tr>
<tr>
<td>PomDPs</td>
<td>Partially Observable Markov Decision Processes</td>
</tr>
<tr>
<td>RPI</td>
<td>Robot Persona Inventory</td>
</tr>
<tr>
<td>SDK</td>
<td>Software Development Kit</td>
</tr>
<tr>
<td>ZPD</td>
<td>Zone of Proximal Development</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

“The robot would be unable to understand children because it would lack reasoning skills that require cognitive, social, and emotional intelligences. Teaching requires that teachers understand what students already know and use that to help them make new connections. It also requires a relationship between student and teacher, one that allows and encourages risk-taking. A teacher's job is to balance a push for new knowledge and a stay for students to gain mastery. This takes a lot of intuition and personal judgment.”

- Educator

The above quote was taken from a survey we conducted on educators’ opinions on a robot teaching assistant and mirrors the thoughts of many of the respondents [1]. To address this concern, it is crucial to (1) understand the state of mathematics education in the US, particularly for at risk youth; (2) assess the potential of Human-Robot Interaction (HRI) for education; and (3) consider the Artificial Intelligence (AI) needed to develop a socially adaptive robot tutor. This three-pronged approach (education, HRI, and AI, respectively) lays the foundation of this dissertation.

1.0 EDUCATION

Mathematics is among the core academic subjects identified by the US Department of Education [2]. Math competence in early education leads to career and college readiness as it prepares students for undergraduate courses in college [3] and plays a critical part in
the competency for workers in the technical workforce and the nation’s economic development [3].

Although math proficiency is extremely important, many students are not excelling in the field [4]. The ACT Aspire test is administered in South Carolina to 3rd through 8th grade students. This statewide assessment tests students on grade level standards. Student scores are categorized into four levels based on readiness benchmarks: in need [of support], close, ready, and exceeding. Students who score below the cut off score are classified as “in need”. Those who score at or above the low cut off score and below the benchmark are classified as “close”. Those who score below the high cut off score and at or above the benchmark are classified as “ready”. The students who scored at or above the high cut off score are classified as “exceeding”.

Based on 2015 ACT Aspire scores for 5th grade students in South Carolina [4], 51.8% (N=54,601, M=418.2, SD=5.4) were classified as either in need or close. Thus, more than 50% of South Carolina 5th grade students were struggling with math. 4th grade test scores were similarly substandard. Figure 1.1 shows the 5th grade test percentages for each category.

![Figure 1.1 5th Grade mathematics test results](image-url)
Sadly, these results were worse for rural schools. 8.71% more rural students were classified as “in need” or “close” than urban students. In contrast, 10.85% more urban students were classified as “ready” or “exceeding” than rural students. Figure 1.2 shows the percentage of both rural and urban students in “in need” and “close” categories combined and “ready” and “exceeding” categories combined.

Figure 1.2 Percentage of rural and urban students by grouped score classifications

*Tutoring.* Tutoring is one approach to help students perform better in mathematics as it is often used to assist students who may show weaknesses in academic areas. Tutoring is a supplemental aid in the learning process that can further enhance a student’s academic ability [5]. Benjamin Bloom found that students who receive one-on-one tutoring outperformed students who receive traditional classroom instruction by two standard deviations (two-sigma problem) [6].

A tutoring interaction is comprised of an academic component and social component [7], [8]. Academically, tutors provide immediate and specific feedback. Socially, tutors provide positive reinforcement and guidance [8]. Together, these components are critical
for success in tutoring [5]. Further, this academic and social interaction fosters student engagement [5].

*Engagement.* In education, student engagement influences student motivation and progress in learning. The term student engagement encompasses the student’s attention, curiosity, interest, optimism, and passion when learning. There are many facets of engagement as it relates to education including intellectual engagement and emotional engagement [9].

*Intellectual engagement.* Intellectual engagement focuses on a student’s cognitive state during learning [9]. Teaching strategies are often employed for the maximum benefit of intellectual engagement. Two effective techniques that are encouraged are guided practice and progress monitoring [9].

Guided practice, also called scaffolding, is a process that allows a student to solve a problem with assistance from an expert [tutor] that the student would not be able to solve independently [10]. During instruction, the tutor provides support to the student as the student works to master a skill [11]. This strategy allows the student to master new skills in small increments (zones) by building on previous knowledge and with the help provided by the tutor. With this, a student is able to work within his/her zone of proximal development (ZPD) [12] and shift from watching the tutor model a skill to being able to perform the skill independently [8], [13].

While guided practice teaches a student a skill (or set of skills), progress monitoring is a complementary approach that allows the student to practice newly learned skills. Progress monitoring is a technique that can be used to assess student performance, predict future outcomes, and allow instructors to develop effective instruction [14]. As the name
suggests, progress monitoring is the act in which the tutor monitors a student’s progress. During this process, the tutor asks the students questions on the subject matter. This process allows students to demonstrate their understanding [8], [15].

Emotional engagement. Emotional engagement describes a student’s affective state during learning. Student emotions impact cognition and positive emotions stimulate attention [16]. It is important to organize emotions in a way that makes emotion groupings meaningful. Scherer et al. labels emotions by valence and control/power [17].

Valence refers to how the student feels. High valence (positive) refers to a pleasant and enjoyable experience that is likely to have positive and desired outcomes for the student. Conversely, low valence (negative) suggests an unpleasant and joyless experience that is likely to have negative and undesired outcomes [18].

Control/power refers to the student’s perceived ability to influence a situation. High power refers to the strong belief that a student can change a situation whereas low power refers to the student’s belief that they cannot change a situation [18].

Tutoring technology. Educational technology has grown significantly in recent decades [20] and has been used in many mathematics learning environments. While technology does not fix issues in mathematics education alone, (human teachers are still needed) its use has made significant contributions to learning, including social and emotional development [20].

Tutoring technology is helpful because it can provide individualized support, immediate reinforcement, unbiased feedback, and self-paced instruction. It is a powerful tool that can be used for management, communication, evaluation, motivation, and cognition [20].
**Onscreen agents.** Onscreen agents (also known as avatars or virtual humans) have become increasingly popular for tutoring activities [21], [22]. One reason for this is that they can employ all of the advantages of standard tutoring technology with the added benefit of exhibiting social behaviors [21]. An onscreen agent is a simulation of an animated object (usually a human) that displays many realistic traits for interaction such as facial expressions, emotions, personality, and communication [23], [24].

**Robot tutors.** Many educational options exist from standard tutoring technology to onscreen agents; however, robots differ because the physical embodiment of the robot adds an additional degree of sociability, which results in higher performance for students [25], [26]. Even in applications outside of tutoring, the robot was more favorable than an onscreen agent. For example, robots were perceived to be more enjoyable, more credible, and more informative during a moving task [27]; more attentive and more helpful during a drumming game [28]; and more engaging as a therapist for older adults and individuals suffering from dementia [29].

Although robots are not a replacement for human teachers, robots have potential as pedagogical agents in education [30]. Academically, robots can conduct various learning tasks such as recalling lessons and reinforcing facts; and socially, they can create a positive learning environment through social actions such as attention guiding and communicativeness. To implement academic and social capabilities for a robot, HRI and AI are both applicable areas of interest. While HRI can address the interaction between the robot and the student for sociability, AI can address robot adaptability as it pertains to academic outcomes.
1.1 HUMAN-ROBOT INTERACTION

HRI is the field of study that involves understanding, designing, and evaluating robotic systems that communicate with humans [31]. HRI is applied in areas in which it is necessary for the robot to interact with the user [32], [33]. This is exactly the case in a tutoring scenario where a social interaction between the tutor (the robot) and the student is necessary for effective learning to occur [34].

Robots and education. Several research studies have investigated the use of robots for education. These studies have shown that social robots are useful supplemental tools for education. Yun and colleagues documented a study where students were instructed via a robot tele-operated by a teacher, which led to learning gains for students [35]. Another study investigated the conceptual design of an educational robot that engaged students in a lesson about historical ancient cultures [36]. Though the robot’s sociability has been shown to contribute to student achievement, little has been done to illustrate the specific aspects of the robot that facilitate learning and retention [37], [38].

Social robots have also been widely used to support mathematics education. Brown and Howard used verbal cues to minimize idle time and decrease boredom during a mathematics test [39]. In another study, researchers used personalization to students while playing an adaptive arithmetic game with a robot [40]. Ramachandran and colleagues used a social robot that aided students while practicing fractions [41]. Socially responsive feedback (i.e., task-related feedback, motivational support), was effective in a robot learning companion that helped students practice mathematics problems [42].

Robots have also demonstrated positive trends among student perception and engagement [43]. One study documented how a robot’s perceived sociability increased
from the pre- to post-questionnaires during a mathematics tutoring session [44]. Howley et al. documented that students were more willing to ask the robot questions over a human tutor in most situations due to varying perceptions of the robot’s social role during a tutoring session [45]. Kanda et al. concluded that the social behavior of the robot aided in facilitating a better relationship with the student and increased the student’s social acceptance of the robot during a mathematics lesson [46]. The implementation of adaptive robots is an important topic in HRI; however, AI can be applied to develop robots that adapt and respond to a student’s needs.

1.2 ARTIFICIAL INTELLIGENCE

AI is the field of study that involves synthesizing and analyzing computational agents that can act intelligently. An intelligent agent can make decisions about its actions based on factors such as goals/values, prior knowledge, observations, past experiences, and the environment [47]. Figure 1.3 illustrates an agent that uses inputs to influence its actions.

![Agent system diagram](image)

**Figure 1.3 Agent system [47]**

AI can be thought of in four ways: systems that think like humans; systems that act like humans; systems that think rationally; and systems that act rationally. Systems that think like humans automate processes that require human thinking such as decision-
making and problem solving [48]. Systems that act like humans are systems that mimic human actions [49]. Systems that think rationally use information to perceive, reason, and act [50]. Lastly, systems that act rationally automate intelligent behavior and use information to achieve the maximum goal [51].

An effective human tutor adapts to the student (tutee) by gathering information about the student (e.g., capabilities, motivations, etc.) and tailoring real time instruction to meet the learning needs of the student [52]. This adaptability makes AI a probable approach to intelligent tutoring systems. Agents rely on an array of inputs such as student’s prior knowledge, common student errors, or facial expressions which can be used to conduct activities (i.e., assess student knowledge and provide relevant feedback). Figure 1.4 shows a sample agent system as a tutor.

Figure 1.4 Agent system (tutor)
Adaptive techniques have been applied using artificial intelligence in several studies across multiple domains for learning. Many popular AI models including Bayesian networks and Partially observable Markov decision processes (PomDPs) have been widely used, but they are not ideal to use for this application of work. Based on its capabilities, decision trees are potentially more effective in robot education; however, very little research has been conducted on using decision trees to develop a socially adaptive robot tutor.

Previous work has focused on adaptive tutoring and the robot’s [or computer’s] response once information is inferred. In some cases, social responses are reactions to a student’s state to aid in academic success [53], [54], [55], [56].

Decision trees. A decision tree is a model used for classifying data and is one of the most effective methods used for supervised classification learning\(^1\). A tree is built per its training data, which it uses to make classifications. The internal nodes in a decision tree represent the tree’s features and its classes are represented by the tree’s leaves [57]. Figure 1.4 shows a sample decision tree that uses four predictors (outlook, temperature, humidity, and wind) to determine a decision (yes, no) to play golf.

1.3 SUMMARY

Due to the need for student enrichment in the math, and the benefits of using robots for education, socially adaptive robots are ideal as a teaching tool for mathematics education. Social robots are not only capable of delivering mathematics content, but they are also capable of socially interacting with students to promote an enriching educational experience. However, how do we develop a socially adaptive robot with reasoning skills and an intuition about the student’s emotional state?

\(^1\) Supervised classification learning uses labeled training data to construct a classifier.
Figure 1.5 Sample decision tree [58]
CHAPTER 2

IMPETUS OF RESEARCH

2.0 PROBLEM

Educators have expressed that to best serve students, a robot tutor must possess reasoning skills and the robot must be capable of having an intuition about the student’s emotional state [1]. To date, there is a lack of literature that describes implementation of a socially adaptive robot tutor that uses a decision tree model to predict student emotion and performance for practicing multiplication via effective teaching techniques (i.e., guided practice and progress monitoring).

2.1 PAST STUDIES

We have conducted several studies that investigated the use of a social robot for education that will be applied to this proposed body of work. Our research introduced the robot tutor, Ms. An (Meeting Students’ Academic Needs) and investigated student’s perceptions and academic outcomes when practicing multiplication with Ms. An.

Study 1 was a study about attitudes. In this early study, we investigated student’s attitudes and perceptions toward a social robot. We developed social behaviors on the robot and students did, in fact, perceive the robot as a social entity that they enjoyed studying with [59], [60].

The next logical step was to then investigate how this social robot compared to commonly used educational technology. Therefore, Study 2 was a study about viability. In that study, we made a direct comparison between the social robot (Ms. An) and a tablet
workbook. Again, we found that the students perceived the robot as social. Importantly, students indicated a preference for studying with the robot compared to the workbook. This study verified that the implementation of a social robot tutor is in fact a viable option for education. We did not find any educational gains; however, this needs to be further investigated [61].

The final preliminary study shifted toward understanding educator’s (i.e., teachers, principals, interventionists) perceptions toward the robot. Even if students enjoy the tutor, educators will directly impact whether the robot is adopted into schools and used as a classroom aid. Therefore, study 3 was a study about acceptance. Educators indicated several uses for a robot-teaching assistant, including motivation for students, assistance with classroom activities, and encouragement/emotional support to students. While educators saw many benefits of a robot tutor, they were not without any concerns. Educators indicated that a robot tutor lacked reasoning skills and intuition. They also felt that a robot tutor would decrease personalization. Lastly, educators had doubts about being able to operate the robot [1].

2.2 RESEARCH GOAL

The goal of this research was to use a decision tree model to develop a socially adaptive robot tutor that predicted and responded to student emotion and performance to actively engage students in mathematics education.

Research questions. To assess the research goal (i.e., effectiveness of a robot’s ability to educate and engage students), this study addressed the following research questions:

[Q1] How well can a decision tree model classify a student’s emotion and performance?

[Q2] How well can a socially adaptive robot tutor engage 5th grade students to
practice multiplication?

a) How do students perform academically by studying with a socially adaptive robot tutor?

b) How do students respond emotionally by studying with a socially adaptive robot tutor?

[Q3] What social perceptions do students have of a socially adaptive robot tutor while practicing multiplication?

To address these research questions, we conducted a study in which students interacted with a robot during multiple tutoring sessions. We recorded information (such as delay in answer) that was needed to help the robot make predictions about the student. We collected information about each student’s mathematics performance before, during, and after the tutoring sessions as well as information about each student’s emotional states throughout the study. We also gathered information about the student’s perceptions and opinions of the robot tutor.
CHAPTER 3

METHODOLOGY

3.0 PLATFORM

We used the NAO humanoid robot (see Figure 3.1) as the robot tutor named Ms. An (Meeting Students’ Academic Needs). The NAO humanoid robot is an ideal platform for delivering education because of its multimodal capabilities such as speech and gesture. The NAO stands 58 cm tall. It has 25 degrees of freedom, 2 cameras, various touch sensors, and 4 microphones. The robot is also capable of voice and vision recognition.

Figure 3.1 NAO humanoid robot [62]
3.1 LESSON

State Standards. The multiplication tutoring session covered problems that addressed the South Carolina state standards:

- (4th grade) 4.NSBT.5 Multiply up to a four-digit number by a one-digit number and multiply a two-digit number by a two-digit number using strategies based on place value and the properties of operations.

- (5th grade) 5.NSBT.5 Fluently multiply multi-digit whole numbers using strategies to include a standard algorithm [63].

Content. The content of the lessons spanned across the different ways in which multiplication can be described through equal groups, area arrays, and comparison [64]. Table 3.1 shows an example of the representation of each problem type with corresponding context, computation, and pictorial interpretations.

Table 3.1 Multiplication problem types

<table>
<thead>
<tr>
<th>Context</th>
<th>Computation</th>
<th>Pictorial</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Equal groups</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>There are 4 bags of apples with 6 apples in each bag. How many apples are there in all?</td>
<td>24</td>
<td>[6] [6] [5] [6]</td>
</tr>
<tr>
<td><img src="image" alt="Equal groups pictorial" /></td>
<td><img src="image" alt="Equal groups pictorial" /></td>
<td></td>
</tr>
<tr>
<td><strong>Area arrays</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>There are 3 rows of desks with 4 desks in each row. How many rows are there?</td>
<td>12</td>
<td><img src="image" alt="Area arrays pictorial" /> <img src="image" alt="Area arrays pictorial" /> <img src="image" alt="Area arrays pictorial" /> <img src="image" alt="Area arrays pictorial" /></td>
</tr>
<tr>
<td><img src="image" alt="Area arrays pictorial" /> <img src="image" alt="Area arrays pictorial" /> <img src="image" alt="Area arrays pictorial" /> <img src="image" alt="Area arrays pictorial" /></td>
<td><img src="image" alt="Area arrays pictorial" /> <img src="image" alt="Area arrays pictorial" /> <img src="image" alt="Area arrays pictorial" /> <img src="image" alt="Area arrays pictorial" /></td>
<td></td>
</tr>
</tbody>
</table>
What is the area of a 3 inch by 5-inch rectangle?  

Comparison  
A string of yarn is unrolled to 2 cm long. How long will the string be if unrolled to 3 times as long?  

Students practiced multiplication with problems that included multiplying whole numbers by up to four digits and one digit and multiplying two-digit numbers by two-digit numbers. To ensure record of a wholistic multiplication experience, students solved problems with different combinations of multiplication question and answer types. For example, Figure 3.2 shows a session question that was given as a context question type and pictorial (equal groups) answer type.

Figure 3.2 Example session question
Common misconceptions. Common misconceptions or error patterns are mistakes students could make while solving multiplication problems [20]. We embedded common misconception or error patterns in each multiple choice incorrect response. Figure 3.3 shows a problem and details the correct solution using the standard multiplication algorithm and misconception for each answer choice [65].

![Sample problem with common misconceptions](image)

Figure 3.3 Sample problem with common misconceptions

Progress monitoring. During progress monitoring, Ms. An prompted students to answer multiple choice questions. They used a dedicated screen space on the tablet for scratch work. Figure 3.4 shows the progress monitoring interface.
Guided practice. Guided practice contained two parts. In part 1, Ms. An demonstrated a problem to the student while he/she followed along. In part 2, Ms. An and the student worked on a problem together. During this time, Ms. An guided the student through each intermediate step to solve the problem. Once complete, instruction returned to progress monitoring so that the student could solve a similar problem independently. Figure 3.5 shows a guided practice example using the standard multiplication algorithm.
Figure 3.5 Guided practice example

3.2 DEVELOPMENT

*Robot.* Ms. An was programmed using python and the Aldebaran Python software development kit (SDK).

*Robot sociability.* Robot sociability was implemented to mimic the behaviors and intuitions of a human tutor. For robot sociability, Ms. An predicted each student’s affective state and performance and responded accordingly. See Table 3.2.

Table 3.2 Example robot responses

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Example response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happy</td>
<td>“I am really enjoying working on these math problems with you.”</td>
</tr>
<tr>
<td></td>
<td>“Practicing math makes me happy.”</td>
</tr>
<tr>
<td></td>
<td>“I’m glad we are doing this.”</td>
</tr>
<tr>
<td>Angry</td>
<td>“Right now, this math is frustrating me.”</td>
</tr>
<tr>
<td></td>
<td>“This work is making me feel a little angry.”</td>
</tr>
<tr>
<td></td>
<td>“Practicing math can sometimes be irritating.”</td>
</tr>
<tr>
<td>Emotion</td>
<td>Examples</td>
</tr>
<tr>
<td>---------</td>
<td>----------</td>
</tr>
</tbody>
</table>
| Sad     | “This work is making me feel a little down.”  
          “Right now, this math is making me sad.”  
          “I’m a little sad working on these problems right now.” |
| Surprised | “Wow, this is great!”  
             “Practicing math with you is so much fun!”  
             “This is exciting!” |
| Neutral | No response |

For affect, each emotion (i.e., happy, angry, sad, surprised, neutral) was classified by valence (i.e., high, low) and control/power (i.e., negative, positive), which corresponded to five possible robot states: \{high, low\} \times \{negative, positive\} + \{neutral\}. Each emotion (apart from neutral) fell into one of four categories as described by the Geneva Emotion Wheel (GEW) [18]. See Figure 3.6.

![GEW Emotion Wheel](image)

Figure 3.6 GEW emotion classifications

Figure 3.7 shows the state diagram of Ms. An’s response per the student’s projected emotion.
Via a performance decision tree (see Section 3.2 Development: Decision Trees), if Ms. An predicted that the student would likely answer the upcoming question incorrectly, she provided guided practice before presenting a problem; however, if Ms. An predicted the student would answer correctly, she presented the problem without any intervention. Figure 3.8 shows the state diagram of the Ms. An’s response per the student’s predicted performance.

Decision trees. The decision trees were built using Weka, a machine learning software package [66] that used the C4.5 algorithm [67]. The C4.5 algorithm uses a training set to build decision trees by recursively calculating the entropy to determine
which features are most useful for splitting the data [67]. To prune the tree, Weka uses a post-pruning technique that removes nodes that are not statistically significant [68].

The emotion decision tree predicted the students’ emotion and the performance decision tree predicted the students’ performance. We used data from a previous study [61] for the training set to build both trees for this study. The training set for both trees contained 120 instances.

The features of the training set (see Tables 3.3 and 3.4) for the decision trees were the students’ gender, coded emotion (excluded in emotion decision tree), previous session score, answer delay, and percent correct (current session score). Researchers analyzed videos of the study sessions to code students’ emotions. The pre-test score was used for the previous session score for the first session. For pre-test score and percent correct,
scores greater than 0.80 were categorized as high; scores between 0.80 and 0.50 were categorized as medium; and scores below 0.50 were categorized as low. For delay in answer, we categorized times greater than 80 seconds as high; times between 80 seconds and 50 seconds were categorized as medium; and times below 50 seconds were categorized as low. The classification for the emotion decision tree was the students’ emotion and the classification for the performance decision tree was the students’ performance.

Table 3.3 Decision tree features for robot adaptability (emotion)

<table>
<thead>
<tr>
<th>Features</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Pre-test score</td>
</tr>
<tr>
<td></td>
<td>Delay in answer</td>
</tr>
<tr>
<td></td>
<td>Percent correct</td>
</tr>
<tr>
<td>Gender</td>
<td>Emotion</td>
</tr>
<tr>
<td>male, female</td>
<td>high, medium, low</td>
</tr>
<tr>
<td></td>
<td>high, medium, low</td>
</tr>
<tr>
<td></td>
<td>high, medium, low</td>
</tr>
<tr>
<td></td>
<td>neutral, happy, angry, sad, surprised</td>
</tr>
</tbody>
</table>

Table 3.4 Decision tree features for robot adaptability (performance)

<table>
<thead>
<tr>
<th>Features</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Emotion</td>
</tr>
<tr>
<td></td>
<td>Pre-test score</td>
</tr>
<tr>
<td></td>
<td>Delay in answer</td>
</tr>
<tr>
<td></td>
<td>Percent correct</td>
</tr>
<tr>
<td>Gender</td>
<td>Performance</td>
</tr>
<tr>
<td>male, female</td>
<td>neutral, happy, angry, sad, surprised</td>
</tr>
<tr>
<td></td>
<td>high, medium, low</td>
</tr>
<tr>
<td></td>
<td>high, medium, low</td>
</tr>
<tr>
<td></td>
<td>high, medium, low</td>
</tr>
<tr>
<td></td>
<td>correct, incorrect</td>
</tr>
</tbody>
</table>

Figures 3.9 and 3.10 show the emotion decision tree classified by individual emotions and rules, respectively. Figures 3.11 and 3.12 show the performance decision tree and rules, respectively. The pruned tree for emotion only included percent correct and gender.
features. The pruned tree for performance only included the percent correct feature.

![Decision Tree](image)

Figure 3.9 Emotion decision tree classified by individual emotions

- R₁: IF (percent_correct=high) THEN emotion = neutral
- R₂: IF (percent_correct=medium) AND (gender=female) THEN emotion = neutral
- R₃: IF (percent_correct=medium) AND (gender=male) THEN emotion = sad
- R₄: IF (percent_correct=low) THEN emotion = sad

Figure 3.10 Emotion rules classified by individual emotions

![Performance Decision Tree](image)

Figure 3.11 Performance decision tree

- R₁: IF (percent_correct=high) OR (percent_correct=medium) THEN performance=correct
- R₂: IF (percent_correct=low) THEN performance=incorrect

Figure 3.12 Performance rules

The rate of correctly classified instances for the performance tree was 87.5%. The rate of correctly classified instances for the emotion tree classified by individual emotions
was low (50%), meaning the tree only guessed correctly half the time. Thus, we decided to also consider an emotion decision tree that classified emotion by valence (non-negative, negative). Figure 3.13 shows an emotion decision tree classified by valence (non-negative, negative). Figure 3.14 shows the rules for the emotion decision tree classified by valence.

![Emotion decision tree](image)

**Figure 3.13 Emotion decision tree classified by valence**

R₁: IF (percent_correct=high) THEN emotion = non-negative  
R₂: IF (percent_correct=medium) AND (gender=female) THEN emotion = non-negative  
R₃: IF (percent_correct=medium) AND (gender=male) THEN emotion = negative  
R₄: IF (percent_correct=low) THEN emotion = negative

**Figure 3.14 Emotion rules classified by valence**

The emotion decision tree classified by valence had a better rate of predictability (76.77%) than the emotion tree that classified each individual emotion (50%).

*Interface.* The interface (see Figure 3.15) was developed using Python and its PyQt package. The interface was designed to allow educators (who will most likely be novice robot users) to quickly and easily use Ms. An in a tutoring session, where an expert user (e.g., developer) will likely not be present for setup. Further, the interface adheres to human-computer interaction (HCI) design principles (visibility, feedback, affordance, mapping, and consistency).
Current tools that are available to operate the robot are best used by experts and not ideal for the target population (i.e., educators): 1) Choregraphe software suite (see Figure 3.16) and 2) Python (see Figure 3.17).

Figure 3.15 Interface, start page

Figure 3.16 Choregraphe interface
3.3 HRI STUDY DESIGN

To analyze the effectiveness of the adaptive robot, we conducted a between-subjects study\(^2\) with two conditions: A) with a non-adaptive robot (control group); and B) with a socially adaptive robot (adaptive group). Table 3.5 shows a comparison of robot traits and behaviors for each condition.

### Table 3.5 Adaptive robot versus non-adaptive robot

<table>
<thead>
<tr>
<th>Non-Adaptive Robot (Control)</th>
<th>Adaptive Robot</th>
</tr>
</thead>
<tbody>
<tr>
<td>▪ Static emotional state (neutral)</td>
<td>▪ Dynamic emotional state to match student’s emotional state (happy, angry, sad, surprised, neutral)</td>
</tr>
<tr>
<td>▪ Asked multiplication questions [progress monitoring] without instructional support</td>
<td>▪ Asked multiplication questions [progress monitoring] with instructional support [guided practice]</td>
</tr>
</tbody>
</table>

In the adaptive robot condition, Ms. An predicted the student’s emotion and

---

\(^2\) Between subjects study is one in which two or more groups are tested under different conditions.
performance and proactively determined next actions before presenting a question to be solved. For emotion classification, Ms. An used social responses that corresponded to each student’s emotional state (happy, angry, sad, surprised, neutral) before asking a question. For performance, if Ms. An predicted that the student would answer the upcoming question correctly, it proceeded with progress monitoring; however, if Ms. An predicted that the student would answer the upcoming question incorrectly, the robot proceeded with guided practice. In the non-adaptive condition, Ms. An behaved in a neutral emotional state and completed only progress monitoring activities (i.e., asked mathematics questions to be solved without any intervention) despite the student’s affective state or competency. Figures 3.18-3.21 illustrate example scenarios of both conditions.

Figure 3.18 Ms. An predicts that the student is sad and that the student answer the upcoming question incorrectly (adaptive)

Figure 3.19 Ms. An predicts that the student is happy and that the student will answer the upcoming question correctly (adaptive)
Student-Robot Interaction. In addition to gestures and movements, Ms. An communicated with the students verbally using speech and visually through the tablet. Ms. An performed actions such as reading each multiplication question, prompting the students at various points during the lesson, and giving feedback on answer choices (verbal communication). These actions corresponded with the question and activity display on the tablet (visual communication). Students could press buttons on the touch screen interface to select answer choices and enter values using a keypad. Ms. An received that data from the tablet and responded accordingly. Figure 3.22 shows the student-robot communication diagram.
Participants. Twenty 4th and 5th grade students were included in the study (10 males, 10 females), ranging from 9-12 years old ($M=9.95$, $SD=0.84$). Participants were recruited from Blenheim/Elementary Middle School in rural Blenheim, SC. Of those participants, 50% identified themselves as Black/African American; 30%, White/Caucasian, 5%, other; and 5% opted not to report race/ethnicity.

Participants completed a Technology Experience Profile that measured their familiarity with and use of different technologies. While the students rated an overall familiarity ($M=3.59$, $SD=1.12$) with technology, their experience with robots, specifically, was low ($M=2.95$, $SD=1.28$). The top technologies which the student reported using on at least a weekly basis (e.g., $M=4.0$ or higher) were video games, tablet, smart board, smart phone, music player, and social media. The least used technologies ($M=3.0$ or below) were webcam, electronic book reader, LCD projector, student response systems, robots, and a camera. See Figure 3.23.
Eleven students were assigned to the control group and nine students were assigned to the adaptive group. To ensure the groups were equally split by student performance, we used the pre-test 3 scores to assign students to each condition. Students in the control group had a 32% (SD=14) average score and students in the adaptive group had a 28% (SD=9) average score.

3.4 MATERIALS AND MEASURES

Workbook. Students used the Dell Inspiron 13 7000 series 2-in-1 laptop/tablet during each session for both conditions. They interacted with Ms. An using the tablet by completing activities such as reviewing content and answering questions. This form of input increased the integrity of the data as it eliminated errors that could have occurred from other methods of input such as voice recognition and vision recognition.

Student demographic form. Students completed a demographic form (Appendix A). We used data from this form for descriptive statistics.

Pre-/post-test. Students completed two pre-/post-tests for the study. They completed...
the pre-tests at the beginning of the study. In test 1 (Appendix B), students were asked to select the different ways to represent a multiplication problem. This was a test to see if students would recognize the equivalence of mathematics problems as computation, context, and pictorial. In test 2 (Appendix C), students solved multiplication problems. This test was used to establish a baseline for the student’s multiplication performance as well as ensure equivalent groups for performance.

At the end of the study, after all sessions were finished, students completed post-tests that were analogous to the pre-tests. The tests were multiple choice and developed and administered through SurveyMonkey (https://www.surveymonkey.com/).

Emotions questionnaire. Students were asked to best describe their emotions (i.e., happy, angry, sad, surprised, neutral) at the beginning, in the middle, and at the end of each session (Appendix D). Students rated their emotions using emojis to ensure student understanding of the meaning of each emotion which has been shown to increase accuracy in student responses [69]. The students completed the emotion questionnaire directly on the tablet when prompted by the robot that said, “Tell me how you’re feeling. Choose the emoji that best describes how you’re feeling.”

Robot persona inventory. We adapted the Agent Persona Inventory (API) [70] to create the Robot Persona Inventory (RPI) to measure the participant’s attitudes toward the robot tutor (Appendix E). API is a reliable, validated instrument that measures four pedagogical agent persona factors (credibility, facilitated learning, engagement, and human-likeness) and two latent variables (informational usefulness and affective interaction) [70].

Facilitated learning consists of how the agent enables learning and reflection; credible
focuses on the value of the instruction from the agent; human-like addresses how natural the agent’s communication enhances the its personality and emotional expression; and engaging relates to how well the agent motivates the student [70].

Information usefulness corresponds to the agent’s ability and influences the agent factors facilitated learning and credible. Affective interaction corresponds to the agent’s personality and influences the agent factors human-like and engaging. Figure 3.24 shows the relationship among the agent persona factors and latent variables [70].

![Figure 3.24 Agent characteristics and relationship to agent factors](image)

RPI captured data related to student’s perceptions of the robot’s social presence, personality, and method of instruction [70]. We worked with elementary school educators to ensure that the content was on a 5th grade reading level. The RPI was developed and administered through SurveyMonkey.

**Interview.** We developed a three-part interview script to collect qualitative data regarding the participants’ opinions about studying with Ms. An (Appendix F). The first part asked questions that focused on HRI; the second part, Education; and the third part, AI. The interviews were conducted one-on-one in a closed office. Students engaged in generative questioning that allowed each student to ask themselves further questions to develop meaning and deepen their comprehension of their session with Ms. An. This
allowed the students to better communicate their thoughts and feelings about their interaction with Ms. An.

3.5 PROCEDURE

Students engaged in one-on-one tutoring sessions to practice multiplication per the South Carolina 4th and 5th grade mathematics state standards. Mainly due to time and resource constraints, most education interventions using robotics are short-term interventions (some being as short as one interaction) [71]. To have a longer intervention and multiple interactions, students worked with Ms. An for two sessions, having one session a week for two weeks. Sessions lasted approximately 30-45 minutes.

Prior to the study session, students completed a student demographic form, technology experience profile, and multiplication pre-test. Then, each student was asked to sit in a small room and the students worked at a desk with the robot. The student began each session by answering the emotions questionnaire. Next, they interacted with the robot. For both the adaptive and non-adaptive groups, the robot acted as a tutor and completed progress monitoring activities. The robot asked students multiplication questions. Each question was displayed on the tablet and students answered questions via the tablet interface. Contrary to the non-adaptive robot, the adaptive robot employed proactive behaviors and executed those behaviors when needed (see sections 3.2 and 3.3). Students completed the emotions questionnaire again, halfway through the study session. Once the tutoring session was complete, students completed the emotions questionnaire a final time.

After all sessions were completed, students completed a final session on solving multiplication using the partial products technique. In this session, students began with
guided practice then concluded with progress monitoring.

At the end of all three sessions, students completed a post-test, RPI questionnaire, and interview. Students were given a retention test after all students completed the study.

Figure 3.25 details the study procedure for each student.
CHAPTER 4
RESULTS

4.0 DATA ANALYSIS

Unless otherwise noted, alpha was set at .05 for all statistical tests. Due to the small sample size in each group, we report this data with guarded generalizations. We indicate all data that are statistically significant with an asterisk (*).

4.1 DECISION TREE MODEL

Research question 1 (How well can a decision tree model classify a student’s emotion and performance?) addressed the accuracy of a decision tree model. To evaluate the robot’s ability to make classifications for emotion and performance and to better understand where misclassifications could occur: 1) for emotion, we compared the robot’s prediction for each student’s emotion to the student’s self-reported emotions throughout the session and 2) for performance, we compared the robot’s prediction for each student’s performance to the student’s actual performance. We also considered emotion classifications that were grouped by valence: negative (sad, angry) and non-negative (happy, surprised, neutral).

Although closely related, there were differences in the study from which we used the training data [60] and the current study. The previous study consisted of one session where students interacted with the robot and the pre-test score was the score used for pre-test feature. In the previous study, the emotion feature was derived from researchers coding the students’ emotions by reviewing videos of the sessions. Lastly, the previous
study only contained computation multiplication problems. In contrast, this current study consisted of two sessions where students interacted with the robot. The pre-test score was used for the pre-test feature in session 1 and the session 1 score was used as the pre-test feature in session 2. In this current study, the emotion feature was self-reported by the students.

Our use of data from a different, but similar study to build a tree for the current study is a popular technique known as transfer learning. Transfer learning in artificial intelligence is a technique in which the data used for a training set to solve one problem is applied as a training set to solve a similar problem [72]. While using this technique is common, it may have contributed to the low prediction accuracy for both training models to the new models.

These comparisons for emotion classifications are shown in the confusion matrices in Tables 4.1 and 4.2. The values along the diagonal of the matrices are the success rates for predictions.

Table 4.1 Confusion matrix for each emotion

<table>
<thead>
<tr>
<th></th>
<th>angry</th>
<th>happy</th>
<th>neutral</th>
<th>sad</th>
<th>surprised</th>
</tr>
</thead>
<tbody>
<tr>
<td>angry</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>happy</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>neutral</td>
<td>0.04</td>
<td>0.51</td>
<td>0.19</td>
<td>0.05</td>
<td>0.21</td>
</tr>
<tr>
<td>sad</td>
<td>0.00</td>
<td>0.40</td>
<td>0.20</td>
<td>0.00</td>
<td>0.40</td>
</tr>
<tr>
<td>surprised</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 4.2 Confusion matrix for emotion by valence

<table>
<thead>
<tr>
<th></th>
<th>negative</th>
<th>non-negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>negative</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>non-negative</td>
<td>0.09</td>
<td>0.91</td>
</tr>
</tbody>
</table>
The results show that the model is not accurate for each individual emotion (as expected from results of the training set described in Section 3.2). Despite the students exhibiting other emotions, the robot only predicted neutral and sad emotions. Happy was most commonly classified as neutral (51%). Happy was also misclassified as sad 40% out of all sad classifications. Surprise was also misclassified as neutral (21%) and sad (40%).

The classifications for non-negative emotions improved (91%) when they were grouped by valence. The groupings are helpful because they compensate for the subtlety of the students’ emotions. Further, since we would like to apply this model to tutoring interactions to improve emotional engagement, grouping by valence is adequate as emotional engagement is measured by the ratio of positive and negative emotions.

These comparisons for performance classification are shown in the confusion matrix in Table 4.3. The values along the diagonal of the matrix are the success rates for predictions.

<table>
<thead>
<tr>
<th></th>
<th>correct</th>
<th>incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>correct</td>
<td>0.54</td>
<td>0.46</td>
</tr>
<tr>
<td>incorrect</td>
<td>0.40</td>
<td>0.60</td>
</tr>
</tbody>
</table>

Incorrect performance was classified correctly at a higher rate than correct performance. However, the classifications were correct a little over half the time, which is only slightly better than choosing randomly.

To eliminate any losses from transfer learning, we built a new tree for performance using the data from this current study. Table 4.3 shows the confusion matrix for the new tree.
Table 4.4 Confusion Matrix for performance – data from current study

<table>
<thead>
<tr>
<th></th>
<th>correct</th>
<th>incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>correct</td>
<td>0.64</td>
<td>0.36</td>
</tr>
<tr>
<td>incorrect</td>
<td>0.31</td>
<td>0.69</td>
</tr>
</tbody>
</table>

This tree performed better than the tree used in the study. The classifier predicted correct performance at a rate of 0.64 and incorrect performance at a rate of 0.69.

4.2 ENGAGEMENT

Research question 2 (How well can a socially adaptive robot tutor engage 5th grade students to practice multiplication?) emphasized student engagement. To consider two aspects of engagement, research question 2 was comprised of two sub-questions. To address question 2a, how do students perform academically by studying with a socially adaptive robot tutor, we report average learning gains and percent correct by answer type. To address question 2b, how do students respond emotionally by studying with a socially adaptive robot tutor, we report the frequency of emotions exhibited throughout the study sessions.

Learning gains. The difference in pre-test and post-test scores is a measure of each participant’s learning gain during the study. We also calculated the difference in session 1 and session 2 scores to measure each participant’s learning gains. To allow for a reliable analysis for our between-subjects design, we calculated the normalized learning gain in each group [74].
normalized gain = \frac{\text{post-test}_\text{score} - \text{pre-test}_\text{score}}{1 - \text{pre-test}_\text{score}}

Figure 4.1 Formula for normalized gain (pre-/post-test)

**Pre-/post-test 1.** Pre-/post-test 1 was a test on students’ ability to identify the different ways to represent multiplication problems. Figure 4.2 shows the normalized average learning gains for pre-/post-test 1 for each condition. We conducted Wilcoxon Signed Rank tests to compare pre- to post-test scores in each condition. The adaptive conditions did show a statistically significant ($z=-2.06, p<.05$) improvement from pre- ($M=3.67, SD=1.32$) to post-test ($M=5.44, SD=2.83$) scores. Therefore, the adaptive robot did, in fact, promote learning gains in the students’ ability to identify the different ways to represent multiplication problems. The control condition did not have a significant change ($z=-0.239, p=.81$) from pre- ($M=4.63, SD=2.01$) to post-test ($M=4.63, SD=2.06$) scores. Therefore, the control condition did not yield learning gains in this skill.

Mann-Whitney U test was used to compare the learning gains between conditions. While the adaptive group had higher learning gains from pre- to post-test1 ($M=0.44, SD=0.22$) than the control group ($M=-0.15, SD=0.67$), this difference between groups was not statistically significant ($z=-1.62, p=.10$). It is important to note that although this is a promising trend, the there is no significant difference likely due to the variance in the control group being higher, and due to the small sample size.
Figure 4.2 Pre-/post-test 1 average learning gains

Figure 4.3 shows pre-/post-test 1 scores for each participant, separated by condition. Pre-/post-test 1 scores were calculated by the number of items correctly selected divided by twelve total correct options. Thicker lines represent more than one participant with the same score.

More students in the adaptive group (5 students) showed an increase in scores from pre- to post- than in the control group (4 students). The adaptive group did not contain any student’s scores to decrease from pre- to post- but the control group (5 students) had some students to show a decrease. Two students showed no change in the control group and 4 students showed no change in the adaptive group.
Pre-/post-test 2 was a test on students’ ability to correctly solve multiplication problems. Figure 4.4 shows the normalized average learning gains for pre-/post-test 2 for each condition. We conducted Wilcoxon Signed Rank tests to compare pre- to post-test scores in each condition. The adaptive conditions did not show a statistically significant ($z=-1.13$, $p=.26$) improvement from pre- ($M=0.28$, $SD=0.09$) to post-test ($M=0.39$, $SD=0.29$) scores. The control condition also did not have a significant change ($z=-1.66$, $p=.10$) from pre- ($M=0.32$, $SD=0.15$) to post-test ($M=0.46$, $SD=0.21$) scores. Therefore, the control condition also did not yield learning gains in this skill.

Mann-Whitney U test was used to compare the learning gains between conditions. Figure 4.4 shows the normalized average learning gains for pre-/post-test 2 for each condition. There was not a decrease in learning gains for either group in test 2; thus, the adaptive session did not negatively impact the students. There was not a statistically significant (via Mann-Whitney U test) difference in learning gains between the two conditions for test 2 ($z=-.34$, $p=.73$).
Figure 4.4 Pre-/post-test 2 average learning gains

Figure 4.5 shows pre-/post-test 2 scores for each participant, separated by condition. Pre-/post-test 2 scores were calculated by the number of items answered correctly divided by 18 total questions. Thicker lines represent more than one participant with the same score.

Figure 4.5 Pre-/post-test 2 scores

*Session scores.* Figure 4.6 shows the normalized average learning gains for session 1
and session 2 for each condition. Session scores were calculated were by the number of items answered correctly divided by 10 total questions for each session. More students in the adaptive group showed an increase from session 1 to session 2 than in the control group. The adaptive group also had less students to show a decrease from session 1 to session 2 than the control group. For the control group, 4 participants had an increase from session 1 to session 2; 3 participants showed no change; and 4 participants had a decrease from session 1 to session 2. For the adaptive group, 6 participants had an increase from session 1 to session 2; 1 participant showed no change; and 2 participants had a decrease from session 1 to session 2. The adaptive group had higher learning gains from session 1 to session 2 ($M=0.13$, $SD=0.29$) than the control group ($M=-0.49$, $SD=1.23$), although this difference was not statistically significant (Mann-Whitney U test, $z=-1.34$, $p=.18$) probably due to sample size.

![Figure 4.6 Session 1 to session 2 average learning gains](image)

**Figure 4.6 Session 1 to session 2 average learning gains**

*Percent correct by answer type. We categorized each item by answer type (computation, context, pictorial (area array), and pictorial (equal groups)) in pre-/post-test*
2. Figure 4.7 shows the percent correct by answer type for pre-/post-test 2 for each condition. Percent correct for each answer type was calculated by the number of students who answered correctly divided by the number of occurrences for the answer type. Seven questions had a Computation answer; nine, Context; three Pictorial (area array); and one Pictorial (equal groups).

![Figure 4.7 Pre-/post-test 2 percent correct by answer type](image)

Participants in the adaptive group increased performance from pre- to post- for all answer types. However, participants in the control group increased performance from pre- to post- for all answer types except Context. The major improvements occurred in Pictorial (equal groups) (\(\Delta_{\text{adaptive}} = 0.46, \Delta_{\text{control}} = 0.32\)) and Computation (\(\Delta_{\text{adaptive}} = 0.14, \Delta_{\text{control}} = 0.27\)) for both groups. Participants in the adaptive group had a higher rate of increase for Pictorial (area array) and Pictorial (equal groups) than participants in the control group. This could be a result of the adaptive’s robot effectiveness in addressing the highest ranked common misconception in this category, reversed group and count.

*Delay in answer*. The delay in answer was the time (in seconds) from when the robot read the mathematics question to the time that the student provided an answer. Table 4.5
details the delay in answer for each session per group. Students in the adaptive group had a higher delay in answer (session 1: $M=51.47, SD=35.83$; session 2: $M=91.84; SD=97.78$) than in the control group answer (session 1: $M=43.84, SD=23.89$; session 2: $M=74.62; SD=33.11$) for both sessions. The delay in answer also increased in session 2 for both groups. The delay in answer data were not statistically significant but the data show an encouraging trend.

Table 4.5 Average delay in answer (seconds)

<table>
<thead>
<tr>
<th></th>
<th>Session 1</th>
<th>Session 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>43.84</td>
<td>74.62</td>
</tr>
<tr>
<td>Adaptive</td>
<td>51.47</td>
<td>91.84</td>
</tr>
</tbody>
</table>

Frequency of emotions. We assumed that students were equally likely to select any of the 5 emotions (happy, angry, sad, surprised, neutral), and calculated a Pearson’s Goodness of Fit Chi Square. The chi square for both the control ($X^2=37.77$) and the adaptive ($X^2=42.62$) were significant ($p<.001$), suggesting that the distribution of reported emotions were not evenly reported. Students significantly reported happiness more often than other emotions. Students were more likely to feel surprised in the control condition. This could be because they had less feedback/coaching on how they were doing. The emotions sadness and anger were not commonly reported.

4.3 ROBOT SOCIABILITY

Research question 3 (What social perceptions do students have of a socially adaptive robot tutor while practicing multiplication?) focuses on students’ perceptions of a robot tutor. To address research question 3, we used the results of the RPI questionnaire. We analyzed the RPI by overall mean for each group, agent factors, agent characteristics, and by each individual item.

We conducted Wilcoxon Signed Rank tests to compare how the mean value for each
RPI category (rating scale: 1=strongly disagree; 3 = neither agree nor disagree; 5= strongly agree) differed from neutral (neither agree nor disagree).

Figure 4.8 Overall RPI rating

Figure 4.8 shows the overall RPI rating for both groups. There was not a noticeable difference for the overall RPI in the control group ($M=3.72, SD=0.55$) and adaptive group ($M=3.86, SD=1.03$). Both groups had favorable perceptions about neutral.

RPI by agent factor was categorized by the following categories: facilitated learning, credible, human-like, and engaging. As shown in Figure 4.9, participants viewed the adaptive condition as facilitating learning ($z=-2.20, p<.05$) and more credible ($z=-2.14, p<.05$) than neutral. The adaptive condition for engaging was not significantly different than neutral, but the result was marginally significant ($z=-1.80, p=.07$). Similarly, for the control condition, participants viewed the robot as facilitating learning ($z=-2.81, p<.05$), credible ($z=-2.71, p<.05$), and engaging ($z=-2.95, p<.05$) compared to neutral. Neither condition was significantly viewed as human like ($p>.05$).
RPI by agent characteristics is categorized by informational usefulness (ability) and affective interaction (personality). Figure 4.10 shows the results of RPI by agent characteristics. Students in the adaptive group rated Ms. An similarly for informational usefulness ($M=4.10$, $SD=0.39$) as those in the control group ($M=3.92$, $SD=0.20$). Students in the adaptive group rated Ms. An similarly for affective interaction ($M=3.52$, $SD=0.43$) as those in the control group ($M=3.50$, $SD=0.20$).

Last, we wanted to compare each RPI item. We conducted Mann-Whitney U tests to compare the control vs. adaptive condition for each questionnaire item. None of these
comparisons were significant, most likely due to too small of a sample size for between-groups comparisons.

We then conducted Wilcoxon Signed Rank tests to conduct a within-group comparison for each individual questionnaire item – comparing the mean to 3.00 (neutral). These findings are shown in Tables 4.6 and 4.7.

Table 4.6 RPI by item for control

<table>
<thead>
<tr>
<th>RPI Item</th>
<th>Control</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>The robot helped me think about multiplication more deeply.</td>
<td>3.64</td>
<td>1.07</td>
<td>4</td>
<td>1-5</td>
<td>11</td>
<td>-1.658 .097</td>
</tr>
<tr>
<td>The robot made multiplication interesting.</td>
<td>4.18</td>
<td>1.11</td>
<td>4</td>
<td>1-5</td>
<td>11</td>
<td>-2.240 .025</td>
</tr>
<tr>
<td>The robot encouraged me to think about what I was learning.</td>
<td>3.82</td>
<td>0.54</td>
<td>4</td>
<td>1-5</td>
<td>11</td>
<td>-2.179 .029</td>
</tr>
<tr>
<td>The robot kept my attention.</td>
<td>3.36</td>
<td>0.48</td>
<td>4</td>
<td>1-5</td>
<td>11</td>
<td>-3.085 .002</td>
</tr>
<tr>
<td>The robot showed me the information effectively.</td>
<td>3.36</td>
<td>1.15</td>
<td>4</td>
<td>1-5</td>
<td>11</td>
<td>-1.027 .305</td>
</tr>
<tr>
<td>The robot helped me to concentrate on the lesson.</td>
<td>3.27</td>
<td>1.42</td>
<td>4</td>
<td>1-5</td>
<td>11</td>
<td>-1.512 .131</td>
</tr>
<tr>
<td>The robot helped me focus on the important information.</td>
<td>4.18</td>
<td>0.83</td>
<td>4</td>
<td>1-5</td>
<td>11</td>
<td>-2.667 .008</td>
</tr>
<tr>
<td>The robot improved my knowledge of multiplication.</td>
<td>3.27</td>
<td>1.29</td>
<td>4</td>
<td>1-5</td>
<td>11</td>
<td>-0.690 .490</td>
</tr>
<tr>
<td>The robot was interesting.</td>
<td>4.18</td>
<td>0.57</td>
<td>4</td>
<td>1-5</td>
<td>11</td>
<td>-2.919 .004</td>
</tr>
<tr>
<td>The robot was enjoyable.</td>
<td>3.82</td>
<td>1.27</td>
<td>4</td>
<td>1-5</td>
<td>11</td>
<td>-1.727 .084</td>
</tr>
<tr>
<td>The robot seemed knowledgeable.</td>
<td>4.09</td>
<td>1.16</td>
<td>5</td>
<td>1-5</td>
<td>11</td>
<td>-2.364 .018</td>
</tr>
<tr>
<td>The robot seemed intelligent.</td>
<td>4.09</td>
<td>1.31</td>
<td>5</td>
<td>1-5</td>
<td>11</td>
<td>-2.059 .039</td>
</tr>
<tr>
<td>The robot was helpful.</td>
<td>4.27</td>
<td>0.86</td>
<td>4</td>
<td>1-5</td>
<td>11</td>
<td>-2.697 .007</td>
</tr>
<tr>
<td>The robot was useful.</td>
<td>3.64</td>
<td>1.57</td>
<td>4</td>
<td>1-5</td>
<td>11</td>
<td>-1.204 .229</td>
</tr>
<tr>
<td>The robot seemed like a teacher.</td>
<td>4.00</td>
<td>1.13</td>
<td>4</td>
<td>1-5</td>
<td>11</td>
<td>-2.251 .024</td>
</tr>
<tr>
<td>The robot had a personality.</td>
<td>3.36</td>
<td>1.30</td>
<td>3</td>
<td>1-5</td>
<td>11</td>
<td>-0.553 .581</td>
</tr>
<tr>
<td>The robot’s emotion was natural.</td>
<td>3.55</td>
<td>1.16</td>
<td>3</td>
<td>1-5</td>
<td>11</td>
<td>-1.350 .177</td>
</tr>
<tr>
<td>The robot seemed like a human.</td>
<td>3.00</td>
<td>1.13</td>
<td>3</td>
<td>1-5</td>
<td>11</td>
<td>-0.073 .942</td>
</tr>
<tr>
<td>The robot’s movement was normal.</td>
<td>3.27</td>
<td>1.25</td>
<td>4</td>
<td>1-5</td>
<td>11</td>
<td>-0.548 .594</td>
</tr>
<tr>
<td>The robot showed emotion.</td>
<td>3.36</td>
<td>1.30</td>
<td>3</td>
<td>1-5</td>
<td>11</td>
<td>-0.933 .351</td>
</tr>
<tr>
<td>The robot was expressive.</td>
<td>3.55</td>
<td>0.89</td>
<td>4</td>
<td>1-5</td>
<td>11</td>
<td>-1.732 .083</td>
</tr>
<tr>
<td>The robot was enthusiastic.</td>
<td>4.00</td>
<td>0.74</td>
<td>4</td>
<td>1-5</td>
<td>11</td>
<td>-2.598 .009</td>
</tr>
<tr>
<td>The robot was entertaining.</td>
<td>3.27</td>
<td>1.14</td>
<td>3</td>
<td>1-5</td>
<td>11</td>
<td>-1.122 .112</td>
</tr>
<tr>
<td>The robot was motivating.</td>
<td>3.45</td>
<td>1.23</td>
<td>4</td>
<td>1-5</td>
<td>11</td>
<td>-1.221 .222</td>
</tr>
<tr>
<td>The robot was friendly.</td>
<td>4.18</td>
<td>0.57</td>
<td>4</td>
<td>1-5</td>
<td>11</td>
<td>-2.919 .004</td>
</tr>
</tbody>
</table>

Those items in Tables 4.5 and 4.6 with p-values less than .05 are considered statistically significant. For ease, the significant questionnaire items are listed in Table 4.8. As depicted in this table, more items from the facilitated learning construct were statistically significant in the adaptive condition. The robot was interesting was statistically significant for both groups. More items from the credible construct were statistically significant in the control condition. The robot seemed knowledgeable and the robot seemed like a teacher were statistically significant for both groups. No items in the human-like construct were statistically significant. Lastly, more items from the
engagement construct were statistically significant in the control condition. The robot 
was motivating was statistically significant for the adaptive group, which directly 
addressed the engaging persona factor (how well the agent motivated the student).

Table 4.7 RPI by item for adaptive

<table>
<thead>
<tr>
<th>RPI Item</th>
<th>Adaptive</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Facilitated Learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The robot helped me think about multiplication more deeply.</td>
<td></td>
<td>4.22</td>
<td>1.05</td>
<td>5</td>
<td>1-5</td>
<td>9</td>
<td>-2.399</td>
</tr>
<tr>
<td>The robot made multiplication interesting</td>
<td></td>
<td>3.67</td>
<td>1.33</td>
<td>4</td>
<td>1-5</td>
<td>9</td>
<td>-1.292</td>
</tr>
<tr>
<td>The robot encouraged me to think about what I was learning.</td>
<td></td>
<td>4.22</td>
<td>1.23</td>
<td>5</td>
<td>1-5</td>
<td>9</td>
<td>-2.771</td>
</tr>
<tr>
<td>The robot kept my attention.</td>
<td></td>
<td>4.00</td>
<td>0.94</td>
<td>4</td>
<td>1-5</td>
<td>9</td>
<td>-2.165</td>
</tr>
<tr>
<td>The robot showed me the information effectively.</td>
<td></td>
<td>4.22</td>
<td>1.31</td>
<td>5</td>
<td>1-5</td>
<td>9</td>
<td>-1.960</td>
</tr>
<tr>
<td>The robot helped me to concentrate on the lesson.</td>
<td></td>
<td>4.11</td>
<td>0.87</td>
<td>4</td>
<td>1-5</td>
<td>9</td>
<td>-2.332</td>
</tr>
<tr>
<td>The robot helped me focus on the important information.</td>
<td></td>
<td>3.89</td>
<td>1.45</td>
<td>5</td>
<td>1-5</td>
<td>9</td>
<td>-1.613</td>
</tr>
<tr>
<td>The robot improved my knowledge of multiplication.</td>
<td></td>
<td>3.78</td>
<td>1.31</td>
<td>4</td>
<td>1-5</td>
<td>9</td>
<td>-1.643</td>
</tr>
<tr>
<td>The robot was interesting.</td>
<td></td>
<td>4.33</td>
<td>1.25</td>
<td>5</td>
<td>1-5</td>
<td>9</td>
<td>-2.000</td>
</tr>
<tr>
<td>The robot was enjoyable.</td>
<td></td>
<td>4.00</td>
<td>1.15</td>
<td>4</td>
<td>1-5</td>
<td>9</td>
<td>-2.008</td>
</tr>
<tr>
<td>Credible</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The robot seemed knowledgeable.</td>
<td></td>
<td>4.22</td>
<td>1.31</td>
<td>5</td>
<td>1-5</td>
<td>9</td>
<td>-1.060</td>
</tr>
<tr>
<td>The robot seemed intelligent.</td>
<td></td>
<td>3.89</td>
<td>1.29</td>
<td>4</td>
<td>1-5</td>
<td>9</td>
<td>-1.857</td>
</tr>
<tr>
<td>The robot was helpful.</td>
<td></td>
<td>4.00</td>
<td>1.63</td>
<td>5</td>
<td>1-5</td>
<td>9</td>
<td>-1.473</td>
</tr>
<tr>
<td>The robot was useful.</td>
<td></td>
<td>4.44</td>
<td>0.96</td>
<td>5</td>
<td>1-5</td>
<td>9</td>
<td>-2.516</td>
</tr>
<tr>
<td>The robot seemed like a teacher.</td>
<td></td>
<td>4.22</td>
<td>1.47</td>
<td>5</td>
<td>1-5</td>
<td>9</td>
<td>-2.053</td>
</tr>
<tr>
<td>Human-like</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The robot had a personality.</td>
<td></td>
<td>3.56</td>
<td>1.71</td>
<td>5</td>
<td>1-5</td>
<td>9</td>
<td>-1.055</td>
</tr>
<tr>
<td>The robot's emotion was natural.</td>
<td></td>
<td>3.44</td>
<td>1.64</td>
<td>4</td>
<td>1-5</td>
<td>9</td>
<td>-0.807</td>
</tr>
<tr>
<td>The robot seemed like a human.</td>
<td></td>
<td>3.22</td>
<td>1.81</td>
<td>4</td>
<td>1-5</td>
<td>9</td>
<td>-1.375</td>
</tr>
<tr>
<td>The robot's movement was natural.</td>
<td></td>
<td>2.89</td>
<td>1.52</td>
<td>4</td>
<td>1-5</td>
<td>9</td>
<td>-1.365</td>
</tr>
<tr>
<td>The robot showed emotion.</td>
<td></td>
<td>3.11</td>
<td>1.66</td>
<td>3</td>
<td>1-5</td>
<td>9</td>
<td>-0.690</td>
</tr>
<tr>
<td>Engaging</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The robot was expressive.</td>
<td></td>
<td>3.11</td>
<td>1.91</td>
<td>4</td>
<td>1-5</td>
<td>9</td>
<td>-0.664</td>
</tr>
<tr>
<td>The robot was enthusiastic.</td>
<td></td>
<td>3.78</td>
<td>1.31</td>
<td>4</td>
<td>1-5</td>
<td>9</td>
<td>-1.444</td>
</tr>
<tr>
<td>The robot was entertaining.</td>
<td></td>
<td>3.67</td>
<td>1.56</td>
<td>4</td>
<td>1-5</td>
<td>9</td>
<td>-1.027</td>
</tr>
<tr>
<td>The robot was motivating.</td>
<td></td>
<td>4.33</td>
<td>1.03</td>
<td>5</td>
<td>1-5</td>
<td>9</td>
<td>-2.420</td>
</tr>
<tr>
<td>The robot was friendly.</td>
<td></td>
<td>4.11</td>
<td>1.29</td>
<td>5</td>
<td>1-5</td>
<td>9</td>
<td>-1.833</td>
</tr>
</tbody>
</table>

Table 4.8 Statistically significant RPI items by condition and agent factors

<table>
<thead>
<tr>
<th>Control</th>
<th>Adaptive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facilitated Learning</td>
<td></td>
</tr>
<tr>
<td>• Made multiplication interesting</td>
<td></td>
</tr>
<tr>
<td>• Kept student’s attention</td>
<td></td>
</tr>
<tr>
<td>• Helped student focus on important information</td>
<td></td>
</tr>
<tr>
<td>• Interesting</td>
<td></td>
</tr>
<tr>
<td>Facilitated Learning</td>
<td></td>
</tr>
<tr>
<td>• Made student think about multiplication more deeply</td>
<td></td>
</tr>
<tr>
<td>• Encouraged students to think about what they were doing</td>
<td></td>
</tr>
<tr>
<td>• Kept student’s attention</td>
<td></td>
</tr>
<tr>
<td>• Showed information effectively (p=.05)</td>
<td></td>
</tr>
<tr>
<td>• Helped student concentrate on lesson</td>
<td></td>
</tr>
<tr>
<td>• Interesting</td>
<td></td>
</tr>
<tr>
<td>• Enjoyable</td>
<td></td>
</tr>
<tr>
<td>Credible</td>
<td></td>
</tr>
<tr>
<td>• Knowledgeable</td>
<td></td>
</tr>
<tr>
<td>• Intelligent</td>
<td></td>
</tr>
<tr>
<td>• Helpful</td>
<td></td>
</tr>
<tr>
<td>• Seemed like a teacher</td>
<td></td>
</tr>
<tr>
<td>Credible</td>
<td></td>
</tr>
<tr>
<td>• Knowledgeable (p=.05)</td>
<td></td>
</tr>
<tr>
<td>• Useful</td>
<td></td>
</tr>
<tr>
<td>• Seemed like a teacher</td>
<td></td>
</tr>
</tbody>
</table>

51
4.4 INTERVIEW FINDINGS

The results of the student interview were helpful in further understanding students’ perceptions and attitudes about Ms. An.

_HRI._ The HRI questions addressed how students felt about their interactions with Ms. An. All students in both groups responded that they liked Ms. An because she helped with math, and students mostly described her as nice. Other adjectives that were used were cool, fun, and kind. All words used to describe Ms. An were positive. The top responses for students in the control group about their feelings about themselves during the interaction were joyful and grateful. The top response in the adaptive group was smart. When asked “How did Ms. An help you? How would you describe what Ms. An did in the session?”, both groups indicated that Ms. An did everything. Students in the control group also responded that Ms. An gave feedback as to if they answered a question correctly or incorrectly. Student in the adaptive group responded that in addition to feedback, Ms. An showed students the steps to solve the mathematics problems. The top subjects proposed in the adaptive group were English and language arts (ELA) and science. The top subject proposed in the control group was social studies. Students in both groups wanted to work with Ms. An frequently, both once a week and every day.

Table 4.9 highlights the keywords and phrases mentioned during the interview and example quotes.
Table 4.9 Key findings from interview (HRI)

<table>
<thead>
<tr>
<th>HRI</th>
<th>Control</th>
<th>Adaptive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feelings about Ms. An</td>
<td>Helpful</td>
<td>Helpful</td>
</tr>
<tr>
<td></td>
<td>“-wanted to help me get better at what I was doing”</td>
<td>“She made me a better math student and always helped me if I don't get the questions.”</td>
</tr>
<tr>
<td></td>
<td>“She helped me understand and also was very nice. She was helpful and felt like a great teacher.”</td>
<td>“-went step by step with the math and helped me with what I did wrong and helped me to get it right”</td>
</tr>
<tr>
<td></td>
<td>“She is awesome.”</td>
<td>“She helped me with my work. I know Ms. An can help more kids.”</td>
</tr>
<tr>
<td>Feelings about self during interaction</td>
<td>Joyful</td>
<td>Smart</td>
</tr>
<tr>
<td>Type of help provided</td>
<td>She did everything</td>
<td>She did everything</td>
</tr>
<tr>
<td></td>
<td>Gave feedback</td>
<td>Showed step-by-step</td>
</tr>
<tr>
<td></td>
<td>“She asked how I was feeling and told me if I answer right or wrong. She did everything”</td>
<td>Gave feedback</td>
</tr>
<tr>
<td></td>
<td>“Ms. An told me if I got a question right or wrong. She did everything.”</td>
<td>“Ms. An helped with everything that my math teacher is helping me with, she did everything I needed.”</td>
</tr>
<tr>
<td>Other subjects to study with Ms. An</td>
<td>Social Studies</td>
<td>ELA</td>
</tr>
<tr>
<td></td>
<td>“I think it would be awesome if she could help me with other subjects. I think it would help me a lot – She made it a lot fun and a little easier.”</td>
<td>“I need help on hard words and Ms. An can help me spell some words I don't know how to spell. If she knows how to do math, she might can help me with ELA because I think Ms. An is very intelligent.”</td>
</tr>
<tr>
<td>Frequency to work</td>
<td>Once a week</td>
<td>Once a week</td>
</tr>
</tbody>
</table>
with Ms. An Every day Every day

“I would like to study with Ms. An every day.”

“I would like for Ms. An to help me at home every day because she can help me get my grades up.”

Education. The education questions addressed how well the students felt they learned something from Ms. An. Students in both groups agreed that Ms. An helped them learn math. In both groups, students studied most with their teachers and parents. In the adaptive group, students preferred to study with Ms. An over other types of study support. In the control group, students equally preferred to study with Ms. An and their parents over other types of study support. Table 4.10 highlights the keywords and phrases mentioned during the interview and example quotes.

Table 4.10 Key findings from interview (Education)

<table>
<thead>
<tr>
<th>Education</th>
<th>Control</th>
<th>Adaptive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic support</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Current study support</td>
<td>Teacher Parents</td>
<td>Teacher Parents</td>
</tr>
<tr>
<td></td>
<td>“I don’t like to study with my classmates because they are too playful.”</td>
<td>“When I study with my teacher, she helps me a lot.”</td>
</tr>
<tr>
<td>Preferred study support</td>
<td>Ms. An Parents</td>
<td>Ms. An Parents</td>
</tr>
<tr>
<td></td>
<td>“I would like for Ms. An to help me in school because that is where I do most of my work. I think she will help me do better.”</td>
<td></td>
</tr>
</tbody>
</table>
AI. The artificial intelligence questions addressed how well the students felt that Ms. An. acted intelligently and offered personalized support. Student mostly used the word smart to describe intelligence. Most students also indicated that their parents were intelligent and that Ms. An was too. Students in the control group expressed that Ms. An was caring and asked how they were feeling. Students in the adaptive group felt that Ms. An respected their feelings and that Ms. An was concerned with them answering problems correctly. Table 4.11 highlights the keywords and phrases mentioned during the interview and example quotes.

Table 4.11 Key findings from interview (Artificial Intelligence)

<table>
<thead>
<tr>
<th>AI</th>
<th>Control</th>
<th>Adaptive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description of intelligence</td>
<td>Very smart Parents</td>
<td>Very smart Parents</td>
</tr>
<tr>
<td>“I do think Ms. An Intelligent because she helps kids with math.”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Ms. An is intelligent because she helps me. She is very nice and smart.”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personalization</td>
<td>Caring</td>
<td>Respectful</td>
</tr>
<tr>
<td>“She cared about me. She made me feel good by saying that I did good.”</td>
<td>“Ms. An respected me and my feelings. She gave me one-on-one time in math.”</td>
<td></td>
</tr>
<tr>
<td>“I think Ms. An cared about how I felt because she asked me how I felt. I was very shy at first because I never saw a robot before. Then, I became comfortable because she was very nice.”</td>
<td>“She respected my feelings.”</td>
<td></td>
</tr>
<tr>
<td>“She cared about me because every time I got problem wrong, she helped me get it right.”</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

55
CHAPTER 5
DISCUSSION

This study investigated the educational impact of a social robot tutor by investigating adaptive strategies to measure and maintain/increase student engagement while practicing multiplication. The novelty of this multi-disciplinary project is the combination of the fields of HRI, AI, and education.

Based on the research questions, we developed and tested the following hypotheses:

1) a socially adaptive robot tutor will be able to predict a student’s emotion and performance; 2) a socially adaptive robot tutor will engage students academically and emotionally as they practice multiplication; and 3) students will have positive perceptions of the socially adaptive robot tutor. From the results of the research study, we found that

1) the decision tree model did not accurately classify student emotion and performance; 2) the socially adaptive robot tutor successfully engaged students emotionally and academically as they practiced multiplication; and 3) students had positive perceptions of the socially adaptive robot tutor.

5.0 DECISION TREE MODEL

[Q1] How well can a decision tree model classify a student’s emotion and performance? Our decision tree models for emotion and performance did not perform as expected and were not as predictive as we had hoped. The emotion classifier did not do well for individual emotions or valence. Percent correct and gender were the two factors used to predict emotion. On several occasions throughout their interactions with Ms. An,
students expressed being happy; however, they were classified as neutral. Upon review of the video data, the students did in fact appear neutral in facial affect, as shown in Figure 5.1. Human emotions are often subtle, especially during an education task [74] which would understandably render a neutral classification although students indicated happy. In fact, our previous studies have shown that students are mostly positive when interacting with the robot [59], [60], [61]. Thus, the model to predict emotion may not be necessary for mathematics learning tasks.

![Figure 5.1 Images of students who self-reported happy emotion](image)

Similarly, the performance classifier was inexact. Correct performance was only correctly classified a little over half the time, and incorrect performance was correctly classified slightly better than correct performance. Percent correct was the factor used to predict performance. Although, previous studies produced better outcomes for personalization [56], [75], these models were not applied in a mathematics education context. Thus, our model was the first of its kind and has provided important future insights as we move forward in this exploratory area of research.

As we consider this huge design implication for future work, it is also important to explore the features that we used to develop the decision trees. The training data contained a low variability of features as it mostly consisted of non-negative emotions. This is not surprising as we used data from similar studies with students practicing multiplication with Ms. An [59], [60], [61] where most students had positive interactions.
with the robot. Further, the features may not be very predictive of the results. Conducting a user study with educators to better understand what factors they consider when evaluating a student’s state will help us uncover better suited features that could be used in the model.

5.1 ENGAGEMENT

**[Q2] How well can a socially adaptive robot tutor engage 5th grade students to practice multiplication?** a) How do students perform academically by studying with a socially adaptive robot tutor? Per the results, the adaptive strategies employed to increase academic engagement were effective. Due to the faulty decision tree model, the adaptive behavior occurred randomly; however, it occurred most of the time and made an educational difference for the students. Students in the adaptive group had higher significantly different learning gains in post-test 1. Furthermore, their scores tended to be higher in session 2 than the control group. Previous research has also shown that adaptive robots are capable of helping students achieve cognitive gains [39], [53].

_Multiplication problem types._ Students in the adaptive group showed higher gains for identifying the different ways to represent multiplication problems in pre-/post-test 1. Before the intervention, students in both groups did not do well understanding representation of multiplication problems, as indicated by pre-test 1. Based on student work during study session, students in the control group tried to solve various problem types as computation problems despite them being context or pictorial. Figure 5.2 shows examples of this. Students in the adaptive group received additional instruction and practiced examples on solving all multiplication problem types; therefore, students in the adaptive group never tried to apply computation for context or pictorial solutions.
A large van can seat 15 people. How many people can 14 large vans seat?

![Diagram showing different answer options and a calculation example.]

Figure 5.2 Examples of student work using wrong answer type
Solving multiplication problems. Pre-/post-test 2 and the study sessions addressed students’ ability to correctly solve multiplication problems. We expected that students in the adaptive group would show a higher increase than students in the control group for test 2, but students in both groups had similar gains. However, students in the adaptive group showed slightly higher gains for correctly solving multiplication problems during the study sessions (although not statistically significant). Per the session results, we saw that the adaptive strategies helped students solve multiplication problems; however, learning gains for the adaptive group were not reflected in test 2. This may be a consequence of social facilitation. Social facilitation describes the concept in which individuals perform familiar tasks better in the presence of others than if they were alone [75]. In this case, students performed better in the presence of the robot (during sessions) than without the robot (during pre-/post-test 2).

b) How do students respond emotionally by studying with a socially adaptive robot tutor? Per the results, the adaptive strategies employed to increase emotional engagement may have been effective. This is consistent with previous research that has shown that adaptive robot tutors that respond to student emotion, contribute to positive outcomes in education [54]. Students showed a slight trend to report happiness in the adaptive condition (.56) more than the control condition (.47). While this difference was not statistically significant, this slight trend suggests promising results in future studies. Overall, students in both groups mostly exhibited positive emotions. It is likely that we did not see any differences between the two groups because of the novelty of the robot for students. Students in both groups indicated low experience with robots (M=2.95, SD=1.28). Thus, all students were
excited to interact with the new technology. More, longer-term interactions will reduce novelty and might allow us to see a greater change in emotions between the two groups.

5.2 ROBOT SOCIABILITY

[Q3] **What social perceptions do students have of a socially adaptive robot tutor while practicing multiplication?** Consistent with our previous work [59], [60], [61] and other research [55], our findings show that robot sociability elicits positive social perceptions. Students in both conditions considered the robot social. However, the nature of how they were perceived sociability differed between conditions. For example, the adaptive conditions yielded a higher number of RPI items for facilitated learning that were statistically different than neutral (compared to the control condition). Thus, while both robots were social, this data suggests that the adaptive condition may be perceived as a better teacher, by making students think about multiplication more deeply, keeping their interest, show information effectively, helping students pay attention, and concentrate on the material.

Students in the adaptive group spent more time before answering problems than students in the control group. This might suggest that students were indeed thinking more about the mathematics problems they were trying to solve, as indicated by the RPI results.

5.3 DESIGN RECOMMENDATIONS

The results of this study provided insight in students’ interaction with Ms. An and their social perceptions of the robot. By identifying the factors that contributed to student engagement and positive perceptions, we can develop preliminary recommendations for
the design of a socially adaptive robot tutor for mathematics education. Table 5.1

provides a detailed overview of key findings and design recommendations.

Table 5.1 Key findings and design recommendations

<table>
<thead>
<tr>
<th>Key finding</th>
<th>Design recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Decision tree model</strong></td>
<td></td>
</tr>
<tr>
<td>The data from the current study that was used to build the performance tree produced a tree with higher predictability than the training data used from the previous study [61].</td>
<td>Use data from current study as training set to build new performance tree for improved predictability.</td>
</tr>
<tr>
<td>Students are not particularly expressive when practicing math and emotions will likely always be predicted as neutral. Thus, emotion prediction was not necessary for this application.</td>
<td>Consider emotion prediction for other educational tasks.</td>
</tr>
<tr>
<td><strong>Engagement</strong></td>
<td></td>
</tr>
<tr>
<td>The supplemental instruction was effective and made students think more deeply about their work; however, this altered the nature of the students’ interaction with the robot and also changed their perception.</td>
<td>Since using adaptive strategies to enhance help seeking behavior is effective [44], employ adaptive strategies in three tiers: 1) Offer help only when requested by the student; 2) Prompt to help if incorrect performance is predicted; and 3) Force help if prompts are ignored and performance continues to decline.</td>
</tr>
<tr>
<td>The subtle display of emotions was often classified as neutral; however, students will likely be happy when studying with the robot.</td>
<td>Mimicking emotions for math is not the best approach. Instead, the robot should always be expressive and upbeat to ensure students have a positive learning experience, which will result in higher emotional and academic engagement as well as higher social perceptions.</td>
</tr>
<tr>
<td><strong>Sociability</strong></td>
<td></td>
</tr>
<tr>
<td>Personalized instruction promoted higher level thinking among students; however, it also changed students’ perceptions of the interaction from fun to serious.</td>
<td>Enhance personalized instruction to maintain higher level thinking but include social factors to ensure that the experience remains enjoyable.</td>
</tr>
</tbody>
</table>

*Future directions.* This study introduced many exciting new questions in this research area. This work was somewhat exploratory as we investigated an approach that has never
been used. While the data revealed promising trends, future work should include a larger sample size and longer-term interactions for better statistically significant outcomes.

As previously mentioned, educators use their intuition to best decide what a student needs. Educators are able to make predictions about a student’s state and decision tree models will allow a robot tutor to make predictions about a student’s state as well. Future work should address the use of more features that may be more predictive for emotion and performance. This work should include an extensive user study with educators’ feedback to uncover features that they use to predict student emotion and performance.

This study focused on rural students and their engagement with a socially adaptive robot tutor in mathematics education when practicing multiplication. Future work that includes urban students will expand the student demographic of the study and offer deeper analysis across different groups of students. Future studies should address additional mathematics topics and other academic subjects outside of math, which will allow for correlations between student engagement and different topics/subjects.

Future studies should use more direct strategies for error analysis to guide instruction by focusing on specific answer types that students are struggling with. Instruction will contain information that solely relates to the answer type and the specific errors that the student makes.

In this recent study, the robot tutor checked student answers for correctness and provided instruction when needed; however, future studies should include error analysis techniques so that the robot can provide more direct, personalized instruction, and feedback for student’s responses. We captured student work in this study as a preliminary step to use this raw input in a meaningful way. Next steps should allow for open ended
answers and conduct error analysis on student response. We will develop a system that analyzes the written values and determines if the input is correct or incorrect. If incorrect, the system will determine the common misconception that the student made. Figures 5.3-5.5 show examples of students’ scratch work while solving problems and the error they made.

Figure 5.3 Student work and errors made (No error)

A jeweler made 123 diamonds necklaces. She used 12 diamonds to make each necklace. How many diamonds did the jeweler use?

Figure 5.4 Student work and errors made (Calculation Error: Times table mistake)
Conclusion. This study investigated the use of a socially adaptive robot tutor to engage students in mathematics education. Often, it is difficult to get students to engage in mathematics education [76]. While technology is not a full solution, it can make significant contributions to better engage students in mathematics education [77]. This study was important because it offered strategies to better engage students (emotionally and academically) in mathematics education.

Although our decision tree models were not very predictive, the results gave answers to our current questions and clarity for future directions. Our adaptive strategies to engage students academically were effective. All students enjoyed working with the robot and we did not see a difference in emotional engagement across the two groups. Our adaptive strategies made students think more deeply about their work and focus more. This higher order thinking is preferred in education as it a cognitive process that demonstrates deeper understanding of the academic material [78].

Not only does this study tell us more about education and AI, but it also tells us how...
to improve the methodology for educational HRI in rural areas. Novelty likely played an important role in this study on a rural population due to lack of exposure for students. Future studies that include urban students may yield different results.

This study offered insight for developing a socially adaptive robot tutor to engage students academically and emotionally while practicing multiplication. Results from this study will inform the human-robot interaction (HRI) and artificial intelligence (AI) communities on best practices and techniques within the scope of this work.
REFERENCES


APPENDIX A – STUDENT DEMOGRAPHIC FORM

1. First Name

2. Last Name

3. Gender
   ☐ Male
   ☐ Female

4. Age
   ☐
   Other (please specify)

5. Grade
   ☐

6. School Name

7. Ethnicity/Race
   ☐ Arab
   ☐ Asian/Pacific Islander
   ☐ African American/Black
   ☐ Caucasian/White
   ☐ Hispanic
   ☐ Indigenous or Aboriginal
   ☐ Latino
   ☐ Multiracial
   ☐ Would rather not say
   ☐ Other (please specify)
APPENDIX B – PRE-/POST-TEST 1

Directions: Answer the following questions. Be sure that your answers correspond to the appropriate letters in each question.

Select all of the ways to correctly represent the problem.

1. **587 x 3**

A bike shop has 587 customers each day. How many customers does the bike shop have in 3 days?

2. **239 x 4**

4 friends each have 239 dollars. How many dollars do the friends have all together?
3. **762 x 8**

Tina worked 762 hours a month. How many hours does she work in 8 months?

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>762</td>
<td>762</td>
<td>762</td>
<td>762</td>
</tr>
<tr>
<td>762</td>
<td>762</td>
<td>762</td>
<td>762</td>
</tr>
<tr>
<td>762</td>
<td>762</td>
<td>762</td>
<td>762</td>
</tr>
<tr>
<td>x 8</td>
<td>x 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6096</td>
<td>36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>120</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ 800</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>956</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX C – PRE-/POST-TEST 2

Directions: Answer the following questions. Be sure that your answers correspond to the appropriate letters in each question.

Solve the following problem.

1. **A strawberry farm has 10 rows of strawberries with 2 strawberries in each row. What is the correct representation of the strawberries?**

![Diagram with options A, B, C, D]
2.\[
\begin{array}{c|c|c|c|c}
\text{y} & 1 & 2 & 1 & 2 \\
\text{x} & 3 & 5 & 3 & 5 \\
\hline
6 & 0 & 6 & 0 & 1 & 0 & 0 & 0 & 3 & 6 \\
3 & 6 & 0 & 3 & 6 & 0 & 3 & 6 & 0 \\
\hline
4 & 2 & 0 & 3 & 2 & 0 & 4 & 6 & 0 & 9 & 6 \\
\end{array}
\]

A | B | C | D
\hline

3. Jacob rode his bike for 7 miles in one week. Thomas rode his bike 16 times more than Jacob in that same week. How many miles did the two ride their bikes in total?

Jacob rode his bike for 17 miles in one week. Thomas rode his bike 7 times more than Jacob in that same week. How many miles did the two ride their bikes in total?

Jacob rode his bike for 16 miles in one week. Thomas rode his bike 7 times more than Jacob in that same week. How many miles did the two ride their bikes in total?

Jacob rode his bike for 16 miles in one week. Thomas rode his bike 9 times more than Jacob in that same week. How many miles did the two ride their bikes in total?

| A | B | C | D |
\hline
4. 

\[
\begin{array}{cccc}
17 & 17 & 17 & 17 \\
\times 10 & \times 10 & \times 10 & \times 10 \\
00 & 10 & 00 & 00 \\
+ 170 & + 70 & + 170 & \\
270 & 70 & 170 & \\
\end{array}
\]

5. Julie has 12 marbles. Max has 15 times as many marbles as Julie. How many marbles does Julie have?

\[
\begin{array}{cccc}
1 & 1 & 1 & 1 \\
15 & 15 & 15 & 15 \\
\times 12 & \times 12 & \times 12 & \times 12 \\
30 & 30 & 40 & 34 \\
+ 15 & + 150 & + 150 & + 150 \\
45 & 180 & 190 & 184 \\
\end{array}
\]
6. The bowling team, Strikers, scored 753 points combined. Another team, Rangers, scored 14 times more than team Strikers. How many points did team Rangers score?

\[
\begin{array}{cccc}
2 & 1 & 2 & 1 \\
7 & 5 & 3 & 7 & 5 & 3 & 7 & 5 & 3 \\
x & \times & 1 & 4 & x & \times & 1 & 4 & x & \times & 1 & 4 \\
3 & 0 & 1 & 2 & 3 & 0 & 1 & 2 & 3 & 0 & 1 & 1 \\
+7 & 5 & 3 & 0 & +7 & 5 & 3 & 0 & +7 & 5 & 3 & 0 \\
1 & 0 & 5 & 4 & 2 & 3 & 7 & 6 & 5 & 1 & 0 & 5 & 4 & 1 & 1 & 0 & 5 & 2 & 2 \\
\end{array}
\]

A B C D

7.

A small dairy farm produces 14 gallons of milk in a day. How many gallons of milk will it produce in 87 days?

A small dairy farm produces 88 gallons of milk in a day. How many gallons of milk will it produce in 14 days?

A small dairy farm produces 87 gallons of milk in a day. How many gallons of milk will it produce in 13 days?

A small dairy farm produces 87 gallons of milk in a day. How many gallons of milk will it produce in 14 days?

A B C D
8.

\[
\begin{array}{|c|c|c|}
\hline
7 & 3 \\
\hline
2 & 9 & 4 \\
\times & 2 & 8 \\
2 & 3 & 5 & 8 \\
+ & 1 & 1 & 1 & 8 & 0 \\
1 & 3 & 5 & 3 & 8 \\
\hline
\end{array}
\quad \begin{array}{|c|c|c|}
\hline
7 & 3 \\
\hline
2 & 9 & 4 \\
\times & 2 & 8 \\
3 & 2 & 5 & 2 \\
+ & 5 & 8 & 8 \\
3 & 8 & 4 & 0 \\
\hline
\end{array}
\quad \begin{array}{|c|c|c|}
\hline
7 & 3 \\
\hline
2 & 9 & 4 \\
\times & 2 & 8 \\
2 & 3 & 5 & 0 \\
+ & 5 & 8 & 8 & 0 \\
8 & 2 & 3 & 0 \\
\hline
\end{array}
\quad \begin{array}{|c|c|c|}
\hline
7 & 3 \\
\hline
2 & 9 & 4 \\
\times & 2 & 8 \\
2 & 3 & 5 & 2 \\
+ & 5 & 8 & 8 & 0 \\
8 & 2 & 3 & 2 \\
\hline
\end{array}
\]

9. Horses have four legs. How many legs would 25 horses have?
10. Lucy’s cabbage patch has 12 rows of cabbage. In each row, there are 4 heads of cabbage. How many heads of cabbage does Lucy have in all?

11. Ronald has 6 packs of gum. There are 8 pieces in each pack. How many pieces of gum does Ronald have?
12. A small auditorium has 8 rows. Each row has 13 seats, how many seats are in the auditorium?

A

B

C

D

13.  

\[
\begin{array}{c|c|c|c}
6 & 6 & 6 & 6 \\
39 & 39 & 39 & 39 \\
27 & 27 & 27 & 27 \\
273 & 273 & 275 & 273 \\
+1280 & +780 & +780 & +780 \\
1553 & 1043 & 1055 & 1053 \\
\end{array}
\]

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
</table>

15. A private art gallery sold 18 paintings in one day. The sales averaged 165 dollars per painting. How much money did the art gallery make?

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A private art gallery sold 18 paintings in one day. The sales averaged 165 dollars per painting. How much money did the art gallery make?</td>
<td>A private art gallery sold 162 paintings in one day. The sales averaged 18 dollars per painting. How much money did the art gallery make?</td>
<td>A private art gallery sold 18 paintings in one day. The sales averaged 162 dollars per painting. How much money did the art gallery make?</td>
<td>A private art gallery sold 19 paintings in one day. The sales averaged 162 dollars per painting. How much money did the art gallery make?</td>
</tr>
</tbody>
</table>
16. 

\[
\begin{array}{ccc}
3 & 3 & 3 \\
27 & 27 & 27 \\
\times & 15 & \times & 15 & \times & 15 \\
135 & 135 & 134 & \text{A} & \text{B} & \text{C} \\
+ & 27 & + & 270 & + & 270 \\
162 & 405 & 404 & \text{D} & & \\
\end{array}
\]

17. 

A public library has 54 bookshelves. Each shelf has 136 books, how many books does the library have? 

A public library has 64 bookshelves. Each shelf has 126 books, how many books does the library have? 

A public library has 54 bookshelves. Each shelf has 126 books, how many books does the library have? 

A public library has 126 bookshelves. Each shelf has 54 books, how many books does the library have? 

\[
\begin{array}{cccc}
126 & 126 & 126 & 126 \\
126 & 126 & 126 & 126 \\
126 & 126 & 126 & 126 \\
126 & 126 & 126 & 126 \\
126 & 126 & 126 & 126 \\
126 & 126 & 126 & 126 \\
126 & 126 & 126 & 126 \\
126 & 126 & 126 & 126 \\
\end{array}
\]
18.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>x</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>30</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>+</td>
<td>150</td>
<td>150</td>
<td>150</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>180</td>
<td>170</td>
<td>190</td>
<td>55</td>
</tr>
</tbody>
</table>
APPENDIX D – EMOTIONS QUESTIONNAIRE

Directions: Choose the emoji that best describes how you feel. There is no right or wrong answer.

Happy  Surprised  Neutral  Sad  Angry
APPENDIX E – ROBOT PERSONA INVENTORY

Directions: For each statement, select the choice that best describes how you feel. There is no right or wrong answer.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neither agree or disagree</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>The robot helped me think about multiplication more deeply.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The robot made multiplication interesting.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The robot encouraged me to think about what I was learning.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The robot kept my attention.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The robot showed me the information effectively.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The robot helped me to concentrate on the lesson.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The robot helped me focus on the important information.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The robot improved my knowledge of multiplication.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The robot was interesting.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The robot was enjoyable.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The robot seemed knowledgeable.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The robot seemed intelligent.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>The robot was</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>helpful.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The robot was</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>useful.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The robot seemed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>like a teacher.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The robot had a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>personality.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The robot's emotion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>was natural.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The robot seemed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>like a human.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The robot's movement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>was normal.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The robot showed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>emotion.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The robot was</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>expressive.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The robot was</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>enthusiastic.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The robot was</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>entertaining.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The robot was</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>motivating.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The robot was</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>friendly.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX F – INTERVIEW SCRIPT

Introduction
This interview is for <<say student’s participant id>> This is the final step in the research study. Our goal is to better understand what students think about Ms. an. I’m going to ask you a few questions. This short interview will take about 15 minutes.

Icebreaker/Warm up
• What are some things you enjoy doing?
• What is your favorite subject?

Discussion
Now that you have worked with Ms. An, I would like to talk to you about your experience.

HRI
1. Did you like or dislike Ms. An?
2. Using one word describe how Ms. An made you feel as a student? As a math student?
3. How did Ms. An help you? How would you describe what Ms. An did in the session?
4. What are some other subjects that Ms. An can help you with in school?
5. If you could work with Ms. An would you like to work with her every day, once a week, once a month, once a year, or never?

Education
1. Did Ms. An help you learn math?
   a. If yes, how did she help?
   b. If not, why not?
2. How do you usually study math?
3. Have you studied math in the following ways?
   a. With your parents?
   b. With your classmates?
   c. With your teacher?
   d. Alone?
   e. With a book?
   f. With a computer program?
4. Do you prefer to study math with your parents, with your classmates, with your teacher, alone, with a book, with a computer program, or with Ms. An the most?

AI
1. Was Ms. An intelligent?
2. Did Ms. An adapt to meet your needs?
3. Did Ms. An understand your feelings?
4. Did Ms. An understand how much you knew about math?
**Closing Questions**
- Is there anything else that you would like to tell me about Ms. An?

**Debriefing**
Thank you for your time during this study. Your input will help us develop a robot tutor that is helpful and easy to use. It is very important that you do not discuss this study with anyone else until the study is complete. Thank you again for your participation!