5-2016

An Intelligent Robot and Augmented Reality Instruction System

Christopher M. Reardon
University of Tennessee - Knoxville, creardon@utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Christopher M. Reardon entitled "An Intelligent Robot and Augmented Reality Instruction System." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Computer Science.

Lynne E. Parker, Major Professor

We have read this dissertation and recommend its acceptance:

Bruce J. MacLennan, Bradley Vander Zanden, Mari Beth Coleman, Eric R. Wade

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

An Intelligent Robot and Augmented Reality Instruction System

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Christopher M. Reardon

May 2016

To my sister, Heather Jean Reardon.

Acknowledgements

I would like to thank my advisor, Lynne E. Parker, who has guided me through many, many years of study, and whose encouragement and direction shaped my professional life and future for the better. I thank my family, my friends, and my colleagues at the Distributed Intelligence Lab, for their support through this challenging endeavor. I also would like to thank the faculty, staff, and most especially the wonderful students at the FUTURE program, whose willingness and enthusiasm is endlessly inspiring. Finally, I must thank Rachel Elizabeth Wright, with whom the majority of this interdisciplinary research was conducted in collaboration, particularly for her expertise in education techniques and experiment design, and without whom I believe this research would not have been successful.

Abstract

Human-Centered Robotics (HCR) is a research area that focuses on how robots can empower people to live safer, simpler, and more independent lives. In this dissertation, I present a combination of two technologies to deliver human-centric solutions to an important population. The first nascent area that I investigate is the creation of an Intelligent Robot Instructor (IRI) as a learning and instruction tool for human pupils. The second technology is the use of augmented reality (AR) to create an Augmented Reality Instruction (ARI) system to provide instruction via a wearable interface.

To function in an intelligent and context-aware manner, both systems require the ability to reason about their perception of the environment and make appropriate decisions. In this work, I construct a novel formulation of several education methodologies, particularly those known as response prompting, as part of a cognitive framework to create a system for intelligent instruction, and compare these methodologies in the context of intelligent decision making using both technologies.

The IRI system is demonstrated through experiments with a humanoid robot that uses object recognition and localization for perception and interacts with students through speech, gestures, and object interaction. The ARI system uses augmented reality, computer vision, and machine learning methods to create an intelligent, contextually aware instructional system. By using AR to teach prerequisite skills that lend themselves well to visual, augmented reality instruction prior to a robot instructor teaching skills that lend themselves to embodied interaction, I am able to

demonstrate the potential of each system independently as well as in combination to facilitate students' learning.

I identify people with intellectual and developmental disabilities (I/DD) as a particularly significant use case and show that IRI and ARI systems can help fulfill the compelling need to develop tools and strategies for people with I/DD.

I present results that demonstrate both systems can be used independently by students with I/DD to quickly and easily acquire the skills required for performance of relevant vocational tasks. This is the first successful real-world application of response-prompting for decision making in a robotic and augmented reality intelligent instruction system.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

In this chapter I introduce the background and motivation of this research (Section 1.1), state the problem and discuss challenges and scope (Section 1.2), overview the approach I take to address the problem (Section 1.3), and enumerate the contributions of this work (Section 1.4).

1.1 Background and motivation

1.1.1 Robots and learning

In robot-learning applications, an active research area involves humans teaching robots to perform tasks. For this research, I explore the complement of this research area: intelligent robots teaching humans. This research area is in extremely nascent stages but has a high potential for returns. Previous works involving robots teaching humans are generally focused on two specific areas: 1) therapy robots for children with autism, where these children can sometimes maintain better or longer interactions with robots than human therapists (Weir and Emanuel (1976); Michaud and Théberge-Turmel (2002); Robins et al. (2009); Billard et al. (2007); Scassellati et al. (2012); Dautenhahn and Werry (2004); Feil-Seifer and Mataric (2011)), and 2) more recently, "classroom motivation" robots which typically leverage the novelty of a robot in an interactive system to encourage children to perform a task, e.g., to exercise (Fridin (2014b); Kose and Yorganci (2011); Addo et al. (2013); Lee and Kim (2010); Howard et al. (2012)).

A strong case can be made for the benefits of using intelligent robot instructors (IRIs) to teach human pupils. The more time a teacher spends with a student, the better the student learns. In classroom settings, therefore, if an IRI were capable of assisting a human teacher by providing instruction to students, it could offload some of the tasks of the teacher, thereby increasing the amount of time available for the teacher to spend with individual students. In the face of future teacher shortages and increasing classroom sizes (Watlington et al. (2010); Wilkin and Nwoke (2011), Mckeown (2012)), the ability of an IRI to augment a human instructor's teaching could allow for better use of limited (human) teaching resources.

Robots have several strengths that can be leveraged in an instructor role. A robot is tireless, and a well-engineered robot could have nearly unlimited energy and attention to direct at students. The precision of a robot would enable it to provide perfectly timely instructions, and would avoid issues such as over-prompting and inappropriate fading that human instructors face. A robot's perception is only limited by its sensors and computing capabilities, meaning one robot could potentially observe and instruct large numbers of students simultaneously. Pupils could perceive robots as less judgmental than a human, and therefore would be less reluctant and more likely to request repeat instruction (e.g., ask the question again) until a lesson was fully learned. To youth already comfortable with using technology to learn, a robot could represent an embodied and more physically-interactive tool for learning than a personal computer or mobile device. Indeed, several studies have shown that embodiment is beneficial for human interaction with intelligent systems (Leyzberg et al. (2012); Bainbridge et al. (2008); Krämer and Bente (2010); Kidd and Breazeal (2004); Wainer et al. (2007); Tapus et al. (2009); Kiesler et al. (2008)).

1.1.2 Social validity and life skills

One area of potential intelligent instruction is the teaching of socially valid life skills. The term *social validity* from the Applied Behavior Analysis field refers to the acceptability of the goals, procedures, and outcomes of instruction or treatment (Foster and Mash (1999)). In the context of instruction, it refers to knowledge or skills that increase a person's independence of personal, community, or job life. It is worth noting that true assessment of social validity is a non-trivial task; social validity should be assessed on a multidimensional gradient, and no well-established criteria exist for binary classification of what is and is not socially valid.

The term *life skills* in the context of education refers to the ability of a person to perform problem solving behaviors to manage his or her daily personal life. UNICEF, which has programs dedicated to life skills education (UNICEF (2016)), defines life skills education as "Education that helps young people develop critical thinking and problem solving skills, that builds their sense of personal worth and agency, and teaches them to interact with others constructively and effectively, has transformative potential."

The World Health Organization (Gillespie et al. (2003)), which conducts similar programs, states that "life skills are abilities for adaptive and positive behaviour that enable individuals to deal effectively with the demands and challenges of everyday life.... In particular, life skills are a group of psychosocial competencies and interpersonal skills that help people make informed decisions, solve problems, think critically and creatively, communicate effectively, build healthy relationships, empathise with others, and cope with and manage their lives in a healthy and productive manner. Life skills may be directed toward personal actions or actions toward others, as well as toward actions to change the surrounding environment to make it conducive to health." The WHO concludes that life skills are an essential and important component of an effective education system.

Life skills taught to K-12 students (Morford et al. (2006)) are diverse, and include domains such as communication, social, self-management, home living, community access, vocational, and functional academic skills. Particularly relevant to educational experiments conducted for this work are home living, vocational, and functional academic domain task skills, for example those necessary for performing cleaning, assembly, and cooking tasks that one would encounter in daily life.

Furthering the motivation for intelligent robot instruction is a potential demand for robot instructors. While supplies of available teachers in the United States vary by locality, specialty, and demographics (Watlington et al. (2010); Wilkin and Nwoke (2011)), all across the U.S. there is currently a chronic special education teacher shortage (Boe (2014); Howard et al. (2013)). It is also recognized that worldwide, there is a global shortage of teachers (Mckeown (2012)).

Indeed, students with intellectual and developmental disabilities (I/DD) have a high potential benefit from IRIs. The unique abilities of robots are particularly suited for several reasons. The degree of repetition required by students with disabilities, the preferred instruction methodologies used, and the benefits of individual time with an instructor could all be well addressed by such a system. The types of tasks commonly taught (e.g., life skills) lend themselves well to robot demonstration and observation. These reasons, coupled with the shortage of human teachers, creates a high potential for use of intelligent instruction systems to assist in educational scenarios.

In addition to these motivations, by examining the progress being made in areas of machine learning and robot learning by demonstration, it is very possible to imagine a future where IRIs, for example, form a bridge between expert human teachers and pupils, at least for some purposes. In such an application, human teachers could instruct robots, who then take those lessons learned, apply an intelligent cognitive system to them, and in turn teach human pupils, as illustrated in Figure 1.1. This relationship would benefit from an extremely powerful ability to scale at the robot level, as any knowledge representation learned and teaching ability programmed onto one robot could be scaled without limit, even using techniques to adapt knowledge

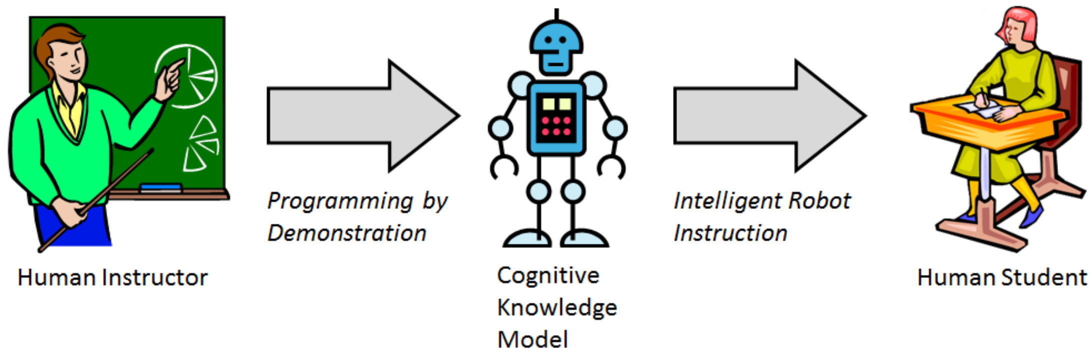


Figure 1.1: Knowledge transference from human instructor, to robot, to human student.

across heterogeneous robots (Zhang et al. (2015)). Thus, via IRIs, one human expert could teach an unlimited number of human pupils, who would benefit from the advantages of those robot instructors (Figure 1.2).

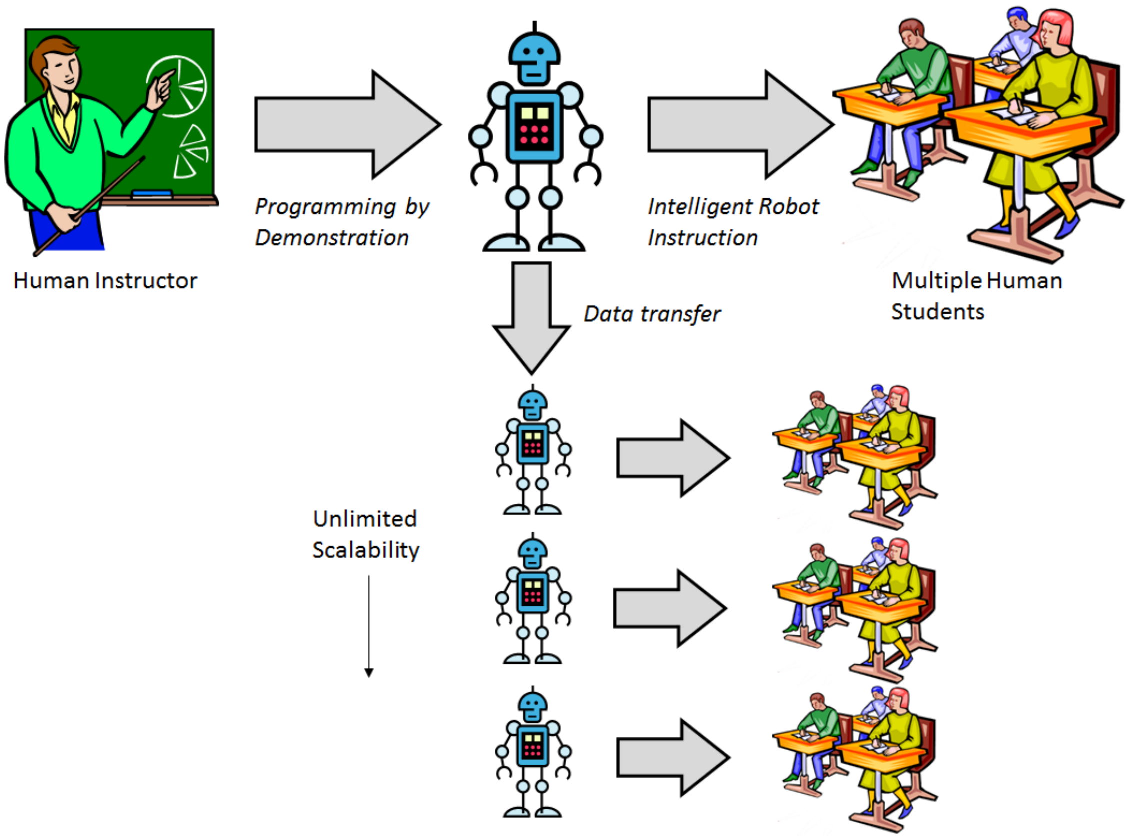


Figure 1.2: Scaling knowledge transference to multiple students and multiple sets of robot instructors and students.

1.1.3 Context-aware augmented reality

Today's young people with I/DD face harsh realities as they enter enter adulthood, such as low employment rates, poor wages and benefits, limited community supports, and low rates of independent living (Grigal and Hart (2010)). There is therefore a strong motivation to provide support for this population to increase their independence in performance of vocational tasks. Assistance provided to help an individual perform a given skill include modeling via job coaches, or printed or digitized materials. Because success in any job requires the performance of multiple skills, persons with I/DD can quickly acquire a plethora of such supports. Though portability of these supports is increased as they are provided through digital storage and mobile devices, the ability to quickly retrieve contextually correct support is extremely challenging, in that it remains reliant on the organizational abilities of the user, a deficit in which is inherently associated with I/DD. Therefore, additional assistance is still needed to ensure that the individual will learn to perform the skill steps correctly and with the greatest level of independence possible (Lancioni and O'Reilly (2001)). Ironically, technology-based approaches are often completely dependent on another person to set up and manage the content, as well as initiate all operations of the support devices. While these types of supports can be successful in training situations, because their effectiveness is contingent on the actions of another person, this creates user dependence, rather than independence, resulting in frustration, disillusionment, and device abandonment.

Augmented reality allows the user to perceive and interact with the real world while simultaneously receiving additional information that is virtualized into their field of perception and has great potential in education (Bower et al. (2014)). AR and wearables eliminate the number of steps needed to access information and resolve the deficiencies in the skills required for basic operation of devices. Of the limited studies involving augmented reality and students with I/DD, all have shown highly successful results, and are discussed further in Section 2.5.

One function of augmented reality is to provide context awareness, which is the ability to provide information appropriate to the situation. The motivating strength of context awareness is precision: it enables the correct information, and only the correct information, to be displayed at the appropriate time. Off-the-shelf context-aware applications are available (e.g., Aurasma (2016); Layar (2016); Junaio (2016); BuildAR (2016)); however, these approaches and research involving context awareness typically uses tags to provide the context (Section 2.5), which is impractical and cumbersome in real-world situations.

An intelligent ARI system that is able to extract the context from the scene without the use of visual codes, would further unencumber the student from dependency on support persons and technology and increase the real-world generalizability of the learning process.

1.2 Problem statement, challenges, and scope

In the face of a worldwide teacher shortage, and a critical shortage of special education teachers in the US, there is an urgent demand for educational resources. For people with I/DD in particular, there is a compelling need to develop tools and strategies in order to facilitate independence, self sufficiency, and address poor employment outcomes in adulthood. The purpose of this research is to address this problem by constructing an intelligent robotic and augmented reality system capable of providing instruction to human pupils, particularly those with I/DD, to teach socially valid life skills.

To maintain an appropriate scope for this research, restrictions must be placed on what is being taught, as shown in Table 1.1. The tasks must be challenging enough that they are not intuitive to the learning audience, be simple enough that they can be taught by the robot with the time and resources available, and be representative enough that we can with confidence trust the generalizability of the results.

Table 1.1: Instructional task criteria

Skills to be taught must:	
1	Be observable by the system.
2a	Be performable by the robot, or,
2b	Be conducive to reality augmentation.
3	Have a non-intuitive solution.
4	Be complex but not an expert skill.
5	Have socially valid benefits to the student.

The tasks being taught must be within the system's capability to both perform and observe what is being performed. The purpose of this research is neither to create highly advanced perception systems nor highly adept manipulation hardware. Therefore, a system for interaction that is capable of interaction with the user and the environment in a manner that is efficient and interpretable is necessary. However, as no such system was available at the outset of this research, it was necessary to create one.

In addition to being able to perceive and interact with the environment, the intelligent instruction system must be able to reason upon the problem, observe the student's performance, and provide appropriate feedback. Because this instruction is delivered to human participants, it is essential that the actions taken by the system be appropriate; therefore, the methods used to select those actions should be proven effective. It would also be very beneficial if this approach would be generalizable across technological platforms. A generalizable, algorithmic approach to the cognitive component of the intelligent instruction system for appropriate action selection is therefore the cornerstone component of this system.

1.3 Approach overview

For this research, I create a robot and augmented reality instruction system to teach socially valid life skills to humans, with a focus on providing instruction to students with I/DD. The goal is to demonstrate that an intelligent instruction system can teach a human such a skill, as well as to examine some of the issues of perception, cognition, knowledge modeling, and human-robot interaction that are important to this domain.

The approach is divided into three main components: (1) Perception, of humans and objects; (2) Cognition, for teaching and evaluation; and (3) Interaction, for instruction, movement, and manipulation as depicted in Figure 3.1.

Component (1) is a combination of object tracking, human perception, and speech recognition (Section 3.2). Component (2), for instructional methodology, leverages response prompting instructional strategies from the Applied Behavior Analysis approaches of the education domain, and creates a cognitive system to teach and evaluate within that framework (Section 3.3). Finally, component (3) includes the methods for providing interactive feedback to the user in interaction scenarios, including speech and gestures for the IRI, and augmented images and speech for the ARI system (Section 3.4).

To limit the scope of this research, it is observed that because the goal of this research is to create a system with a relatively high-level purpose, it must leverage several components that in themselves contain significant large and valid research areas. Because the goal is not to directly advance those areas, pre-existing software and standard approaches are used when available. Despite that, there are significant gaps that must be addressed, so software has been created as necessary to address these gaps.

1.4 Contributions

The fundamental contributions of this research include:

- **The formulation of response prompting for cognitive decision making in real-world, automated, intelligent systems.** I employ response prompting approaches as an modular algorithmic component for decision making and action selection as part of a cognitive system for instruction in systems functioning in the real world.

- **The combination of augmented reality to teach prerequisite skills for more advanced instruction given by an intelligent robot instructor.** I bridge the IRI and ARI systems by using augmented reality to teach prerequisite skills that lend themselves well to visual, augmented reality instruction, prior to a robot instructor teaching skills that lend themselves to embodied instruction and interaction.

- **The comparison of multiple varieties of prompting in the context of intelligent systems.** There are multiple prompting approaches. I implement approaches that include the System of Least Prompts and System of Most Prompts approaches with forward and backward chained tasks as well as non-sequential tasks, and compare the applicability of these approaches to IRI and ARI systems' ability to teach the type of skills selected.

Contributions that demonstrate a broader impact through application include:

- **The application of response prompting for decision making on an intelligent robot instructor.** I create an Intelligent Robot Instructor capable of instructing students in relevant vocational skills.

- **The demonstration of generalizability of response prompting through the creation of an augmented reality instruction system.** By using prompting in an augmented reality system, I show that my approach is generalizable to other intelligent, interactive real-world systems.

- **The use of augmented reality, computer vision, and machine learning to teach vocational tasks.** I show that augmented reality, combined with computer vision and machine learning methods, can be used to teach vocational tasks.

- **The creation of an object detection and tracking system for the purpose of instruction.** In order to address the perception component of an intelligent cognitive system, I create software capable of detecting objects, using known object information (such as color and shape) or image classification via supervised learning.

- **Experimental results proving instructional success teaching students with I/DD using appropriate experimental design and evaluation methods.** All approaches are validated through successful experiments teaching students with I/DD. Single-Case Experimental Design (discussed in Section 4.6.2) is used to control for validity, giving strong confidence in the experimental outcomes.

1.5 Summary

Teaching students with I/DD is a compelling opportunity for intelligent systems, using both augmented reality and robots. The remainder of this work presents a system that accomplishes this. Chapter 2 conducts a review of relevant literature. In Chapter 3, the approach to perception, cognition, and interaction is detailed. Chapter 4 contains a full description of the experiments conducted, the system created, and the methods used to evaluate the outcomes. Chapter 5 presents the results of the experiments, and in Chapter 6 those results, observations, and future directions are discussed. Chapter 7 concludes the work.

Chapter 2

Literature review

There is a rich history of robots learning from humans (Section 2.1), whereas robots as instructors (Section 2.3) and the use of context-aware augmented reality in education (Section 2.5) is more nascent. This chapter explores work in these areas.

2.1 Humans teaching robots

2.1.1 Learning from demonstration

Learning from demonstration (LfD) or programming by demonstration (PbD), is an active research area in intelligent robotics (Argall et al. (2009)) as well as other areas of computer science (Cypher and Halbert (1993)). Robot learning, which can be described as learning the appropriate action for a perceived state, is a particularly crucial component to robot learning.

There are many approaches for learning the policies that map states to actions, but of particular interest to this work are methods that involve robots learning from human instruction, i.e., robot LfD, because robot LfD is complementary to intelligent robots teaching humans. Examples of this supervised learning approach include classification methods (e.g., GMMs, HMM, Baysian approaches), regression methods, reinforcement learning, and plan-based methods (Argall et al. (2009)).

While my approach is inspired by the methodologies of these systems, what it shares is the need for a knowledge representation and cognitive capabilities of these systems. That is, in my case, how should the knowledge to be conveyed via instruction to the student be organized and represented, and how should that conveyance be realized?

In learning, the correspondence problem involves the difficulty of the learner in mapping the teacher's demonstrations to appropriate states and actions. This is particularly challenging in robot LfD because the robot's perception and articulation abilities are dwarfed by a human's. However, by formulating the IRI problem, we avoid most of the correspondence issues inherent in LfD, as it is observed that in a well-constructed interaction, a human will be able to resolve any correspondence mapping from the robot's limited capabilities to the human's much greater capabilities.

2.1.2 Learning about teaching

Motivated by the need to understand how humans teach, Khan et al. (2011) takes a curriculum learning approach that examines teaching strategies, based on a binary classification task in one dimension, by observing the strategies displayed in the sequence of instances selected by the teacher from the teaching set. In experiments, humans use flashcards to teach a humanoid robot graspabilty of objects, rated on a decision scale from not-graspable to graspable. The robot is used to collect the sequence of objects for further analysis, and a framework to explain the observations is presented.

Cakmak et al. (2012) extends an algorithmic teaching approach to sequential tasks, presents an algorithm to select the best demonstration sets to reduce the hypothesis space as quickly as possible, and leverages that idea to provide guidance to human instructors. This work provides a similar opportunity to learn about teaching and examine the decisions that human instructors make in the context of a framework for instruction applied to an intelligent robotic system.

2.2 Cognitive systems

Artificial cognition has its origins in cybernetics, with the intention to create a science of mind based on logic (Varela and Dupuy (1992)). Cognitivism, which models cognition as reasoning, for the purpose of planning and acting, upon knowledge that has been abstracted from perceived information, is the predominant approach to cognition to date (Vernon et al. (2007)). Within the cognitivism paradigm, several cognitive architectures were developed, including Soar (Laird et al. (1987)), ACT-R (Anderson (1996)), C4 (Isla et al. (2001)), and architectures for robotics (Burghart et al. (2005); Benjamin et al. (2004)), which are relatively independent of the task (Gray et al. (1997)). Since architectures represent the fixed part of cognition, they cannot accomplish anything in their own right and need to be provided with knowledge to conduct a specific task. The combination of a cognitive architecture and a particular knowledge set is generally referred to as a *cognitive model* (Vernon et al. (2007)). The knowledge incorporated in cognitive models is typically determined by human designers (Vernon et al. (2007)). This knowledge can be also learned and adapted using machine learning techniques.

Cognitive models have been widely used in human-machine interaction and robotic vision applications. For example, cognitive modeling was adopted in Duric et al. (2002) to construct intelligent human-machine interaction systems. Cognitive perception systems were also used to recognize traffic signs (Yang et al. (2013)), interpret traffic behaviors (Nagel (2004)), and recognize human activities (Crowley (2006)). Over the last decade, probabilistic models of cognition, as an alternative of deterministic cognitive models, have attracted more attention in cognitive development (Xu and Griffiths (2011)). For example, a cognitive vision system was designed in Buxton (2002) to use dynamic decision networks to interpret activities of expert human operators. Another cognitive model was introduced in Town and Sinclair (2003) to apply adaptive Bayesian networks for video analysis.

Probabilistic models have also been widely used for learning and reasoning in cognitive modeling (Chater et al. (2006)).

Cognitive systems for HRI tasks include Feil-Seifer and Mataric (2008), which was developed for robotic HRI interventions with children with Autism Spectrum Disorder, and Cakmak et al. (2010), which uses a cognitive percept-belief system based on Isla et al. (2001) for action selection in a robot learning by demonstration task. The research presented in this dissertation is also inspired by the cognitive system presented in Isla et al. (2001), and is discussed further in Section 3.3.1.

2.3 Robots as instructors or therapists

2.3.1 The importance of embodiment

When it comes to successful interaction with automated agents, the advantages of embodiment have been established by many sources.

Leyzberg et al. (2012) showed that instruction from a physically present robot, compared to instruction from a video of the robot, from an audio recording, and no instruction at all, performed best when providing puzzle-solving advice, as measured by puzzle-solving time and self-report measures. They concluded that "a physically present robot delivering customized tutelage yields higher cognitive gains for a complex (NP-hard) math game than the same instruction provided by a video tutor, an audio-only tutor, a physically present robot giving randomized advice, or no tutor at all." This suggests that robot instructors, by presence alone, can contribute significantly to human performance in complex tasks.

Bainbridge et al. (2008) discovered that physically-present robot's commands are more likely to be obeyed than a video representation of the robot, and showed that physical robots are afforded more of the aspects of a human, such as obedience of unusual instructions and physical personal space. This further suggests that physically present robots could make better instructors than video instruction, and may be regarded more like human instructors.

Fasola and Mataric (2013) found that older adults preferred a physically embodied robot "coach" over a virtual coach for several key social factors, and under robotic coaching performed at a consistently high level.

From a review of social psychology, Krämer and Bente (2010) concludes that the "effectiveness of instructional communications may be improved by augmenting e-learning environments with embodied virtual pedagogical agents", depending on the function of the instructor and the cognitive load placed on the student, and notes that one major challenge is to create systems that are able to adapt accordingly.

Kidd and Breazeal (2004) found a physically-present robot is perceived as more enjoyable, credible, and informative than a video character; Wainer et al. (2007) determined that an embodied robot is more helpful and attentive than a video or simulated character; Tapus et al. (2009) discovered that cognitively impaired and/or Alzheimers patients are more engaged by robot treatment than virtual agent treatment; and Kiesler et al. (2008) showed that health advice from a physical robot was followed more often than the same advice from a robot video virtual agent.

However, a humanoid robot should not try to appear too human (Mori (1970)). Hegel et al. (2011a) argues that a robot performing human-like signals and cues is viewed as inherently dishonest, risking a loss of credibility, and therefore these signals and cues should be used judiciously. In this context, social signals in robots are deliberate, meant to alter the behavior of another being or guide an interaction, and are always created by the programmer, and social cues in robots are features or signs that convey information, whether intended or not, that are not meant to guide an interaction (i.e., everything except signals). Because trying to appear more human is (arguably) deceptive, and the cost of deceptive signals is high and can impact credibility and perception of reliability, therefore, using human-like signs should be weighed carefully.

Finally, Herberg (2013) examined the physical design of robot tutors and children's expectations and concluded that an animal is the design most favored, slightly above humanoid, and patience, politeness, friendliness, caring, knowledgeability, timeliness, and configurability are the most desired traits.

The large body of work that establishes the benefits of embodiment in interaction strongly supports the potential for this research, particularly in the application of intelligent instruction from a robot.

2.3.2 Robots as teaching tools

Interactive robots (as opposed to robots as a mechanical engineering or programming instruction medium, which are outside the scope of this work) as tools for instruction are increasingly attracting the interest of both roboticists and educators. Recently, in Fridin (2014b) and Fridin (2014a) an interactive robot was employed as a "teacher's assistant." Using pre-recorded and choreographed stories and movements in a small classroom setting, a Nao robot was used for story time for kindergarten-aged students. In this scenario, the robot's behaviors were predefined to approximate intelligence, but no intelligence autonomy was incorporated, as a human operator directed the sequence of behaviors. The robot was accepted by students and showed usefulness as an instructional tool. Similarly, Kose and Yorganci (2011) used a Nao robot to teach a large number of preschool-aged students Turkish sign language; this study employed the robot to both demonstrate the signs and visually recognize sequences of flashcards with the correct signs. Other works (Chin et al. (2011)) also concluded that a robot has potential for use as a teaching tool. By placing the robot in the explicit role of the instructor, this work is a significant contribution to the growing movement for using robots as teaching tools.

2.3.3 Robots for people with Autism

Robots developed for interaction with people with Autism Spectrum Disorders (ASD), particularly children, have been studied since Weir and Emanuel (1976) first discovered that robot interaction can be beneficial in those cases (Michaud and Théberge-Turmel (2002); Robins et al. (2009); Billard et al. (2007); Scassellati et al. (2012)). While it is greatly hoped that development of interactive robotic systems can help educate and provide therapy for children with ASD (Dautenhahn and Werry (2004)), much work remains focused on eliciting positive interactions, e.g., Feil-Seifer and Mataric (2011), rather than providing instruction.

Feil-Seifer and Mataric (2008) presented a control architecture for development of autonomous robots for intervention for children with ASD and observed that the behavior of the robot is essential to the success of intervention. Greczek et al. (2014) is similar to our work in that it borrows an approach from occupational therapy called "graded cueing," that uses a series of prompts most analogous to System of Least Prompts (see Section 3.3.3). This approach was not explictly used to teach but instead was used in a single-step "copy-cat" game as therapy for children with ASD using a Nao robot.

2.3.4 Health and exercise robots

Another breakout area identified by researchers for robots in education is health and exercise. Addo et al. (2013) detailed a prototype implementation and future plan of a cloud-based approach using a Kinect camera, an Aldebaran Nao robot, and a virtual environment to interact via verbal and non-verbal communication with individual children as an exercise coach. Lee and Kim (2010) describes an "interactive robot-based tutoring system" that measures the user performance, generates a model, and provides tasks and feedback; however, the design is somewhat limited by very application-specific methodology, as performance measurement, training, and feedback are all designed specifically for a ball-passing-by-kicking task. In Howard et al. (2012), physical exercise demonstrations are translated to robotic movements via a mixed-reality system for the purpose of teaching exercise. Rector et al. (2013) uses skeleton-tracking via a Kinect to perceive humans and provided auditory-only feedback for yoga instruction. Fasola and Mataric (2013) used a robot to provide motivation for exercise in elderly adults.

The use of robots by these researchers in this physical, real-world domain supports this work's efforts to provide the same embodied instructional presence to teach real-world, socially valid life skills.

2.4 Computer interaction

There is a vast field of research in human-computer interaction, particularly for instruction (Jacko (2012)). One highlight is that prior work has found that people with disabilities can benefit greatly from multimodal human-computer interaction technologies (Jaimes and Sebe (2007); Roth and Pun (2003)).

The line between multimodal human-computer interaction and human-robot interaction is becoming blurred, particularly with the influx of wearable devices (Zhang and Rau (2015); Brewster et al. (2003)), assistive devices (such as automated wheelchairs as in Simpson et al. (2004) and Kuno et al. (2003)), toys (such as those used by Westeyn et al. (2008) to assess child development and recognize disabilities), remote collaboration devices (such as commercially available remote presence devices like the iRobot (2016) Ava 500), and other emerging technologies. The combination and complementary use of technologies, such as those presented in this work, is an area of great potential.

In the domain of human-computer interaction, Begoli et al. (2013) is closely related to this work in that a formulation of Applied Behavior Analysis as a process ontology for intervention by intelligent agents for children with ASD was proposed; however, it was implemented as a human-computer interface instead of a robotic instructor, and no experiments were conducted with human participants. While not implemented in this work, applying the use of behavior measurement and statistical process control to track the student's learning from Begoli et al. (2013) would be an interesting challenge in a robotics application, particularly with regards to perception.

2.5 Context-aware augmented reality

Augmented reality has been shown to have great potential in education (Bower et al. (2014)). In particular, the limited number of studies using AR to teach students with I/DD have been successful. Richard et al. (2007) taught a plant labeling task to both students with disabilities and without using an AR table/camera system and observed that students with disabilities were more motivated and enthusiastic than students without disabilities. McMahon et al. (2013) taught independent living skills, particularly identification of food allergens to individuals with ID using a mobile application. Vocational skills were taught in Gómez et al. (2014), which used mobile devices to show interactive guides to the user and provide location and directions in an office environment, and in Chang et al. (2013), which used AR tags to teach meal preparation in a cafeteria setting. Independent navigation skills using augmented reality were taught in McMahon et al. (2015); Smith (2013). There is a recent growth of AR and VR technologies being used in education, but so far there has been very little use in the area of education and training for students with I/DD (Freina and Ott (2015); Sapargaliyev (2015)).

To provide information appropriate to the situation, "context aware" technologies typically use tags to provide the context; e.g., Chang et al. (2013) created an interactive system for prompting using visual codes detected by overhead cameras, and Gómez et al. (2014) uses QR tags and user-entered context information coupled with an intelligent middlewear layer. Gómez et al. (2011) teaches vocational skills with smartphones. Context awareness was provided by bluetooth tags in Chang and Wang (2010) to teach wayfinding using a PDA. Contrary to these approaches, the approach presented here uses computer vision and machine learning methods to determine the context without the need for coded tags.

Chapter 3

Approach

An intelligent robotic instruction system capable of teaching socially valid life skills to humans is a particularly high-level and ambitious project. A modular approach to development of the system has been taken; The key functional pieces are grouped into perception, cognition, and interaction components.*

3.1 Architecture

This research has the following components (Figure 3.1):

1. A perception component, for object detection and tracking, scene classification, and human tracking for interaction.

2. A cognitive component, for evaluating the behaviors of humans and the task being taught, and providing the appropriate instruction responses.

3. An interaction component, with new software created and publicly available software implemented for speech, speech recognition, gestures, and nonverbal cues.

*Portions of the work discussed here have been collaboratively conducted and published in Reardon et al. (2015) and Reardon et al. (2016).

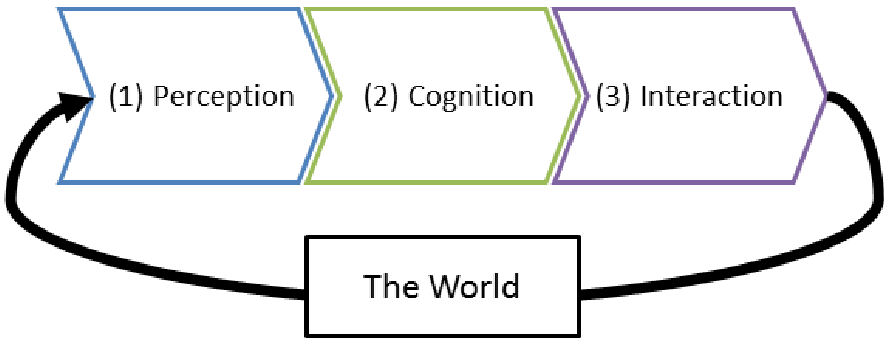


Figure 3.1: Approach overview.

3.2 Perception

Because the purpose of this research is not to advance the field of perception itself, the perception component of this approach uses available technology to create a robust and efficient system for object detection and tracking, image classification, and human tracking for interaction. To reduce complexity and scope, where possible and unobtrusive, publicly available software and simplification through engineered environments are used in this component. Where necessary, custom software has been developed.

Objects in this context mean physically present forms in the environment that are pertinent to the learning task at hand (e.g., items to be addressed or manipulated as part of an experiment); they are neither participants (i.e., humans or robots) nor elements of the surroundings that are part of the environment. In an instructional scenario where both a human and the instructional system must perceive and interact with objects, the instruction system, whether robot or augmented reality, must be able to observe the objects to accurately interpret the performance of the student, make proper decisions, and provide correct instruction; therefore, a fast, accurate method of tracking objects is critical. To approach the object recognition and tracking problem, known assumptions about the objects, such as color and size, are used to simplify the task. Where necessary (for example in the experiment in Section 4.9.1), the objects are modified with colored tags. Custom software created for this research is then employed to localize, identify, and track all objects. We can note that this results in no loss of generality; state-of-the-art object recognition and tracking approaches could be substituted in a modular fashion without loss of performance.

For instruction scenarios involving online classification of images for reality augmentation, such as those in Section 4.8.1 and 4.8.2, image classification using standard supervised learning approaches are employed. Image examples are collected *a priori* and one or multiple classifiers are trained for each classification task.

In robot interaction scenarios, human skeletal tracking is used for the purpose of maintaining gaze with the student participant. Initially, human skeletal tracking for detecting human idle states was attempted (e.g., when the student is finished manipulating objects); however, the resolution and error rates in RGB-D-based skeleton tracking software/hardware were insufficient for that purpose. Instead, object tracking is used to determine when objects are no longer being manipulated (discussed in Section 6.2.7).

3.3 Cognition and instructional methodology

The approach employed to teach socially valid life skills to students with disabilities involves formulating proven instructional methodologies and applying them to robotic and augmented reality instruction systems as part of a cognitive decision making framework. Applied Behavior Analysis (ABA) approaches to instruction, specifically response prompting instructional strategies including Constant Time Delay, System of Most Prompts, and System of Least Prompts (Wolery et al. (1992)), are well-suited for the planned tasks, in that they have been shown to have success teaching discrete and chained (i.e., sequential) tasks, and are formulated in a way that can be applied to robotic instructors. These approaches are commonly used in human-based instruction in the education community.

Generally, the goal of ABA approaches is to modify human behaviors or teach by assessing the environment and acting to stimulate a targeted behavior. One method of action is providing assistance through prompting in an effort to elicit a desired response. One of the key advantages of prompt response strategies is the possibility of different modalities; common types of prompts are vocal, visual, gestural, models (demonstrations), or physical prompts. Another beneficial aspect of prompting is the ability to "fade" or reduce the intrusiveness of the prompts provided to enable individuals to perform the desired behavior independently.

3.3.1 Cognitive framework for instruction

To implement the overall intelligent instruction process, a cognitive framework has been created. Figure 3.2, inspired by Isla et al. (2001), illustrates the framework at a high level. In the cognitive process, an interpretation of the states and actions of the world is created by taking basic sensory information from the world and perceiving information salient to the task at hand. Then, the system reasons on that information, given the predefined (or learned) knowledge of the task at hand, to generate through evaluation (e.g., about the human activities being observed and the correctness of a response) higher-level representations about the scenario. Using that higher-level information, a decision is then made. In the IRI scenario, this involves using the appropriate instructional methodology to select the correct instruction response (e.g., present stimulus, prompt, consequence, reinforcement). Finally, the action is articulated in the world via the navigation and motor system.

The cognitive approaches used are inspired by related work but tailored to the task at hand. Leveraging the response prompting instructional strategies in the cognitive system represent a novel approach to an intelligent robotic and augmented reality system for instruction.

3.3.2 Constant time delay

The Constant Time Delay (CTD) procedure is an instructional strategy that uses prompts, provided after a time delay, or "prompt delay interval," following a "task

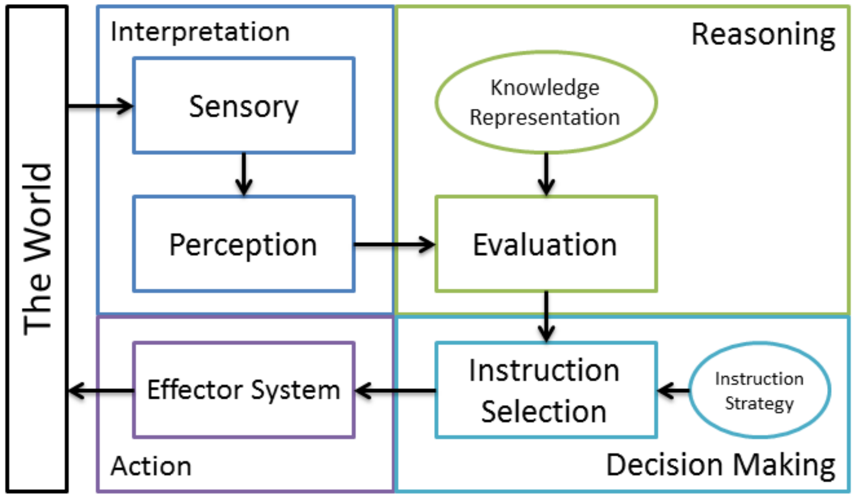


Figure 3.2: Flowchart of the cognitive system for instruction.

direction," which is a question or cue to which the student responds. The description of the task at hand is referred to as the "target stimulus" and is presented to the student. The prompt presented to ensure the task is done correctly is called the "controlling prompt." Initially, the delay between the task direction and controlling prompt is zero, in what are termed "zero-second delay trials." The prompt delay interval is constant for a set of instruction trials until the criterion is met, then systematically increased. There are five possible outcomes of an iteration of the CTD procedure: the student responds correctly before or after the prompt; the student responds incorrectly before or after the prompt; and the student does not respond. A flowchart of the CTD procedure used in this research is presented in Figure 3.3.

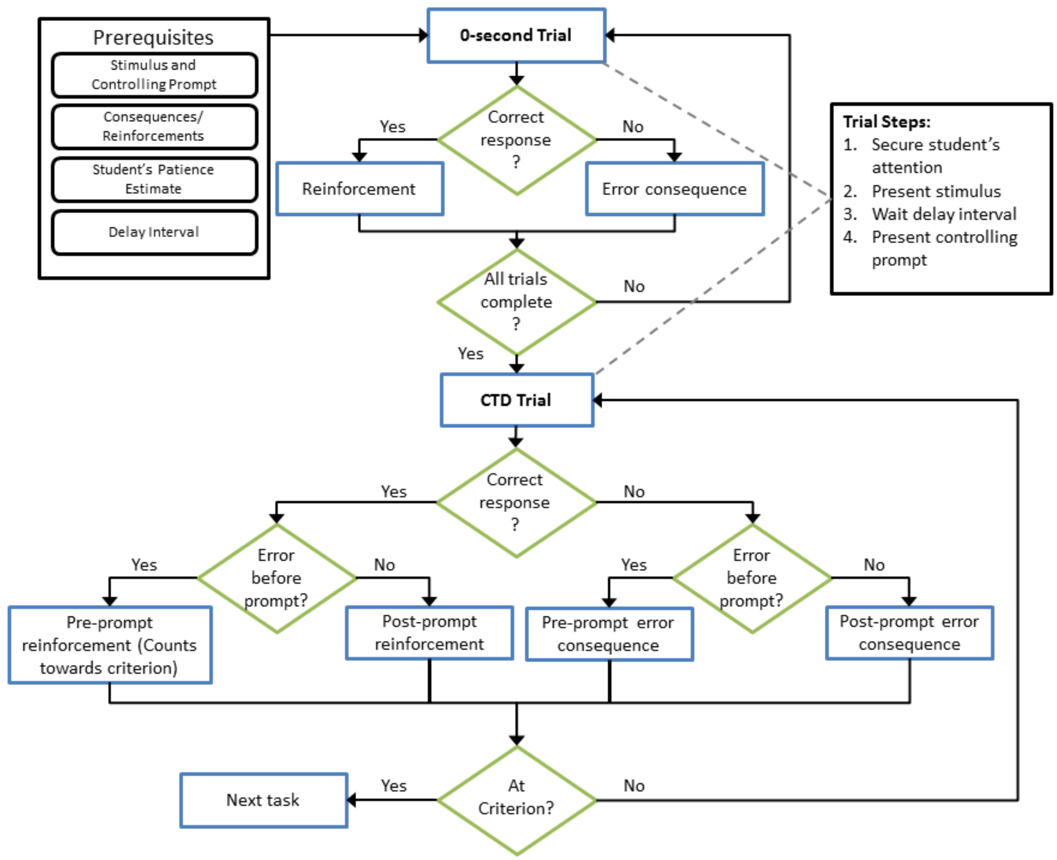


Figure 3.3: Constant Time Delay (CTD) flowchart, adapted from Wolery et al. (1992).

Consistently correct response before the prompt is the goal of CTD. CTD has been successfully used to teach both discrete and chained responses, which require a single response and a sequence of behaviors, respectively (Wolery et al. (1992)).

3.3.3 System of least prompts

Like Constant Time Delay, System of Least Prompts (SLP) is a response prompting strategy. In the SLP instructional methodology, a hierarchy of prompts is arranged from least to most intrusive. At the least intrusive level, no prompt is used. At the most intrusive level, a controlling prompt, i.e., one that assures the task will be correctly performed, is used. The hierarchy of prompts is traversed iteratively to provide more assistance and information as needed. At each iteration, the target stimulus is presented with the prompt for the current level (initially, no prompt). A constant amount of time is allowed to elapse before and after each prompt, known as the "response interval." When a correct response is given, it is reinforced, regardless of when it occurs (i.e., at any point in the hierarchy). When an incorrect response is given, the prompt level is escalated. Possible outcomes of each iteration include: unprompted correct, unprompted error, prompted correct, prompted error, and no response error. Figure 3.4 shows a flowchart of the SLP procedure adapted for this research.

The goal of SLP is for students to respond correctly before any prompt is delivered, at the lowest level of the hierarchy. SLP is considered most effective for teaching chained responses, although it has shown success with discrete responses as well (Wolery et al. (1992)).

3.3.4 System of most prompts

System of Most Prompts (SMP) is a response prompting strategy that is very similar to SLP, except that in the SMP methodology, the hierarchy of prompts is arranged from most to least intrusive. The instruction begins therefore at the most intrusive

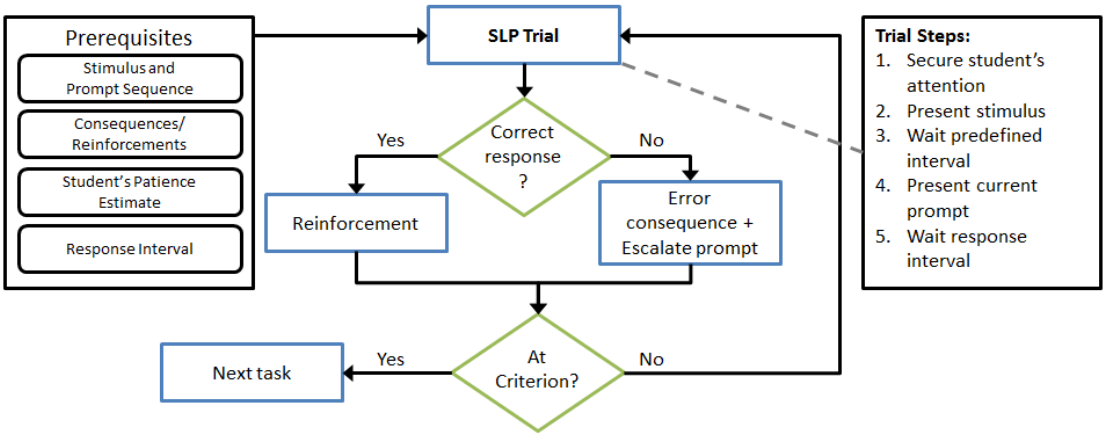


Figure 3.4: System of Least Prompts (SLP) flowchart, adapted from Wolery et al. (1992).

level with the controlling prompt. The hierarchy of prompts is traversed iteratively in decreasing order of intrusiveness. As with SLP, a constant response interval is used, and reinforcement is provided for correct answers. When an incorrect response is given, the prompt level is escalated, as with SLP. Possible outcomes of each iteration include: unprompted correct, unprompted error, prompted correct, prompted error, and no response error.

Figure 3.5 shows a flowchart of the SMP procedure adapted for this research. The intuition behind the SMP approach is to guarantee that the student first makes a successful response (via the controlling prompt), then to fade the intrusiveness of the prompt to work towards full independent behavior.

One observed difference is that with SMP, it is highly likely that the entire prompt hierarchy is traversed for each instruction, which could make the time expended for each instruction longer; however, because the prompts are arranged from most to least intrusive, errors may be less frequent. Further examinations of this interesting tradeoff in the context of experiments conducted on intelligent instruction systems are discussed in Chapter 6.

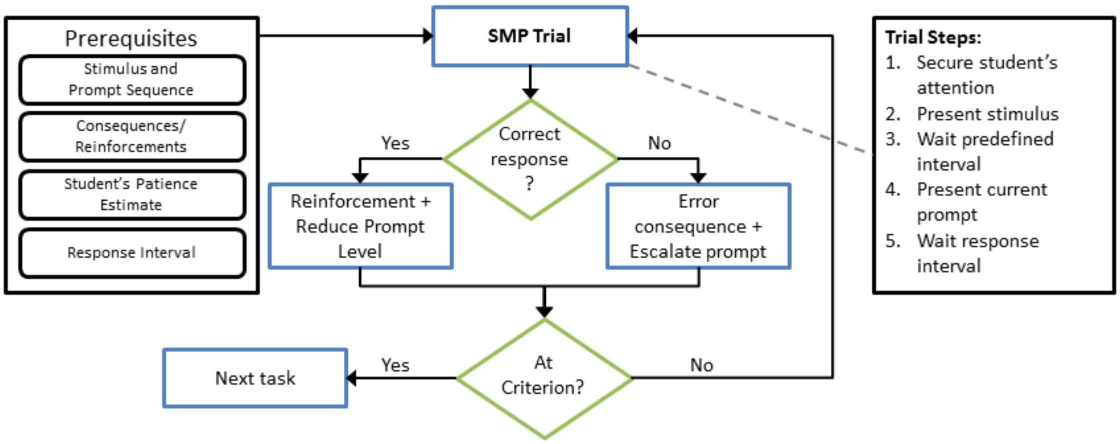


Figure 3.5: System of Most Prompts (SMP) flowchart, adapted from Wolery et al. (1992).

3.3.5 Chained and discrete tasks

Tasks can be subdivided by the manner in which the steps of the task can be taught. Discrete tasks are tasks where a single correct response is expected, such as sight words (commonly used words that students are taught to memorize as a whole by sight). Some discrete tasks can be subdivided into smaller sequences of tasks as necessary for instruction.

Chained tasks are sequential in nature. Instruction on chained tasks is conducted step-by-step in the sequence. Examples of a chained task include most building tasks: building a structure (e.g., from the ground up), assembling an object or puzzle, etc.

Because of their sequential nature, chained tasks can be taught from the beginning of the sequence, in what is known as *forward chaining*, or iterating from the end of the sequence, known as *backwards chaining*. Tradeoffs exist between both; this work employs both approaches as appropriate to the task, and examines the results in Chapter 5.

3.3.6 Applicability of methodology

Response prompting, specifically the CTD, SLP, and SMP methodologies, have been used with strong success to teach pupils with a wide range of disabilities (Wolery et al. (1992)), which provides strong justification for the application of these methodologies to instructing our desired target group.

In addition to being highly applicable to the population in question, the well-defined structures of the instructional procedures are algorithmic in nature and can require physical interaction with a pupil, and are thus particularly well-suited to implementation on a robot. In CTD, one prompt is identified and delivered per trial, the time delay is constant, and decision rules for changing procedure based on student responses are reasonably well-defined. In SLP and SMP, several prompts are represented in a hierarchy, a constant time interval is used as well, and events that trigger traversal of the hierarchy and reinforcement are discrete and well-defined.

3.3.7 Reliability of methodology

The reliability of response prompting lies in the concept of the controlling prompt. The controlling prompt is the most intrusive prompt needed, which provides the highest level of assistance necessary for the students to achieve the task. Because the controlling prompt is selected to be appropriate not only for the task but for the student receiving instruction, the student should always be capable of following the controlling prompt. By following the controlling prompt, a student will always achieve the correct answer. The controlling prompt is, in essence, the solution, delivered in such a way that a willing participant cannot fail to present the solution and achieve the correct response. Because there are a small number of errors allowed (e.g., in SLP) before the controlling prompt is triggered, in a properly designed system there can never be a circumstance where the student remains in an unsuccessful state.

Furthermore, the inclusion of a prompt hierarchy in the response prompting approach means that a system using response prompting is reactively adaptable to the ability level of the student. A student with a higher ability (e.g., using a system designed with SLP) may never or seldom require the highest prompt level, whereas a student with a lower ability may require more intrusive intervention more frequently, but both students can learn from the same system.

Modeling the students' abilities to create a more planned system is an opportunity to advance this approach and is discussed in Section 6.3. Discussion of the strengths of response prompting in the context of the experiments conducted for this research is presented in Section 6.1.

3.3.8 Formulation of response prompting for intelligent robot instruction

This research presents an adaptation of the response prompting methodologies for use by an instructional robot, and incorporates that formulation into the IRI and ARI cognitive system.

The challenging aspects of this endeavor can be grouped into two main categories: those aspects of instruction that are *defined*, i.e., represented by the instruction methodology, and those aspects of the instruction process that are *undefined*, i.e., they are present in the unspecified knowledge, skills, and behaviors of the human instructor but not detailed in the methodology.

The defined aspects of the methodology include the prerequisites, such as the stimuli and controlling prompts, reinforcements, and response intervals, and the general actions involved in the steps of a trial, such as presenting the stimuli, prompts, and reinforcements. The undefined aspects are the essential components that a human instructor innately possesses and brings to the table. They are all of the connecting parts necessary to engage a pupil in instruction.

The defined aspects of instruction are formulated as an algorithm, with mappings between the instructional prerequisites and inputs to that algorithm, and prompts, consequences, and reinforcement deliveries mapped to outputs of the algorithm.

For the undefined instruction aspects, let us consider some examples. The first step of a trial is to secure the student's attention. A human instructor has the ability to not only secure the student's attention, but evaluate whether and to what degree that attention has been secured. Developing an attention-securing strategy was a necessary component of this research, particularly for the IRI. Likewise, we know that there are five possible response outcomes of a trial: unprompted correct, unprompted error, prompted correct, prompted error, and no response error. Determining which outcome occurred is obviously critical to successful instruction via this method, and should be easy to a human instructor. However, the ability to evaluate the outcome of an interaction without ambiguity and with near-perfect accuracy is a non-trivial task for an intelligent system. For example, how can an intelligent system distinguish between a partially correct answer and an entirely incorrect response, particularly when the response is given through object interaction? Further, consider the methods of physical interaction, for example, the presentation of the controlling prompt. The purpose of the controlling prompt is to guarantee a final accurate response from the student, e.g., in the case of SLP when the student is unable to present the correct response from less-intrusive prompts. In many tasks, this could involve physically demonstrating or presenting the correct response. For a human instructor, this should be straightforward and bears little discussion. For a robot, physically presenting a correct response with precision and accuracy all of the time is a challenging problem. From these examples we can observe that there are a number of undefined aspects of instruction that must be explored and formulated in order to construct an intelligent instruction system.

The full adaptation of the response prompting methodologies is incorporated into the overall cognitive system overviewed in Section 3.3. The instruction system spans the Reasoning and Decision Making phases. Representation of the knowledge used as input into the model evaluation includes information necessary to interpret what is perceived. Then, given that evaluation, a decision is made by selecting an appropriate instruction based on the instruction strategy, and which is then articulated by the effector system.

Questions that could be encompassed by this formulation include:

- What is the relationship between human activities (e.g., responses) and robot knowledge, and methods of representing that knowledge?
- How should an accurate evaluation of human responses be ensured?
- What should be the corresponding delivery of reinforcement or consequence to the student?
- How should other instructional criteria, including prompt delivery, delay implementation, and attention securing, be designed?
- How should prompt levels for a robot instructor be selected, and how do they relate or differ from those a human instructor can or would perform?
- Ultimately, are there any behaviors that the robot should or shouldn't perform that could increase the success or rate of learning of a student?

Formulation of response prompting for intelligent robot instruction and subsequent application to a robotic platform therefore represents a significant contribution of this work. Discussion of the consequences of this formulation in the context of experimental results is presented in Chapter 6.

3.4 Interaction

Interaction for instruction is multimodal and varies according to the technological and situational requirements. A overview illustration is presented in Figure 3.6. Generally, the robot instructor has the ability to interact more concretely with objects in the real world, and has the advantages of an embodied agent, as well as the ability to provide social cues such as gaze and reinforcement gestures, whereas the ARI system has

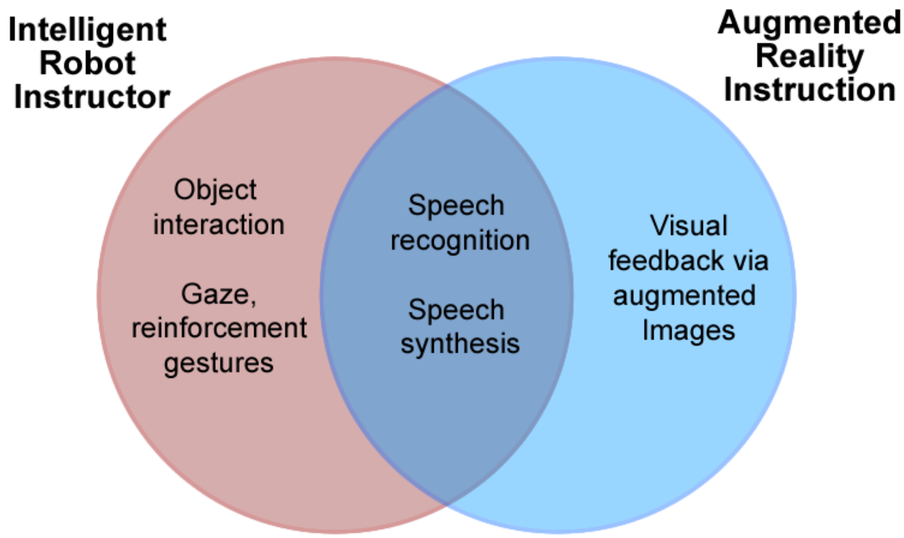


Figure 3.6: Multimodal interaction, IRI vs. ARI system.

the ability to augment images and present them directly into the field of view of the user. Both systems have the ability to recognize speech and provide audio prompts and feedback. Specific details of the interaction components for each experimental scenario are presented in Chapter 4.

3.5 Summary

By dividing my approach into perception, cognition, and interaction components and addressing development in a modular fashion, I have been able to create a complex, high-level system. In particular, this chapter detailed the approach of using response-prompting methodologies formulated for an intelligent, autonomous system. Chapter 4 discusses the implementation of this approach in a series of experiments and the methods used for evaluating the outcomes.

Chapter 4

Experiments and Evaluation

This chapter details the methods and materials used to conduct experiments to validate the approach, including information on participants (Section 4.1), the hardware employed (Section 4.2), the software implemented (Section 4.3), the perception and interaction capabilities (Section 4.4), implementation of the instructional methodologies (Section 4.5), metrics and evaluation (Section 4.6), and experiment details (Sections 4.7, 4.8, and 4.9).

4.1 Participants

Students are recruited from the FUTURE Postsecondary Education Program at the University of Tennessee (FUTURE (2016)). FUTURE is a program for young adults with intellectual and developmental disabilities. As the goal of the FUTURE program is to empower these students with academic, vocational, and decision making skills, the exploration of the potential benefit for intelligent systems to help this population is an excellent interdisciplinary opportunity.

Institutional Review Board (IRB) approval was secured for human participant testing for this research as an addition (IRB form D) to the FUTURE program's ongoing instructional research.

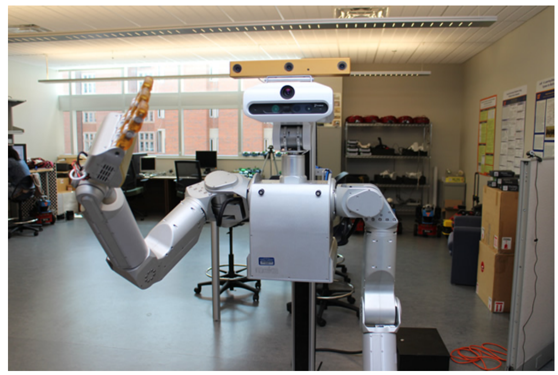
4.2 Hardware

4.2.1 Robot

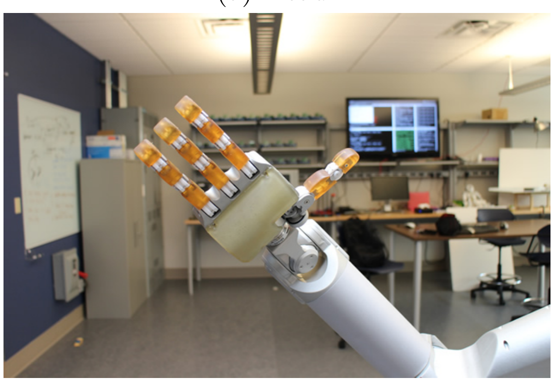
The robotic hardware for this research is a humanoid robot named Rosie (Figure 4.1). Rosie is a custom-designed Meka Robotics M3 mobile humanoid robot, equipped with two Meka A2 series 7 degree of freedom (DOF) elastic arms with 6-DOF force torque sensors; two Meka H3 series 5-DOF (three fingers and one 2-DOF thumb) hands; a Meka M3 sensor head (Figure 4.2) with 2-DOF movement, two PrimeSense (v1.08 and 1.09) cameras, one Point Grey Flea3 8.8 MP color USB 3.0 camera with a wide angle low distortion lens, and one Point Grey Bumblebee XB3 1394 stereo camera; a torso on a prismatic lift mounted on a Meka B1 omnidirectional base; 2 Lavalier 3.5mm uni directional microphones and integrated speakers.

Rosie is equipped with two PCs: one Intel(R) Core(TM) i7-3770S 3.1GHz PC with 4 cores and 3 GB of memory providing real-time functionality of the base, arms, hands, and lift; the second is an Intel(R) Core(TM) i5-3470S 2.9GHz with 4 cores and 8 GB of memory dedicated to the vision and audio components.

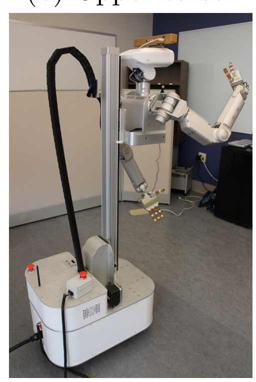
(a) Head



(b) Upper torso



(c) Hand



(d) Full robot (rear view)

Figure 4.1: Rosie, the Meka Robotics M3 mobile humanoid robot.

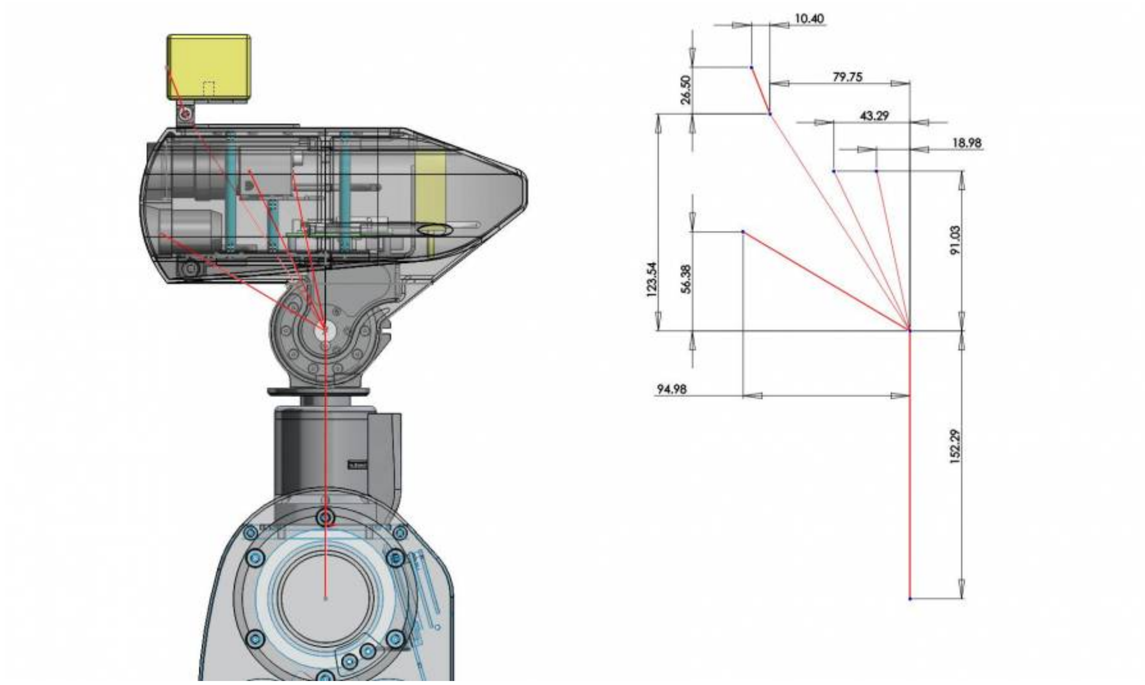


Figure 4.2: A schematic of Rosie's sensor head.

4.2.2 Augmented reality with Google Glass

The augmented reality hardware for this research is Glass, a wearable optical head-mounted display device developed by Google (Figure 4.3). The Glass device has a OMAP 4430 dual-core processor, 2 GB of RAM, 16 GB of flash memory, and runs an extended version of Google's Android 4.4 OS. To interact with the user, a 640x360 pixel prismatic projector is situated over the user's right eye, and a bone conduction transducer is located in the Glass frame over the right ear. Sensors include a 5 megapixel camera, microphone, gyroscope, accelerometer, manetometer, and ambient light sensor. A touch sensor, also located in the right side of the Glass frame, is programmed to accept tap and swipe (left/right/down) inputs.

4.3 Software

Software was developed specifically for this research to provide the IRI and ARI system with the capabilities necessary to instruct students with I/DD in relevant

(a) (b)

Figure 4.3: Google Glass (a) front and (b) rear.

vocational skills. The general software implementation is described in this section. All of the skills chosen involve interaction with objects; specialized perception and interaction software is described in Section 4.4, implementation of the instruction methodology for cognition on an IRI is described in Section 4.5, and software for each instructional task are described in Sections 4.8 and 4.9.

4.3.1 Robot software

Software for this research relies heavily on the Robot Operating System (ROS) (Quigley et al. (2009)) library. Because the Meka robot Rosie possessed very limited software capabilities as delivered from the manufacturer, to create the robot essential for this research, I led the development of a ROS-compatible software stack for Rosie, including motion planning and manipulation, vision, speech synthesis, and speech recognition. Rosie's base software stack is shown in Figure 4.4 and includes the following:

- A common library class with several levels of abstractions for movement commands, from individual joints to entire joint chains, and a common interface to joint state information.
- Simple end-effector movement using the Meka-provided IK and FK solvers.

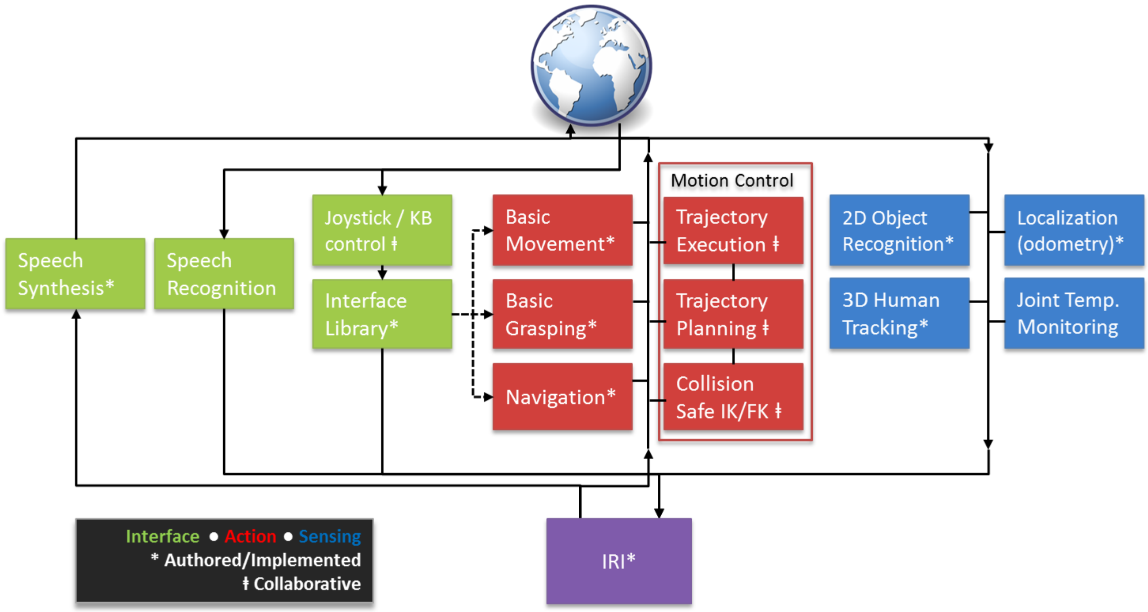


Figure 4.4: Overall software architecture for the IRI.

- Higher-level, self-aware (i.e., collision-avoiding) kinematic solutions for full trajectory planning for joint chain and end-effector movement using the ROS MoveIt! package (Sucan and Chitta (2016)).
- Joystick control of the primary movement functions.
- Localization in the environment (using odometry data).
- Robot navigation in the environment.
- Rudimentary grasping.
- Human skeleton tracking using the RGB-D cameras and the ROS OpenNI package (OpenNI Consortium (2014)).
- Speech synthesis using eSpeak (eSpeak (2016)).
- Speech recognition using Pocket Sphinx (Carnegie Mellon University (2016)).
- Joint temperature monitoring.

- Color blob detection using OpenCV (Bradski (2000)).

Specialized software to perform the IRI tasks are discussed below.

4.3.2 Augmented reality software

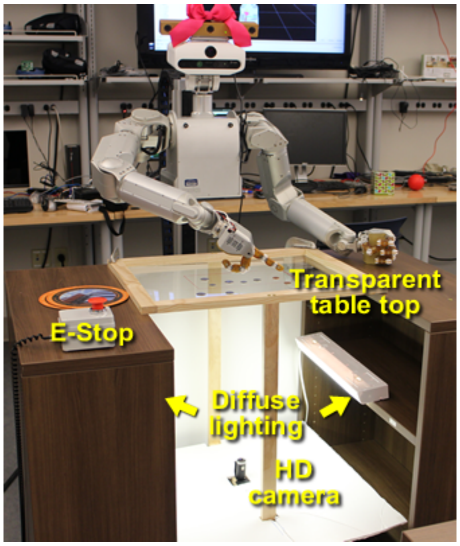
Application software for the Glass has been custom developed for this research in Android 4.4.2 using the Glass Development Kit. Server-side software uses software developed in PHP for communication over HTTP, and the Python implementation of OpenCV (Bradski (2000)) for classification.

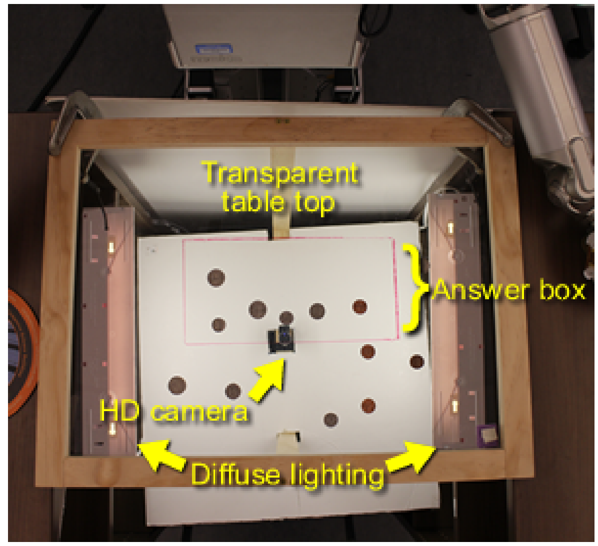
4.4 Perception and Interaction

Perception in the form of human and object recognition and tracking, combined with speech synthesis, speech recognition, and either reality augmentation or gestures by a robot, form the basis for the system's interaction capability.

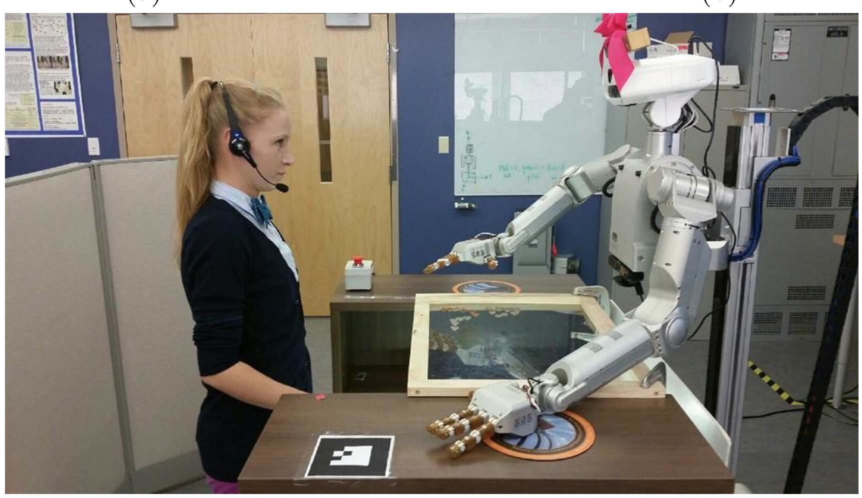
4.4.1 Object recognition and tracking

For the IRI application, the system uses a high-definition camera mounted under a table with a transparent surface (Figure 4.5). To create a simple yet highly accurate and efficient solution to the object tracking problem, known color information about objects being tracked is used. In the event of similarly colored objects, small colored tags are discretely affixed to the bottom of the objects.





(a) (b)



(c)

Figure 4.5: The table setup for the instructional setting from the (a) student view, (b) overhead, and (c) side with participant.

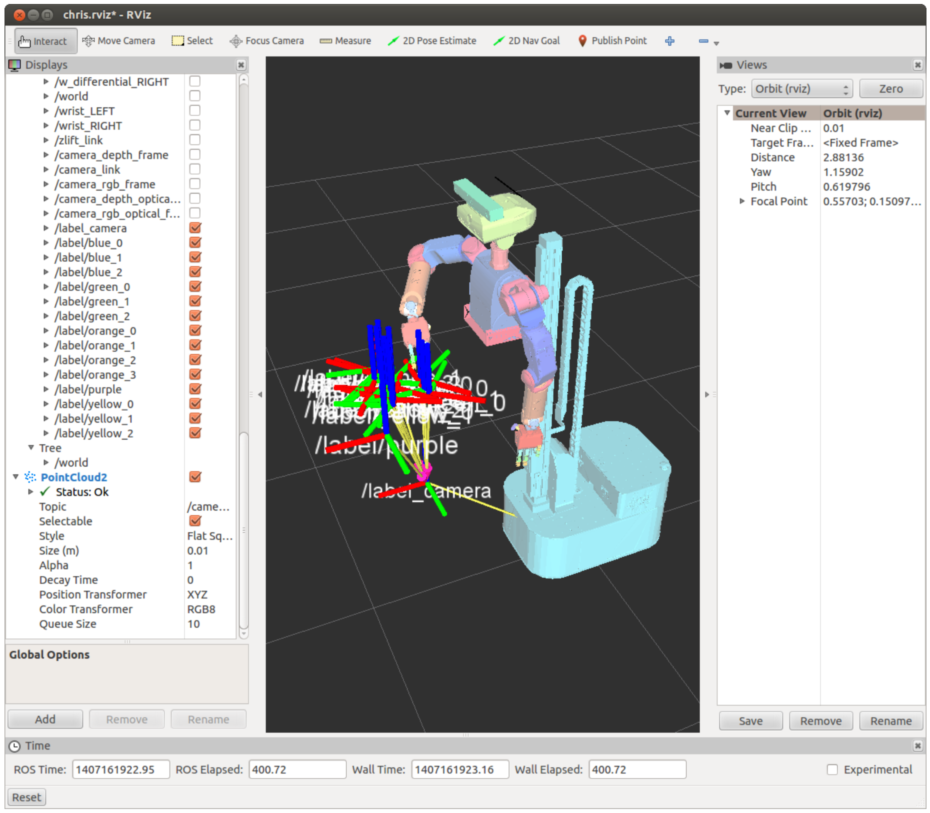


Figure 4.6: Object tracking published via tf and viewed in ROS RViz. Red/green/blue axes represent the location and orientation of each detected object.

Custom object tracking software leveraging OpenCV (Bradski (2000)) was created. Objects are segmented in the image using contours derived from HSV ranges, and positions defined as the contour centroid. Orientations, when applicable, are calculated using methods appropriate for the shape. This approach provides live, highly accurate location and orientation information of the objects on the table surface. Pose information for each object is published into the Robot Operating System (ROS) framework using the tf (Foote (2013)) coordinate frame package. Figure 4.6 shows a live visualization of the published data.

One challenge for any vision system is accurately tuning it for use, particularly when deployed in different locations with varying lighting conditions. To address this, a convenient GUI has been created that provides a live, annotated view of extracted location, orientation, size, and identification of objects and allows online adjustment

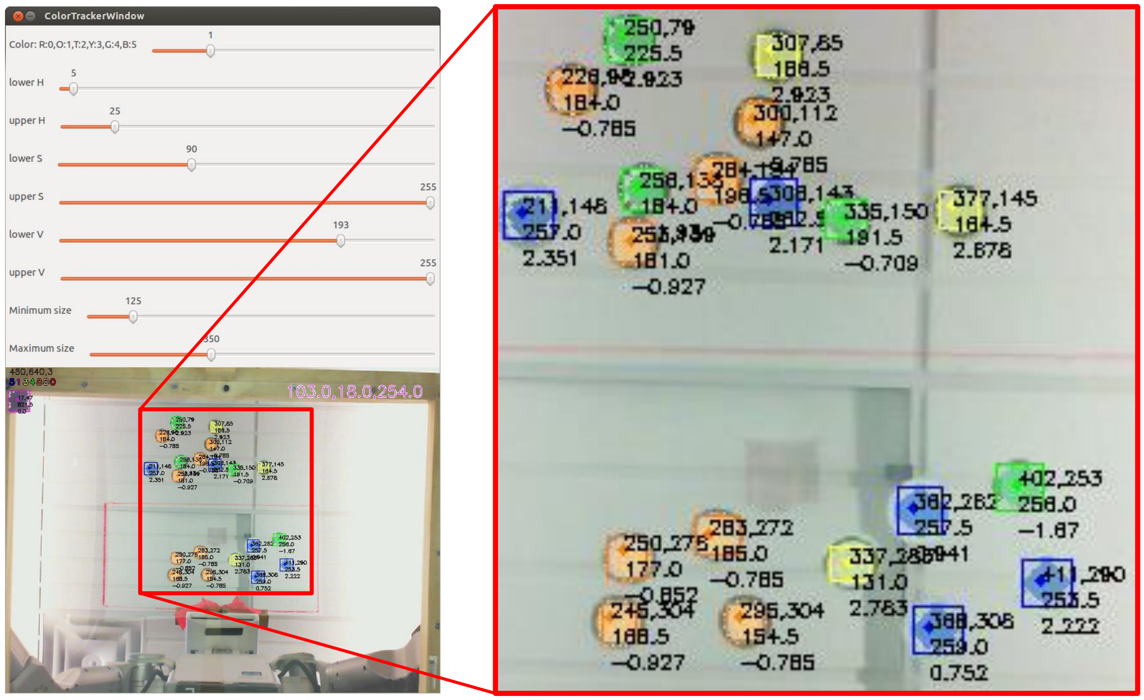


Figure 4.7: The object tracker GUI, with live, adjustable parameters on top and the annotated live image on the bottom. Right zoom shows an enlarged view of the annotated image. Annotations include position, orientation, size, centroid location, and bounding box for each object.

of vision parameters. In the GUI, a live reading of the HSV values at the current cursor location is projected to allow the experiment operator to quickly identify and recalibrate as necessary. Figure 4.7 shows a screen capture of the GUI displaying the annotated image and calibration interface of a scene from the undermounted camera, and Figure 4.8 shows the GUI view of geometric shapes from overhead.

Measured performance of the object tracking system shows it is both highly accurate and efficient. To compute the accuracy of the object tracking system, the error distance between detected and ground-truth centroid is analyzed, as well as a standard object-matching approach using bounding boxes. For matching accuracy, an approach similar to Dollar et al. (2012) is used to evaluate each frame with the PASCAL (Everingham et al. (2010)) detection measure by calculating the area of overlap of the ground truth bounding box BB_{gt} and the detected bounding box BB_{dt}

(a) Initial object recognition (b) Clustered object recognition

Figure 4.8: Annotated object recognition software results seen through the custom object tracker GUI from overhead.

of each detected object. The PASCAL measure states that the overlapping area a_o must exceed 50%, as defined in Equation 4.1, to be considered a match.

$$a_o \doteq \frac{area(BB_{dt} \cap BB_{gt})}{area(BB_{dt} \cup BB_{gt})} > 0.5 \tag{4.1}$$

When tracking objects affixed with colored tags with a radius of approximately 5mm, the object tracking accuracy consistently exceeds the PASCAL matching threshold of 50%, with a mean a_o of 0.86 and minimum a_o of 0.6 (Figure 4.9). The mean error distance (Figure 4.10) between detected and ground truth centroids is less than 2mm.

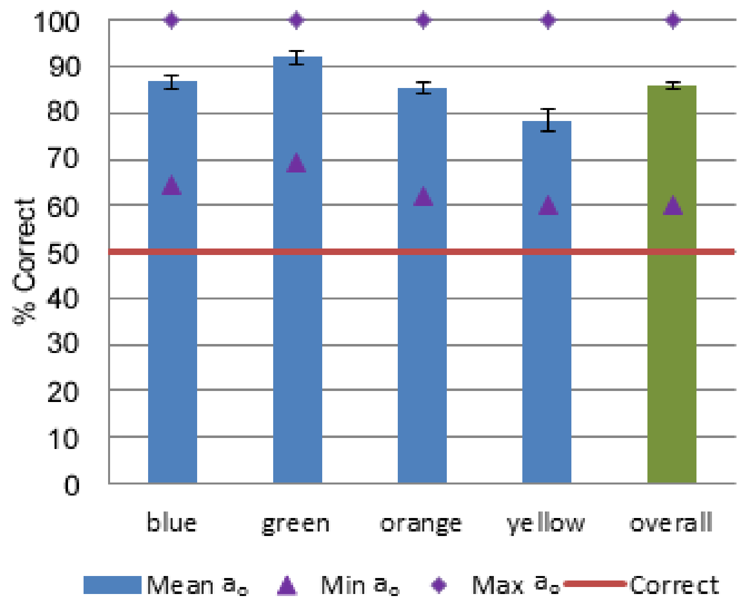


Figure 4.9: Object matching accuracy.

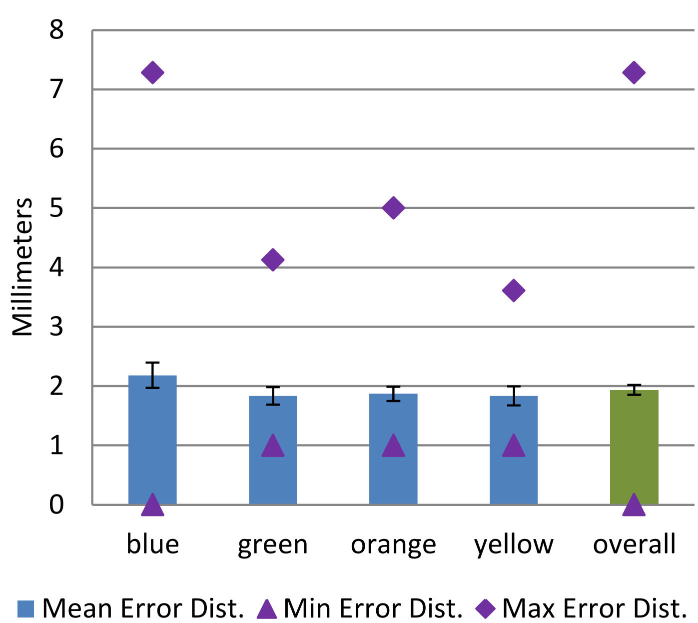


Figure 4.10: Object matching error distance.

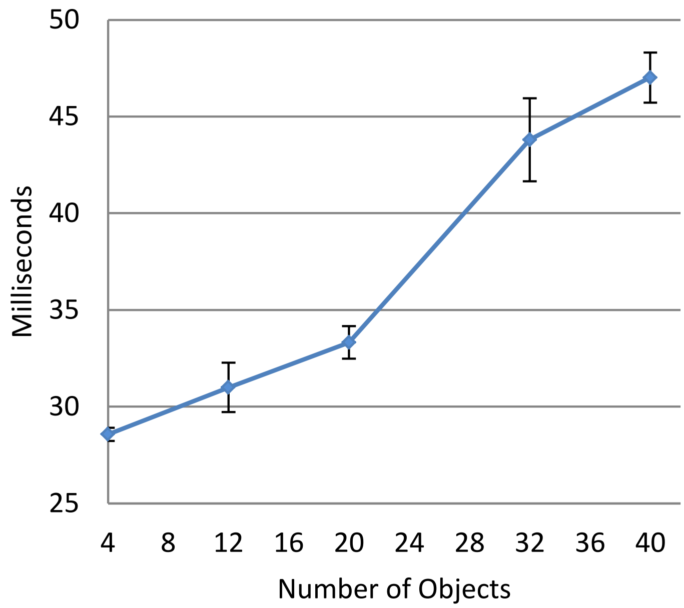


Figure 4.11: Object matching mean tracking time.

To measure the efficiency of the object tracking system, the time until detection of all objects is complete is recorded, using the hardware described in Section 4.2.1, in a live experimental setting for different quantities of objects placed in varying positions throughout the table. Time to detection of 20 objects (Figure 4.11), which has been selected as the upper bound on the number of objects used in any experiments, is 33ms (p=0.95). Using twice that number of objects increases detection time by 20ms (p=0.95). Combined with total system overhead, the tracking system operates at a frequency above 70Hz.

Where appropriate, for certain applications of the ARI system (for example, the experiment detailed in Section 4.8.3) the same software approach to object detection is employed. Instead of an HD camera undermounted on a clear table, a wearable device (i.e., Google Glass) worn by the student is used to collect images of the interaction environment for interpretation.

4.4.2 Robot gestures

An essential need for a competent system for interactive instruction is the ability of the robot instructor to make gestures that are interpretable by, and acceptable to, the human student. The IRI system uses a collection of physical cues to provide a realistic, attentive appearance, as well as gestures to provide explicit, physical instruction.

When the IRI is speaking it uses skeleton tracking to turn the head to face the student, to give the impression of maintaining “eye contact,” and when the student is performing a response, the IRI faces the objects being manipulated. This strategy is part of the approach to create a “securing attention” procedure as part of adapting the prompting methodologies discussed in Section 3.3.

When providing instruction, the IRI uses the poses of objects extracted by the object tracking system to point directly at the objects to which it is referring. It is also able to gesture at objects for other procedural purposes, such as asking the student to reset the table environment in between trials. The IRI also uses gestures when providing differential reinforcement.

To examine gesture interpretability, a simple test where Rosie interacts with a student with I/DD was conducted where the robot attempts to deceive the student half of the time by selecting a coin to gesture to, but then either asks whether the coin is the type of coin being gesturing to, or names a different random coin type, with an equal probability. A confusion matrix is presented in Table 4.1. Over 126 samples, the true positive rate was 0.81, true negative rate 0.87, positive prediction value 0.93, false omission rate 0.12, and overall accuracy 0.90.

Table 4.1: Confusion matrix for gesture test

		Coin named		
		T	F	
Perceived	T	42 (0.33)	3 (0.02)	PPV:0.93
	F	10 (0.08)	71 (0.56)	NPV:0.88
		TPR:0.81	TNR:0.87	ACC:0.90

4.4.3 Perception of context using supervised learning and interaction through augmented reality

Figure 4.12 shows an overview of the ARI system. A student using the ARI (i.e., wearing the Google Glass), when learning to perform a new task, can ask for the next step at any point in the task sequence. The ARI software takes an image from the user's point of view and uploads it to an online server, where the image is processed and an appropriate instructional prompt is pushed back to the user's device. In these experimental applications, the prompt is both audio and an augmented view of the image uploaded, and can include highlighted objects, buttons, or points of interaction, correct models of the solution, and/or text.

Contextual awareness is achieved through perceiving the content of the image and deciding which instructional prompt to present. The image is first parsed for relevant information, then a classifier constructed via supervised learning is used to solve the problem of identifying the correct context of the image. Using the classifier output and the known ground truth, the proper visual and audio prompts for the next step in the task are selected from the knowledge model, which is specific to the task, for example in the form of a lookup table or decision tree. The prompt is then delivered seamlessly to the user through the wearable AR interface.

The key differentiating feature of the context awareness is the ability to only provide prompts for the steps that the user cannot remember. This is especially significant when providing instructional support to this population in that it enables them to quickly learn to perform tasks independently. Such self-directed learning gives the student complete control and propels them towards independence.

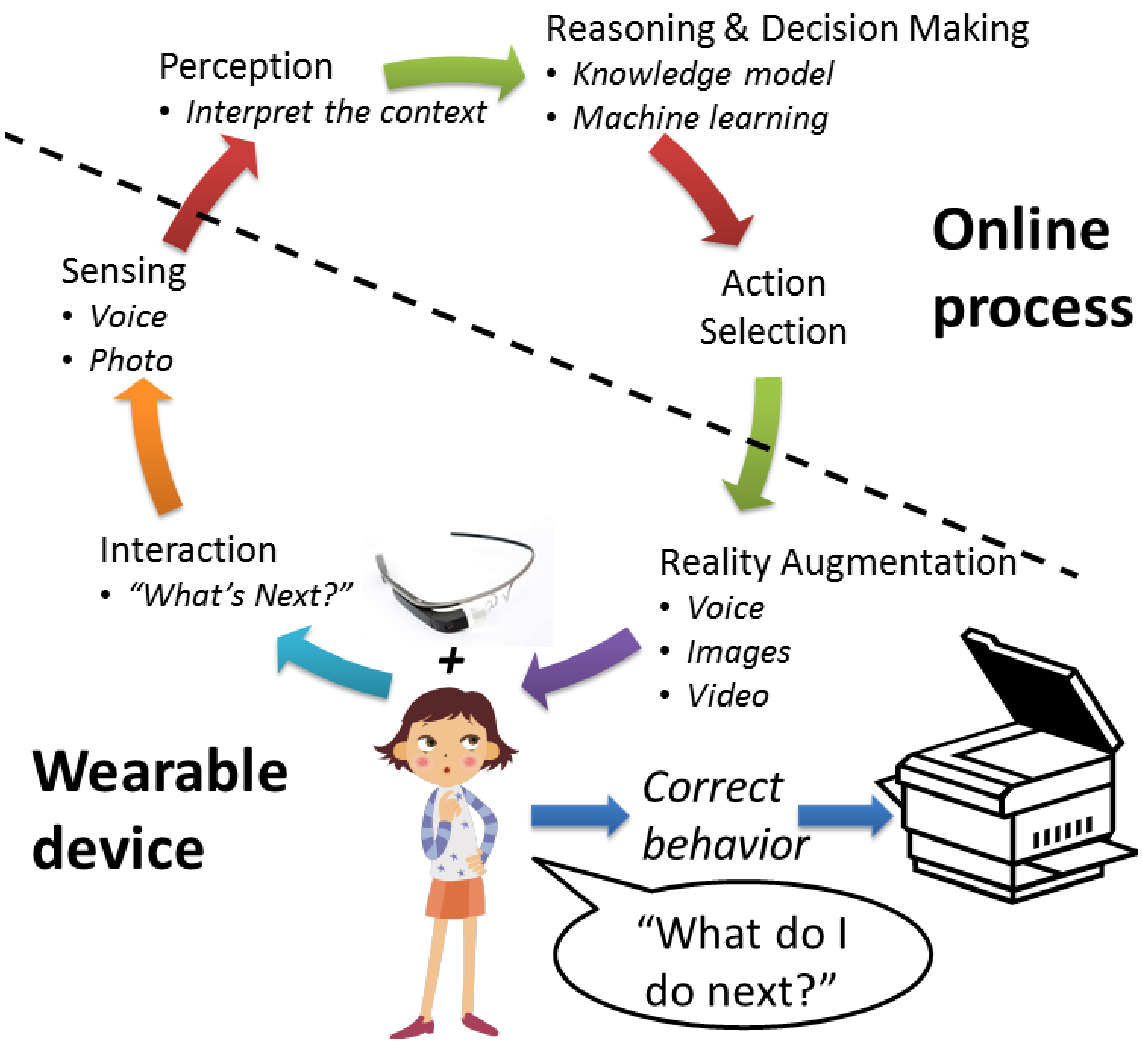


Figure 4.12: ARI system overview. Process begins with user asking for assistance. Using the wearable's sensory information (camera), a picture is uploaded, processed, classified, and the correct augmented image and audio is pushed back to the wearable, containing the proper instructional prompt for the next step of the task.

Because a deficit in organizational skills and technological aptitude is common in persons with I/DD, seamless integration of the various system components and technologies is essential for an efficient user experience and successful learning.

On the wearable side of Figure 4.12, a Google Glass is used for these experiments as the AR device. Being a first-to-market device, the Glass has some limitations. The processing capabilities are limited, particularly in terms of heat dissipation, which under heavy load causes the Glass to automatically shut down. Because of this issue, a client-cloud configuration was selected for the system, where the client is responsible for the user interaction, and the artificial intelligence (perception, reasoning, and decision making) resides on the cloud/server side.

The user experience was streamlined: the simple audio command, "Okay Glass, what's next?" triggers the app. First, the user is presented with a camera viewfinder and the text + audio prompt: "Position the picture and tap." The user is then shown the image they took and asked to tap to confirm. Upon confirmation, from the user's perspective, an image and audio instructional prompt is provided via the Glass display and built-in speaker in around 5 seconds. During this period, the image is uploaded using a multipart HTTP POST request containing the image as an input stream. The request response contains the instruction in the form of an audio prompt and annotated image, which the Android app delivers. Due to limited processing capabilities of the Glass, it was discovered that compression prior to upload actually increased processing time over directly streaming the image via the available 802.11/n network. Another limitation of the Glass is the view screen, which presents a maximum 640x320 resolution image. Because of the limited space, image prompts must be close-in and annotation must be clear and succinct.

On the cloud/server side, intelligent instruction is made possible using currently available tools. OpenCV (Bradski (2000)) is used for image processing and Support Vector Machine (SVM) classifier implementation. The image processing problem is slightly simplified by using known information about the environment such as known color contours to segment and subselect the relevant areas of the image for classification. Images are deskewed and downsampled to a uniform size. Histogram of Oriented Gradient (HOG) features are then extracted, and one or multiple SVMs are used to classify the image. In the event of a failed or low-probability image classification, the user is presented with a simple prompt to try again.

SVMs are trained a priori for the task steps using images of each step taken by the experimenters. A time-consuming challenge was to identify the correct number of images that are able to generalize to sufficiently represent the images the students take, while simultaneously tuning the SVM parameters to achieve a high levels of accuracy without overfitting. While these aspects were manually discovered via trial-and-error; future work could allow for better insight and automation into this process. Prior to the training phase of each experiment (see Section 4.8), the students were instructed on the use of Glass to capture images similar to the training set. In the face of these challenges, it was found beneficial to re-train the classifiers using images captured by the users in between trials to decrease failed classifications for each individual user. Classifier re-training, combined with user instruction, yielded a sufficiently decreasing failed classification rate and increasing user satisfaction.

A major goal of this research was to create a technical framework that allows for multiple different experiments. With modular software development, three different decision workflows are easily incorporated, as described in Section 4.8. Potential future work would include enabling the correct decision workflow to be selected on the fly, thus allowing the user to receive training or assistance for multiple tasks simultaneously, which would further increase independent use of the system.

4.5 Implementation of instruction methodology on an IRI

The instruction process is shown in Figure 4.13 and presented as pseudocode in Algorithm 1. The encoded process is an adaption for an IRI of the SLP / SMP methodologies, which leads to a strongly defined process that has been shown to be successful when used by human teachers. The instructional intervention begins with the IRI giving an introduction and general instructions for the scenario. Task Instruction is the formal term for the step of introducing the task and presenting the target stimulus. Next, the IRI selects a prompt.

In experiments using SLP, initially, there is no prompt; that is, the student is given the opportunity to present an answer independently. In experiments using SMP, the initial prompt is the controlling prompt.

The student then responds, while the IRI observes the student to determine whether the student is idle, the task is complete, or a period of time, formally known as the response interval, has elapsed.

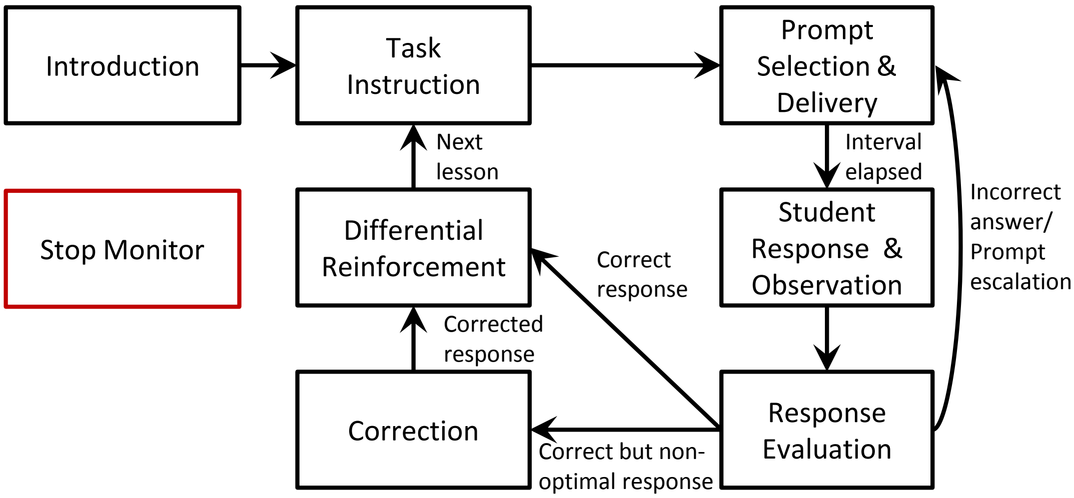
The Response Evaluation determines the result of the step: if an incorrect answer is given, the IRI then uses the evaluated result information as part of the process to select the appropriate response; if a correct answer is given, the IRI provides reinforcement; if a correct but non-optimal answer is given, a correction occurs before reinforcement. All reinforcement is positive. The type of reinforcement is differential, in that it is tailored to the level of intrusiveness of the prompt that was required: students who require less intrusive prompts are rewarded with an increasingly "excited" verbal reaction from the IRI; for the *making change* experiment (Section 4.9.1), an independent correct response (i.e., no prompt was required) also triggered a gesture reinforcement (a "thumbs up").

Algorithm 1 High-level algorithm for IRI prompting implementation

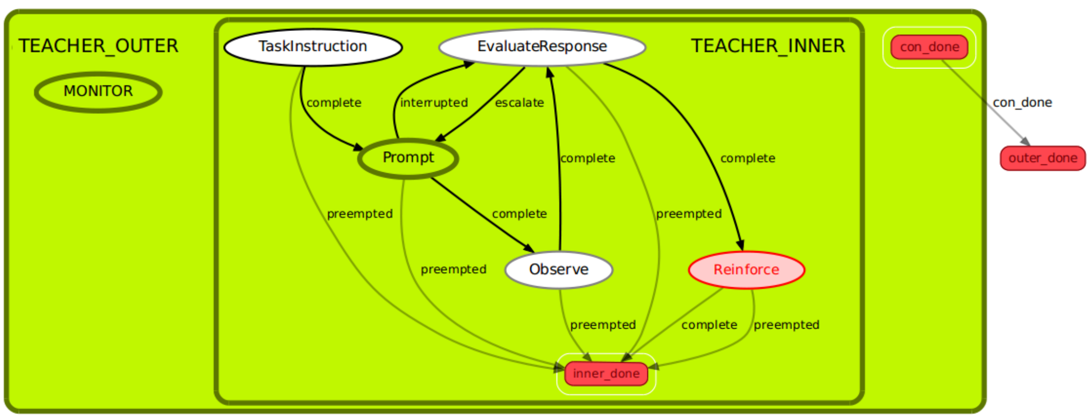
Require: Variables *skill*, *prompt_level*, *prompt_strategy*, *speed*, *step*, *idle_interval*, *prompt_interval*, *first_time* are initialized and treated as global for this pseudocode.

```
function TaskInstruction
    if ¬first_time then
        Introduce(skill, speed)                          ▷ Introduce the skill
        first_time ← True
    problem ← SelectProblem(skill)
    Introduce(problem, speed)                            ▷ Introduce the problem
    Prompt()

function Prompt(results)
    active ← CheckActive()       ▷ Use object tracking system to detect activity
    if active then
        EvaluateResponse()
    if results ≠ 'correct' then
        correct_location ← results[0]
        feedback ← results[1]
        Gesture(correct_location)
        Say(feedback)
    next_prompt ← GetNextPrompt(step)                    ▷ Lookup next prompt
    if prompt_level = controlling_prompt then
        Gesture(next_prompt[0])
    Say(next_prompt[1])
    Observe()

function Observe
    GazeTrackingOffLookDown()                            ▷ Look at the work area
    begun ← False
    while ¬begun do                                      ▷ Wait for the student to begin
        begun ← CheckActive()
    start_time ← Now()
    last_active ← start_time
    while (Now() − start_time) < prompt_interval & (Now() − last_active) <
idle_interval do      ▷ Wait for the student to finish (idle) or run out of time
        last_active ← CheckActive()
    EvaluateResponse()
```

```
function EvaluateResponse
    GazeTrackingOffLookDown()                    ▷ In case coming from Prompt
    results ← CheckResponse(problem)
    evaluated ← EvaluateResults(results)
    GazeTrackingOnLookUp()
    if ProblemComplete(evaluated) then          ▷ Check for all steps completed
        Reinforce(evaluated)
    if evaluated = 'correct' then
        if prompt_strategy = SLP then             ▷ SLP with forwards chaining
            step ← step + 1
        if prompt_strategy = SMP then          ▷ SMP with backwards chaining
            step ← step − 1
        ResetPromptLevel(prompt_level)
                                       ▷ Set prompt_level back to original value
        Prompt()
    else
        prompt_level ← prompt_level + 1
        Prompt()

function Reinforce(evaluated)
    if NonOptimal(evaluated) then ▷ If response was correct but non-optimal
        CorrectNonOptimal(evaluated)                     ▷ Give optimal answer
    DeliverReinforcement(prompt_level)
        ▷ Give speech and gesture reinforcement for highest prompt level reached
    TaskInstruction()

TaskInstruction()                                ▷ Begin with Task Instruction
```



(a)



(b)

Figure 4.13: Implementation overview: (a) shows the overall instruction process based on response prompting methodology; (b) shows the internal representation generated by ROS SMACH, a library for creating hierarchical state machines (Bohren and Cousins (2010)).

The most important decision making the IRI system performs takes place between the Response Evaluation and the Prompt Selection & Delivery steps. Response evaluation begins when the robot has observed the student to be idle over a specified idle interval, or a longer response interval time has elapsed. Response evaluation determines the type of response. That information, coupled with the known information about the previous prompt and human's state (active or inactive), is what the IRI uses to decide the appropriate feedback. In the case where the student is perceived to be actively providing a response during prompt delivery, the IRI immediately reevaluates and presents an updated prompt, if necessary.

A concurrent process is implemented that monitors for a stop command from the student (Figure 4.13). Before beginning the experiment, students were informed that any use of the words "stop" or "exit" in combination with the name of the robot, initiates an immediate shutdown of the robot. This provides an additional layer of safety and comfort for the students beyond the supervision of the experiment operator.

4.6 Metrics and evaluation

4.6.1 Robot performance and acceptance

In order to provide accurate demonstrations and controlling prompts, the robot's ability to perform a desired behavior was thoroughly developed and tested before deployment in a formal instruction scenario. Where appropriate, evaluations to measure the accuracy and effectiveness of the system were conducted (Section 4.4).

In the performance of the first human-robot interaction experiment, the acceptability of the robot's behavior was evaluated with a Likert-type survey, and efforts were made to enhance acceptability.

4.6.2 Single case experimental design

Single case experimental design (SCED), also known as single-subject research design or single subject experimental design, is common in special education research because it allows for the participants to serve as their own control data for the purpose of comparing performances between at least two experimental phases as opposed to comparison between groups or participants (Gast and Spriggs (2010)).

SCEDs are traditionally used in the behavioral sciences as an experimental approach to evaluate the functional relationship between independent and dependent variables. In education studies, the independent variable is the instructional intervention (i.e., teaching), and the dependent variable is the acquisition and maintenance of knowledge and skills for independent behavior (i.e., learning). Similar to group and correlational research designs, single case designs are quantitative in nature.

Different from these methods that are more familiar in other scientific fields, SCEDs use individual participant as his or her own control (Sidman (1960)). In any SCED, a participant's performance (the dependent variable(s)) is measured in a control condition, known as baseline, and at least one intervention condition, and recorded in a time-series fashion. Demonstration of experimental control, that is, that the change in the dependent variable is likely due to the change in the independent variable, is shown through the systematic manipulation of the independent variable. Different case subject designs arise from the manner in which the independent variable is manipulated, for example, through introduction and withdrawal, or when introduction is staggered across people or skills (Horner et al. (2005)). This manipulation allows for the effects of intervention to be evaluated, and a causal relationship to be inferred, thereby establishing the effectiveness of the intervention (Horner and Spaulding (2010)).

The SCED approach is used in all experiments conducted for this work. Because people with I/DD as a group are highly heterogeneous, and the available number of students is small, it is impractical to conduct a large scale study for this research. SCED is appropriate for students with disabilities because it provides a statistically sound measurement of the impacts of the intervention (i.e., intelligent robot and augmented reality instruction) on a small sample set ($n \geq 3$). To accommodate attrition, participant pools on the order of $n = 5$ students were targeted for this research.

4.6.3 Inter-observer agreement

To ensure reliability of results collected, all sessions were observed by a second observer, either in person or from recorded video. Both observers independently collected inter-observer agreement (IOA) data. IOA data were collected during a minimum of 60% of baseline, intervention, and maintenance sessions for each participant. Observers independently and simultaneously recorded the number of steps performed correctly on each vocational task. IOA was defined as the proportion of positive agreements, i.e., the number of agreements divided by the total number of observations. The IOA threshold was set 90% or greater; if the IOA was lower than 90%, then the two observers met and reviewed all test items and responses to achieve consensus.

4.7 Experiment Overview

Table 4.2 provides an overview of the experiments conducted, including prompt strategy (Section 3.3), experimental design employed (Section 4.6.2), and the relation between steps (Section 3.3.5). Five studies were undertaken, three using the ARI system and two with the IRI. The *copy machine* study (Section 4.8.1) using an office copier was selected as a vocational task. In the *student account statement* study (Section 4.8.2) students learned to access account information online, which is relevant to both independent living and employment. The ARI was used to teach foundational prerequisite skills in the *geometric reasoning* study (Section 4.8.3) using puzzle blocks in preparation for the IRI *geometric reasoning and assembly* (Section 4.9.2) study, which both taught different levels of object placement, orientation, discrimination, and assembly. In the *making change* study (Section 4.9.1), the students were taught how to present the correct amount of change for a purchase, which is also relevant as both a job and independent living life skill.

Table 4.2: Experiment overview.

Platform	Experiment	Prompt	Design	Relation
ARI	Copy machine	Self-directed	Multiple baseline across skills	Forward chaining
	Student account	Self-directed	Multiple baseline across skills	Forward chaining
	Geometric reasoning	CTD	Multiple baseline across skills	Discrete and forward chaining
IRI	Making change	SLP	Multiple baseline across particpants	Discrete and forward chaining
	Geometric reasoning and assembly	SMP	Multiple baseline across skills	Backwards chaining

4.8 Augmented reality instruction experiments

Three skills were selected for instruction using the ARI system: using an office *copy machine* (Section 4.8.1), retrieving a copy of one's *student account statement* online (Section 4.8.2), and performing *geometric reasoning* tasks with puzzle blocks (Section 4.8.3).*

*An example video of the ARI system in action for the *copy machine* skill is available online at https://youtu.be/iGs7CX-DSZY.

Participants

Three participants participated in all three experiments: two males and one female ages 19-29 and IQ scores ranging between 57 and 63. The students selected for these studies were unable to perform any of the tasks independently before instruction.

Experimental design

To determine if a causal or functional relation exists between the delivery of the independent variable (IV) (i.e., the instruction system), and significant increases in the dependent variable (DV) (i.e., the acquisition and maintenance of the skills required), a multiple probe across skills design Gast and Spriggs (2010) was employed to evaluate the relationship between the intervention and each student's ability to correctly perform each vocational task. The context-aware intervention was systematically introduced across three vocational tasks. First, the ARI was used to teach the students how to make double-sided copies in the *copy machine* study. Then, after achieving mastery for that skill, the ARI was used to teach the *student account statement* skill. Finally, after mastering the first two skills, the students used the ARI in the *geometric reasoning* study.

This design allows for evaluation of intervention effects while controlling for threats to internal validity (i.e., that the learning is due to the instructional intervention) in situations where withdrawal of skill knowledge is not possible. By introducing the intervention subsequently across a minimum of three replications of skills or tasks, the possibility of any observed change occurring due to extraneous factors (e.g., practice or history effects) is eliminated, which allows for experimental control and the establishment of a causal relationship (Horner et al. (2005)).

Experiments comprised two or three phases. In the baseline phase, probe sessions were performed to collect baseline data on each participant's performance of each target task. For two of the studies, *copy machine* and *student account statement*, the intervention was divided into an instructional phase, in which our system was used to provide instruction for correct performance of each step, as well as familiarize the students with the process of asking the device for assistance; and an independent phase, in which students were instructed to perform the task independently, and access the system for assistance with any step of the task as needed. The third study, *geometric reasoning*, consists of a series of geometric tasks, so no instruction phase was used. Students were asked to perform each prompt independently, where independence was defined as the performance of each task without correction.

4.8.1 Copy machine

The copy machine task consists of making the correct number of double-sided copies of a document. This skill will be useful in office vocational work, which is important in the context of the issues of low employment and poor wages faced by people with I/DD.

A commercial Konica-Minolta Bizhub 363 copy machine was used for this task. The menus of the copy machine were sufficiently complex to make using them non-intuitive for the population in this study. This task was split into 7 steps (Figures 4.14 - 4.20). The task was introduced to the students by presenting them with a two-page document and asking them to make three double-sided copies using the copy machine.

During intervention, the students wore the Glass AR device. First, they were directed to use the ARI for two complete sequences, asking the device for help at each step; this was referred to as the "instruction" phase of intervention. Then, students were told to use the ARI only as needed; this was referred to as the "independent" phase of intervention.

For image classification, image inputs were segmented into two subselected regions: the copy machine image and the region of the image that contains the number of copies entered. Two linear SVM classifiers were trained prior to the instruction, one for each subselection, with 100 representative images for each step / copy count, and

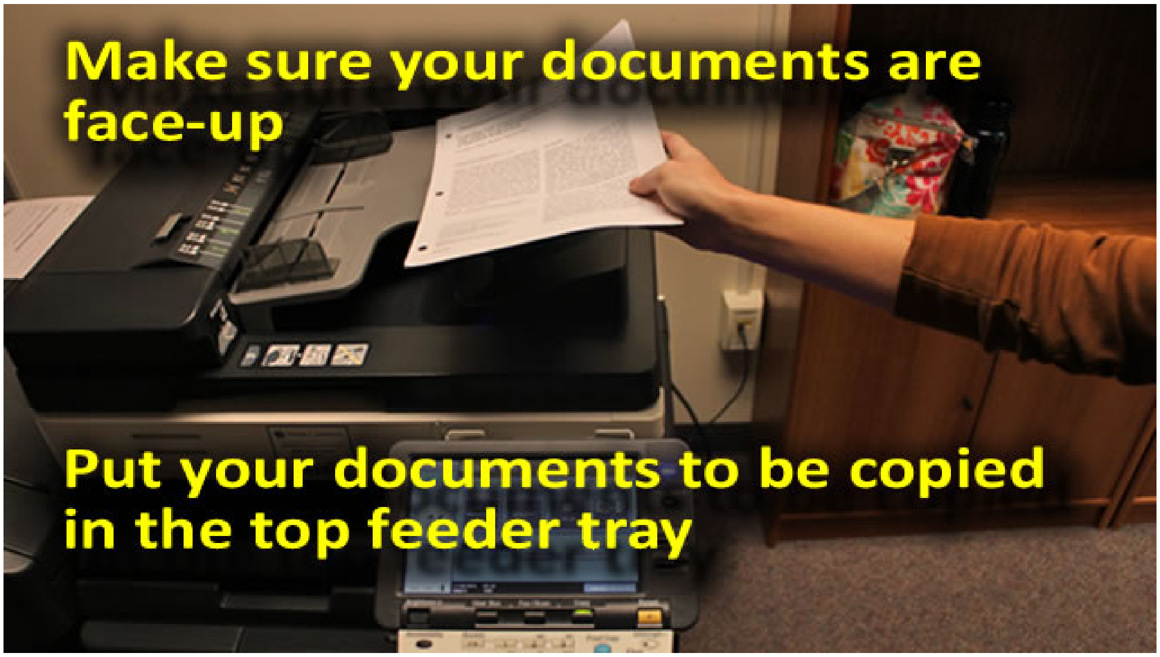


Figure 4.14: Example image annotated with text and instructions pushed to the Google Glass AR device for step 1 in the *copy machine* study, in which the participants learned how to navigate a copier's menus to make double-sided copies.

used in parallel. The combined results of the two classifications were then used to decide which step was currently shown. Testing cross-validation showed accuracy above 95% with cost $C = 2.6$ tuned manually via grid search.

4.8.2 Student account statement retrieval

For this task the students were asked to download a copy of their prepaid student card account statement. The ability to access and retrieve electronic account information is a critical skill in both employment and independent living. For this task, students were asked to use a computer to download a copy of their prepaid student card account ("VolCard") statement.

This task has 9 steps (Figures 4.21-4.28). As with the *copy machine* skill, intervention was divided into instruction and independent phases. The input for the context aware classification for this task was the section of the image containing the computer screen. A single linear SVM classifier was used, trained prior to the

Figure 4.15: Example image annotated with text and instructions pushed to the Google Glass AR device for step 2 in the *copy machine* study.

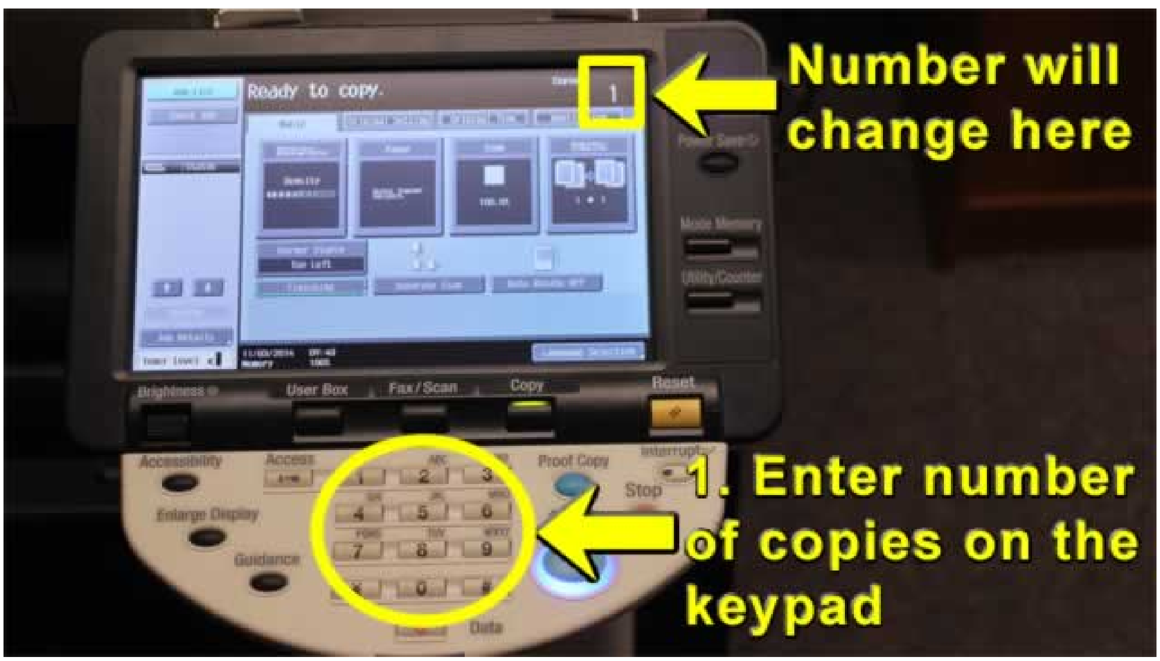


Figure 4.16: Example image annotated with text and instructions pushed to the Google Glass AR device for step 3 in the *copy machine* study.

Figure 4.17: Example image annotated with text and instructions pushed to the Google Glass AR device for step 4 in the *copy machine* study.

Figure 4.18: Example image annotated with text and instructions pushed to the Google Glass AR device for step 5 in the *copy machine* study.

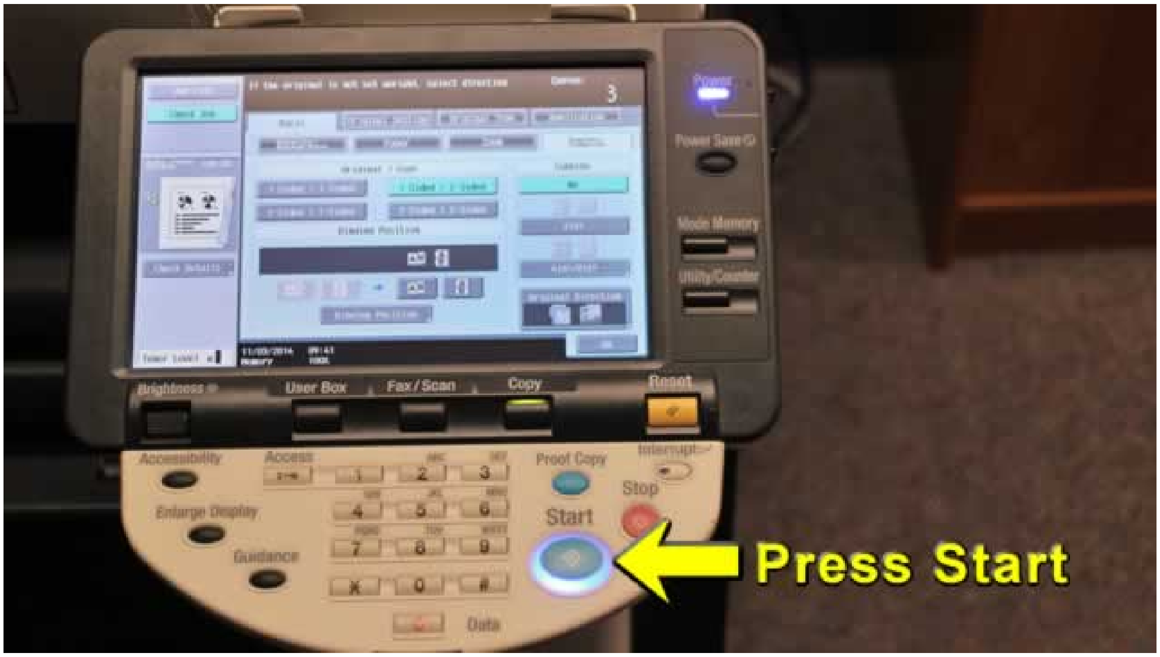


Figure 4.19: Example image annotated with text and instructions pushed to the Google Glass AR device for step 6 in the *copy machine* study.

Figure 4.20: Example image annotated with text and instructions pushed to the Google Glass AR device for step 7 in the *copy machine* study.

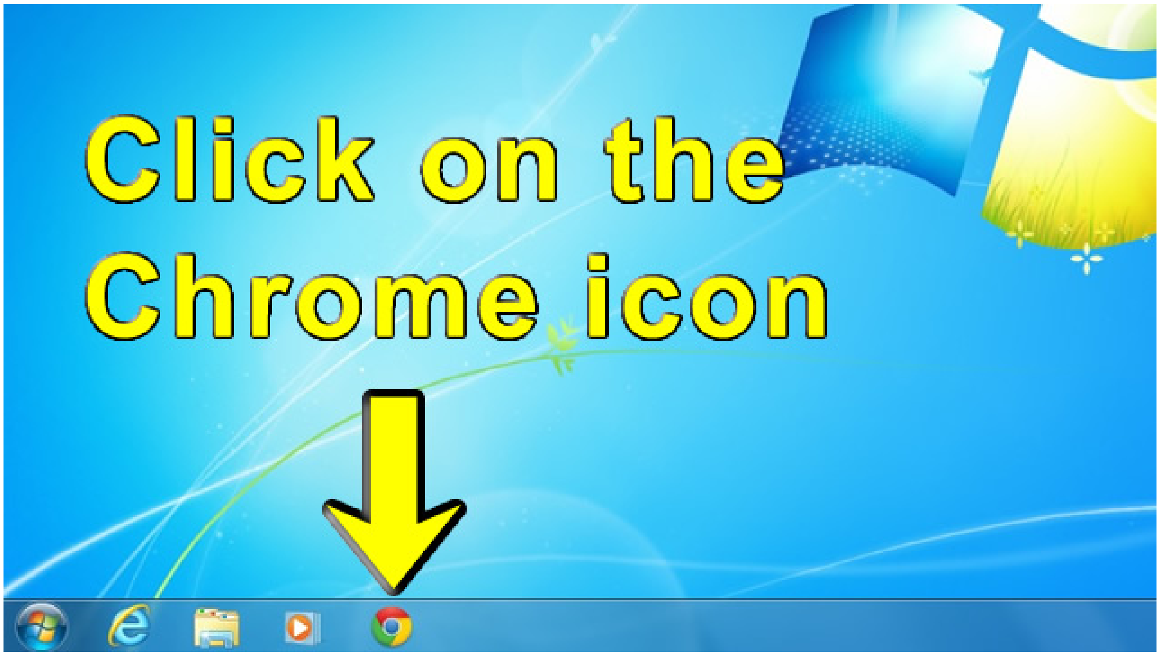


Figure 4.21: Example image annotated with text and instructions pushed to the Google Glass AR device for step 1 of the *student account* study, where the task was to download a copy of their prepaid student card account statement.

instruction with 20 images from each step. Fewer training images were required than for Experiment 1 because being in front of a computer, the view from the students' perspectives was more constrained than at a copy machine. The SVM cost was set to $C = 5.9$ with a cross-validation testing accuracy of 93%.

4.8.3 Geometric reasoning with tangrams

For the third experiment, a geometric reasoning task was selected. The ability to manipulate objects, understand relative object placement and orientation, and use the same set of parts, shapes, or items to follow novel assembly instructions is important in many job settings. Furthermore, this task was chosen to teach prerequisite skills for the IRI *geometric reasoning and assembly* study (Section 4.9). To begin this experiment, students were directed to ask the ARI, "What's first?," then, to ask the ARI, "What's next?," after each successive step. Unlike the *copy machine* and *student account statement* studies, which taught a chained task, the goal of this study was

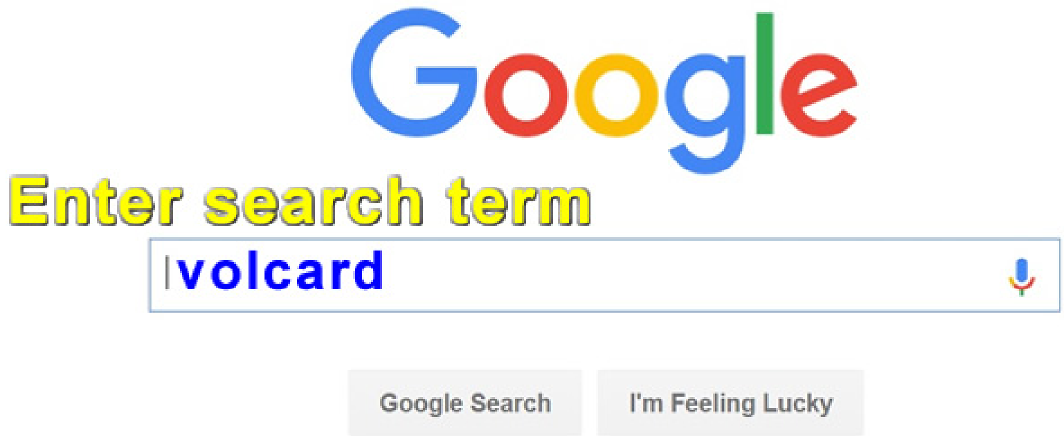


Figure 4.22: Example image annotated with text and instructions pushed to the Google Glass AR device for step 2 of the *student account* study.

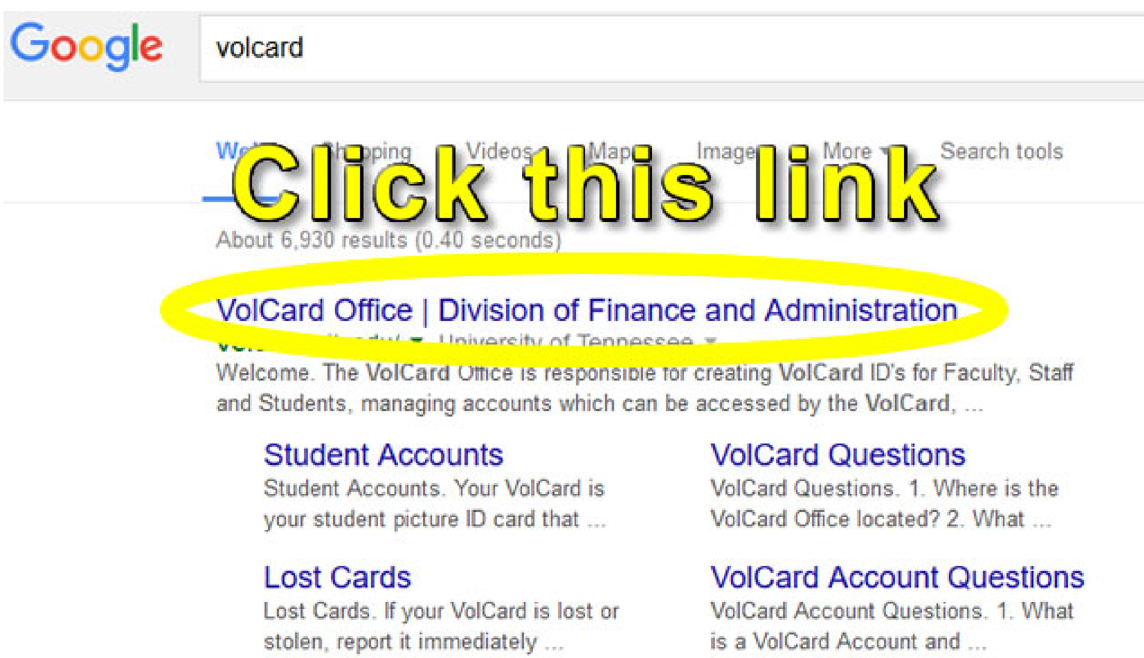


Figure 4.23: Example image annotated with text and instructions pushed to the Google Glass AR device for step 3 of the *student account* study.

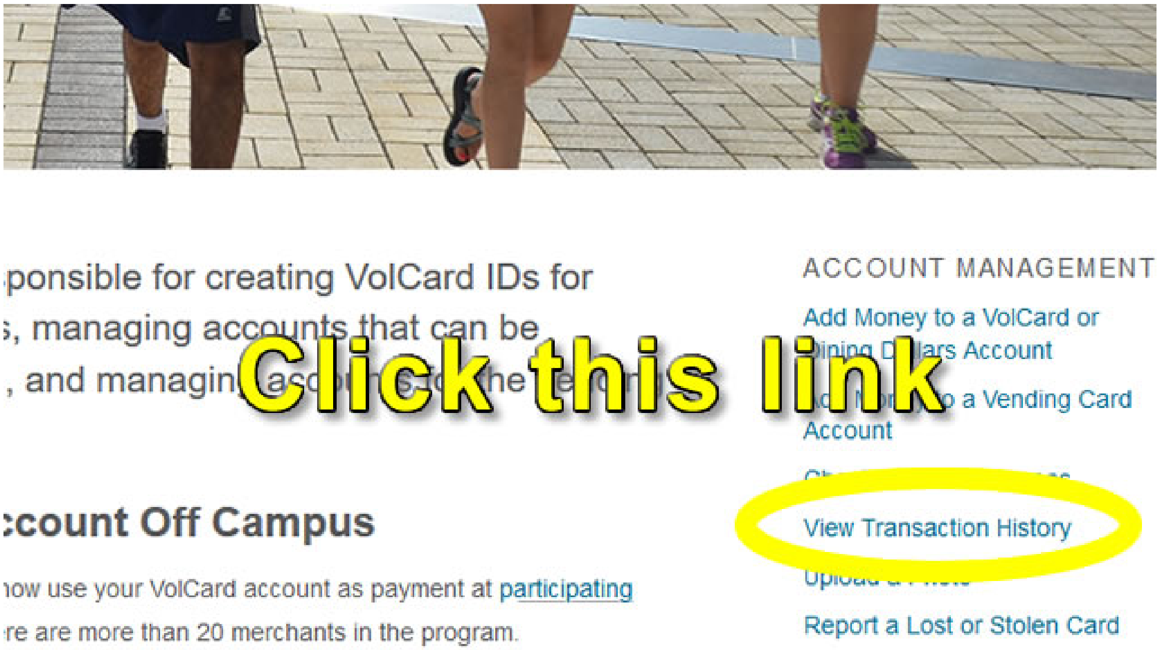


Figure 4.24: Example image annotated with text and instructions pushed to the Google Glass AR device for step 4 of the *student account* study.

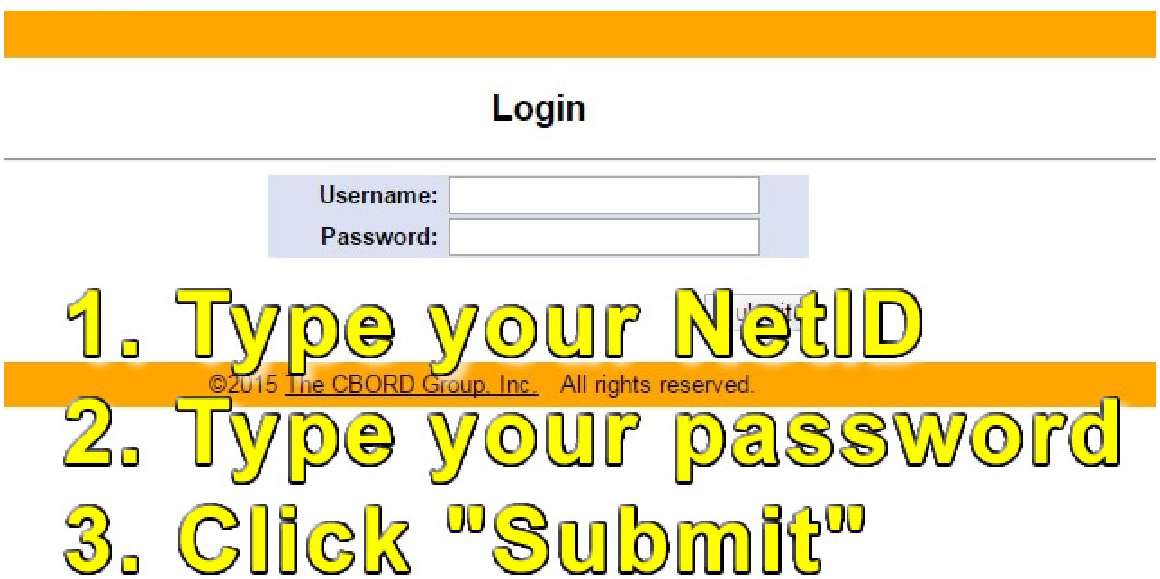


Figure 4.25: Example image annotated with text and instructions pushed to the Google Glass AR device for step 5 of the *student account* study.

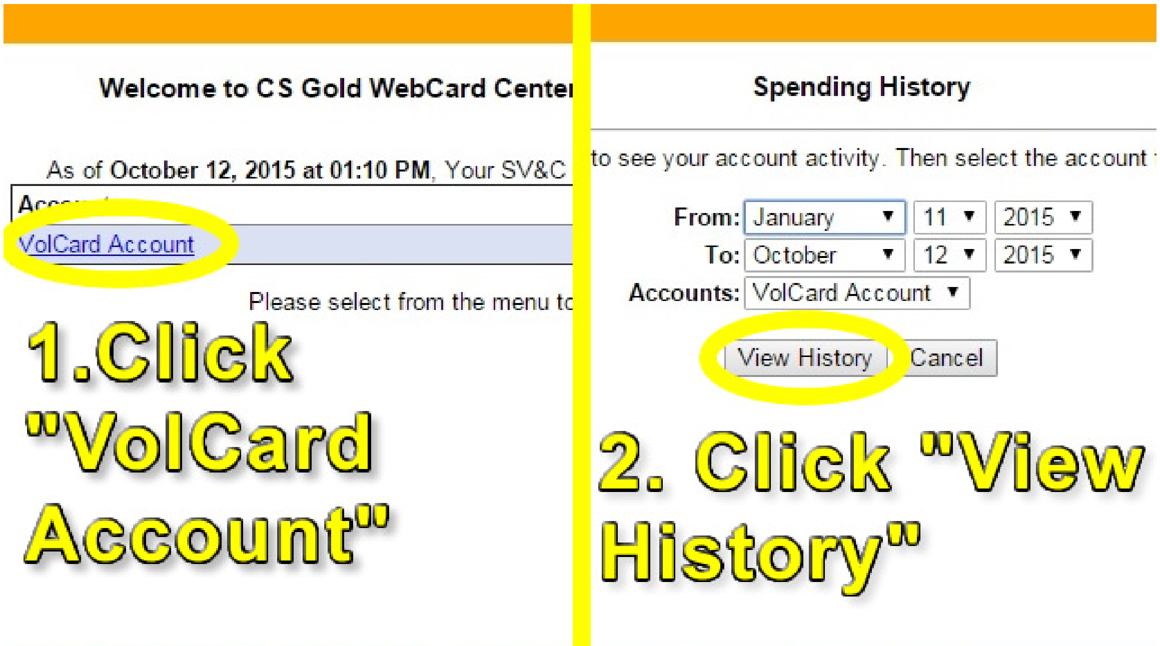


Figure 4.26: Example image annotated with text and instructions pushed to the Google Glass AR device for steps 6 and 7 of the *student account* study.

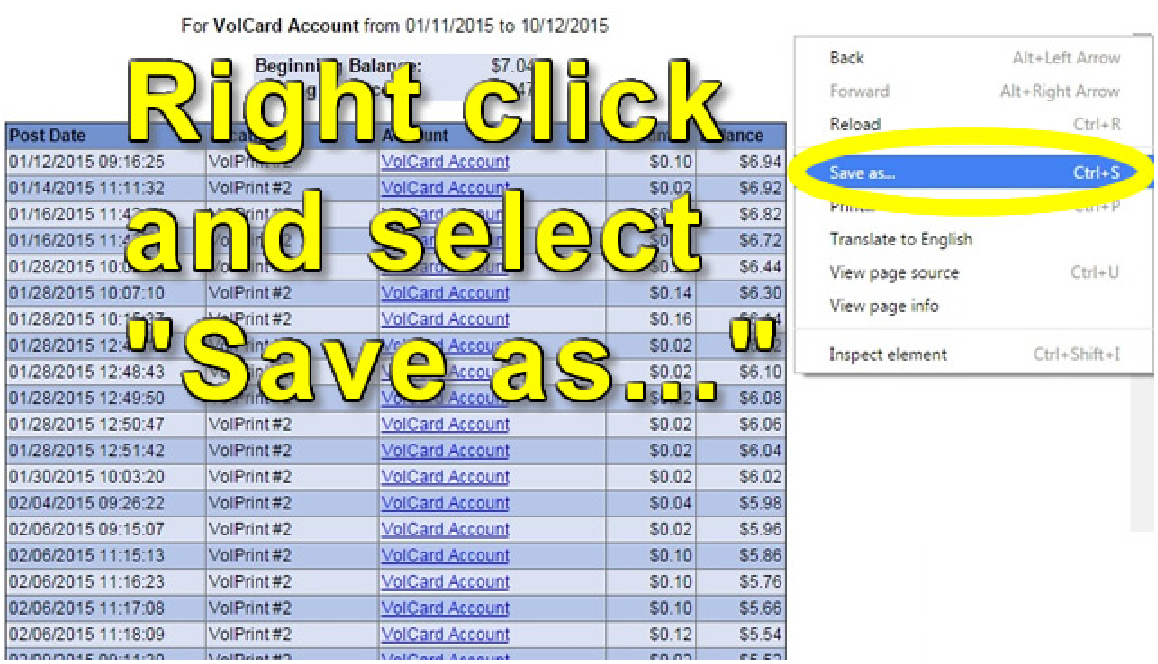


Figure 4.27: Example image annotated with text and instructions pushed to the Google Glass AR device for step 8 of the *student account* study.

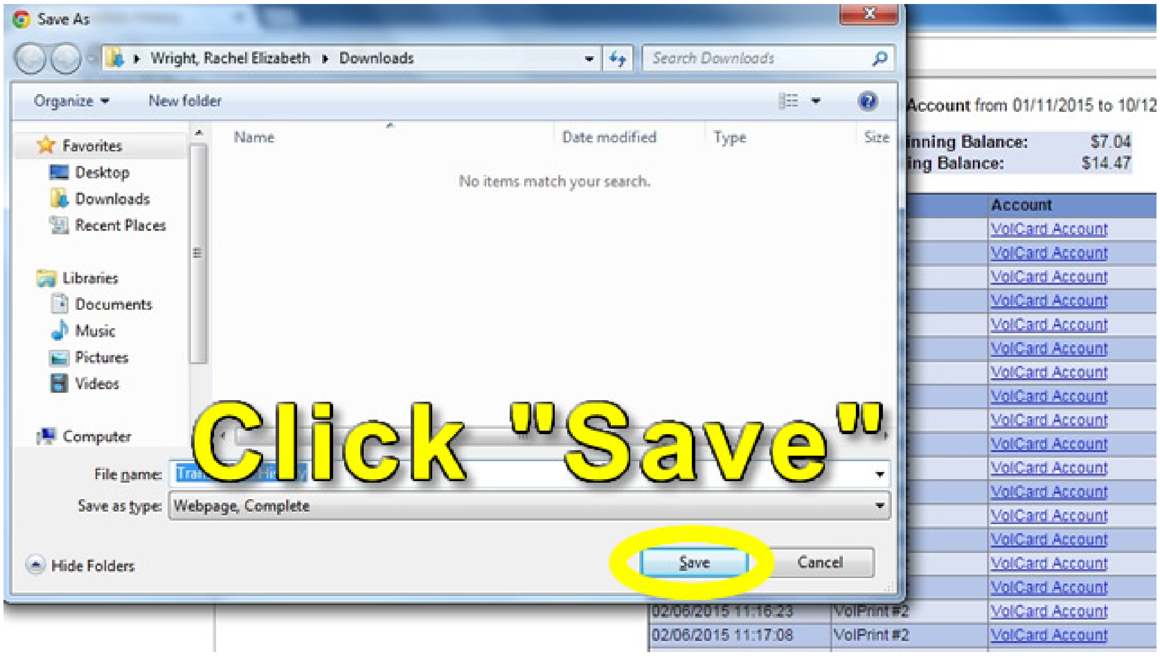


Figure 4.28: Example image annotated with text and instructions pushed to the Google Glass AR device for step 9 of the *student account* study.

for each student to perform each individual step independently without correction. Therefore, no instruction phase was necessary, and the order of steps was changed in between trials to prevent memorization.

For this task, students were asked to identify, place, and rotate colored geometric shapes ("tangrams") on a table. This task has 20 steps, three of which are shown in Figures 4.29-4.32, with an example correction in Figure 4.31. The full outline for the steps is located in Appendix A.1. The difficulty of the steps escalated from simple shape identification to relative placement and rotation, to assembling more complicated shapes. Because this task involves complicated, fine manipulation of shapes, image inputs were processed by first finding color contours, then classification was performed with a decision tree to determine the correct step and corresponding instructional prompt. The knowledge base for this task was generated by inputting correct puzzle solutions in the form of captured images, extracting the object locations programmatically, and combining those locations with a known placement sequence.

Given an image input for classification, the ARI system uses that knowledge base, combined with rules for rotation and placement, to generate the correct prompt for the next step.

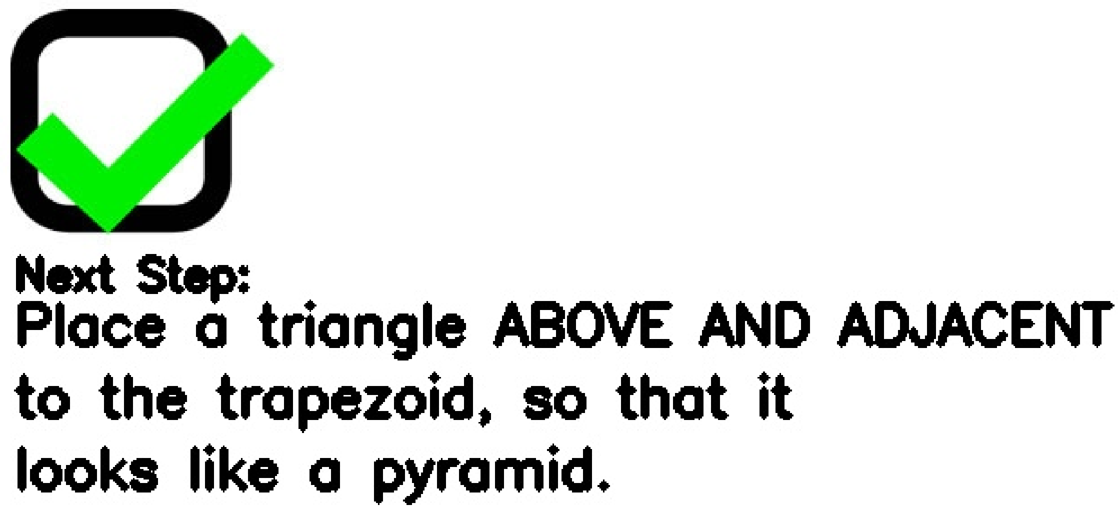


Figure 4.29: Annotated image example for the *geometric reasoning* study. This image shows a prompt to place a triangle above the trapezoid placed in the step prior.

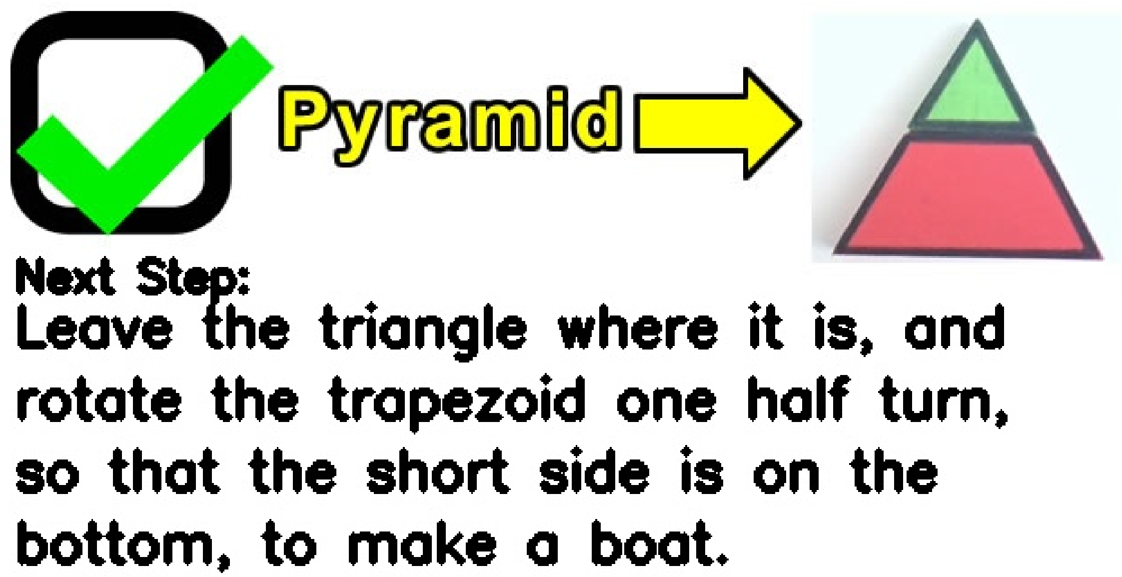


Figure 4.30: Annotated image example for the *geometric reasoning* study. This image shows the reinforcement for creating a pyramid, along with the instructions to rotate the trapezoid one half turn.

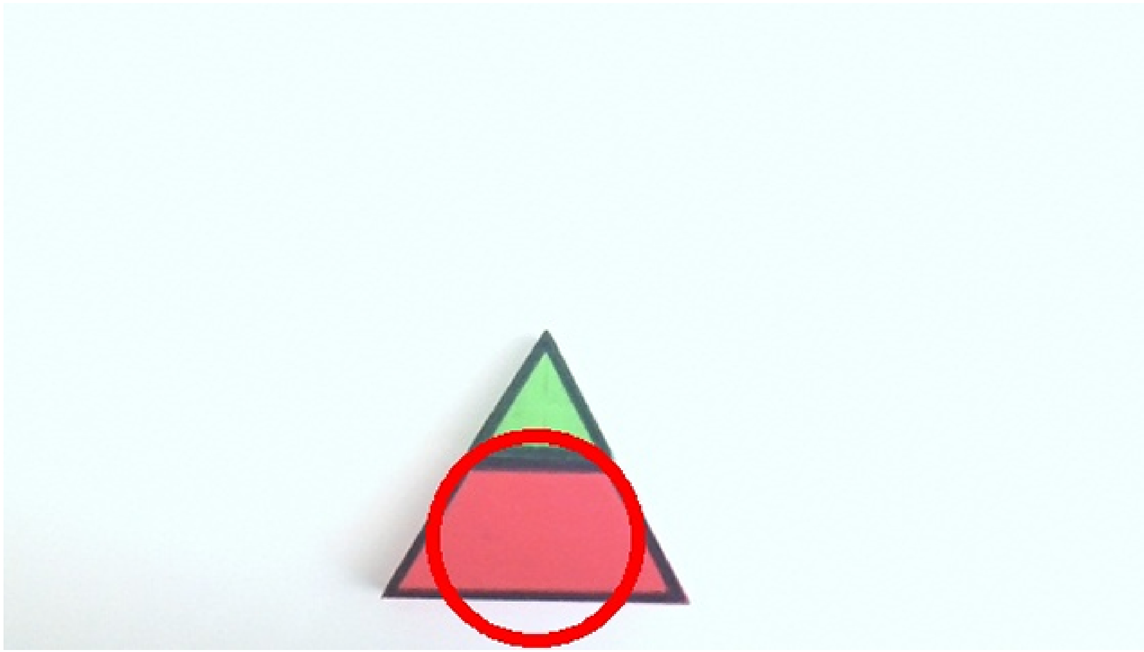


Figure 4.31: Annotated image example for the *geometric reasoning* study. This image shows an annotation of an image taken by a participant where the prior instruction was not followed, which was accompanied by a dynamically-generated audio prompt to rotate the highlighted trapezoid one half turn.

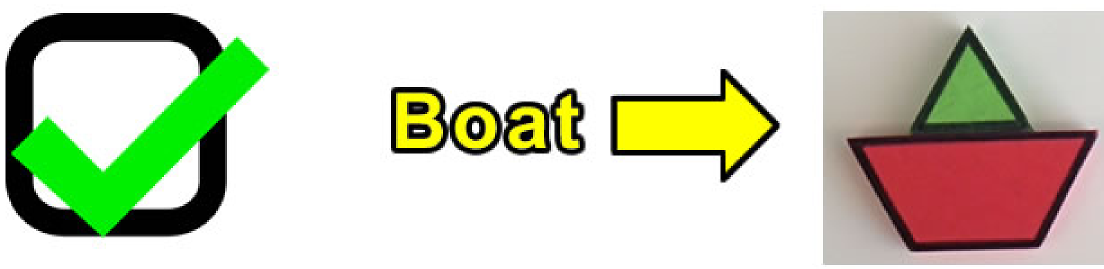


Figure 4.32: Annotated image example for the *geometric reasoning* study. This image shows the reinforcement for the prior completed step.

4.9 Intelligent robot instruction experiments

4.9.1 Making change

The first task selected for the IRI to teach is the task of *making change*. The ability to make change, i.e., to calculate the correct quantities and denominations of currency that should be exchanged after a cash transaction is made, is a valuable skill for both the workplace and independent living, and therefore fits the definition of a socially valid life skill. In this study, students learned the correct combination of coins to make change for a purchase under $1.00.

Participants

Participants in this study are three college-age female students with an IQ between 57 and 67. All three received special education services throughout school under the category of intellectual disability and earned modified high school diplomas. In addition, Student 1 has a dual-diagnosis of emotional disturbance. Prior to instruction, none of the students selected for this study were able to perform any of the steps involved in the *making change* task independently.

Experimental design

A multiple baseline across participants single case experimental design was used to determine whether a causal or functional relation exists between the delivery of the independent variable (IV) (i.e., intelligent robot instruction) and significant increases in the dependent variable (DV) (i.e., the acquisition and maintenance of the skills required to independently make correct change from purchases under $1.00). SCEDs generally involve repeated, systematic assessment of one or more IVs and DVs over time. Because withdrawal of skill knowledge is not possible, this design was selected in order to allow evaluation of intervention effects by controlling for threats to internal validity and to establish a cause-effect relation. Sequentially introducing the intervention across a minimum of three replications of participants allows for experimental control by eliminating the possibility of any observed change occurring due to extraneous factors (e.g., practice or history effects); thus, a causal relation can be established (Horner et al. (2005)).

During baseline, each student was asked to perform a series of trials related to counting amounts of change back to the experimenter for purchases of items under $1.00. No feedback, prompting, or assistance was provided to participants under the baseline condition. At least three trials were conducted until data were considered stable. Stability was defined using the "80%-20%" criteria of the stability envelope, meaning that 80% of the data points fall within 20% of the mean of baseline (Gast and Spriggs (2010)).

Upon verification that Student 1 was unable to correctly perform the skills related to making change independently, and baseline data were considered stable, Student 1 was introduced to the intervention. Students 2 and 3 continued to be assessed periodically to ensure the skills had not been learned through practice or carry over effects. Once Student 1 demonstrated an ascending trend of at least three consecutive scores per skill above baseline mean, the intervention phase was introduced to Student 2 while Student 3 remained in baseline. This continued until all three participants had reached acquisition criterion of at least 3 consecutive trials at 100% independence. Section 5.3.1 discusses the results of this study.

Task description

To set up this task, we use the configuration described in Section 4.4. We use standard U.S. coins (quarters, dimes, nickels, and pennies) and affix a colored tag to one side of each coin, which is placed facing down, towards the camera.

The general outline of an experimental trial begins with Rosie, our IRI, providing the task instruction and then asking the student to show the correct change for a dollar for a randomly selected price less than $1, which is referred to as the target stimulus. After presenting the target stimulus, the student presents an answer by placing coins into a specially delineated area of the table (an "answer box"). Rosie observes and selects the appropriate feedback at the correct time using the encoded decision making approach.

Prompt hierarchy

For the *making change* task, the System of Least Prompts (SLP) strategy was employed (Section 3.3.3) to teach the discrete task of providing the correct amount of change. A prompt hierarchy is encoded as a decision tree and incorporated into the overall cognitive framework. Four levels of increasingly intrusive prompts are defined, with unique interactions for each of the possible response types that could result in a prompt (incorrect, partially correct, and no response). Table 4.3 shows an overview of the resulting prompt hierarchy where prompts are arranged from least to most intrusive. *Verbal Cue 1* is the least intrusive; Prompt level *Direction 2* serves as the *controlling prompt*. Responses include: NR - No Response, PC - Partially Correct response, and I - Incorrect response; PC is a special case where all of the coins in the response are part of the solution, and none are not part of the solution, i.e, the student is progressing towards a solution. The script for the *making change* study is located in Appendix A.2.

Subjective Acceptance Survey

To evaluate the attitudes of the student volunteers towards learning from a robot, Likert-type scale statements and open-ended questions were used to collect subjective data. Students were surveyed both prior to and after working with the IRI system. A five-point Likert-type scale was used for each statement, and optional open-ended follow-up questions appropriate to each statement (e.g., "Why or why not?", "Please explain") were asked. To ensure a uniform understanding of the questions, surveys were performed orally, with visual aids provided for responses. The pre-assessment survey consists of 18 statements divided into 8 categories, the post-assessment 30

Table 4.3: Prompt hierarchy for the *making change* task using SLP.

Prompt Lvl.	Resp.	Prompt Description
Verbal Cue 1	NR	Verbal interaction to determine how much change is due
	PC	Verbal encouragement, verbally provide goal
	I	Same as NR
Verbal Cue 2	NR	Verbal interaction to determine which coin to begin with
	PC	Verbal encouragement, verbally provide goal + shortage between current state and goal
	I	Verbal encouragement, provide goal + excess between current state and goal
Direction 1	NR	Gesture to correct first coin, verbally provide goal
	PC	Gesture to correct next coin, verbally provide goal + shortage
	I	Gesture to coin to remove, verbally provide excess
Direction 2	NR	Gesture to each coin to add, wait until added
	PC	Same as NR
	I	Gesture to each coin to remove, wait until removed, then same as NR

statements in 14 categories, with 2-3 statements each category, and a summative analysis was applied. Section 5.3.2 discusses the results of this survey.

4.9.2 Geometric reasoning and assembly

The purpose of this study was to teach geometric reasoning skills and assembly skills, that is, the ability to consistently assemble a larger object from smaller pieces. Prerequisite skills for this study were taught using the ARI system as described in Section 4.8.3. Three skills were selected to be taught: subcomposition, symmetry, and assembly. Students learned to assemble three puzzles for each concept, for a total of nine tasks taught. Puzzles for each of the tasks are shown in Figure 4.34.

For this study, puzzle blocks were used as an instructional tool. Puzzle blocks have been shown to help "develop higher order, or critical thinking" (Clements et al. (2004)). Bohning and Althouse (1997) found that instruction with tangram puzzles "helps children develop positive attitudes toward geometry, further their shape identification and classification skills, and foster an understanding of basic geometric concepts and relationships" and provide essential developmentally appropriate experiences, which are particularly important for young children. As with the ARI *Geometric reasoning* experiment, the ability to understand these concepts can also be extremely beneficial to assembly-type tasks in workplace environments.

Puzzle blocks, as an instructional medium, were selected for several reasons. Their bright colors and variety of simple geometric shapes are appealing and easily manipulated by both humans and robots. The complexity of problems presented can be manipulated to provide simple to extremely difficult challenges for all ages and skill levels; for example, an outline of a complex shape that is not at the same scale as the physical blocks, such as in Figure 4.33, presents a puzzle that could perplex an individual with very high spatial geometric reasoning ability. Further, their flat, 2-dimensional appearance simplifies the object recognition and tracking problem.

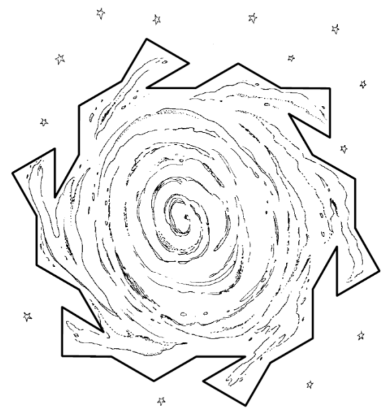


Figure 4.33: A challenging tangram puzzle.

Whereas this research uses puzzle blocks for the IRI task, Chao et al. (2010) used very basic "tangram" representations for robot-learning tasks, specifically recognition of paired configurations of puzzle blocks, and noted that the tangram domain is "interesting because it is related to the general class of assembly tasks."

Participants

This study used the same participants as the ARI study in which prerequisite skills were taught as discussed in Section 4.8.3: two males and one female ages 19-29 and IQ scores ranging between 57 and 63. The students selected for these studies were unable to perform any of the tasks independently prior to instruction.

Task description

This study uses colored geometric blocks like those shown in Figure 4.8. Because the colors of the blocks are irrelevant to the skill and not referred to by the IRI, the blocks were used colored side down (facing the camera) with their black backing side up.

In this study, the IRI teaches students three geometric skill concepts: subcomposition, symmetry, and assembly. Three puzzles are used to teach each concept, for a total of nine tasks taught. Puzzles for each of the tasks are shown in Figure 4.34.

The ultimate goal of each task is for the student to be able to assemble the puzzles for each skill independently when prompted. The general outline of instruction proceeds as follows: Rosie informs the student what puzzle they are building. As the student builds the puzzle, Rosie provides the appropriate prompt for the step the student is on, including gesturing to the appropriate placement location. If the student makes an error, Rosie first provides correction for that error using speech and gestures, then reminds them of the current step.

To simplify reasoning about the puzzle solution, the student is instructed to always place the first piece on a marked location, and build the rest of the puzzle around it.

Experimental design

A multiple baseline across skills single case experimental design was used to determine whether a causal or functional relation exists between the delivery of the independent variable (IV), intelligent robot instruction, and significant increases in the dependent variable (DV), the acquisition and maintenance of the skills required to independently construct puzzles that represented the subcomposition, mirroring, and assembly skills.

During baseline, each student was asked to construct a series of puzzles for each skill type using puzzle blocks. Puzzles used during baseline were not the same puzzles used in intervention, but were of the same variety and difficulty level. No feedback, prompting, or assistance was provided to participants under the baseline condition. As in the *making change* study, at least three trials were conducted until data were considered stable, and stability was defined using the "80%-20%" criteria of the stability envelope.

Unlike the *making change* study, which introduced the intervention across students, because there were multiple geometric skills, the IRI intervention was systematically introduced across each geometric skill. First, the IRI was introduced to teach each student the mirroring skill. Then, once a student demonstrated successful acquisition, an additional baseline reading was taken for the subcomposition skill, and the IRI intervention was introduced for that skill. Lastly, upon acquisition of the subcomposition skill, after an additional baseline reading for the assembly skill was taken, the IRI was introduced to teach the assembly skill.

Prompt strategy

In designing this experiment, the complexity of the possible configuration space for the blocks, and the potentially limitless number of possible avenues for misinterpretation, were issues identified beforehand. For these reasons, a different approach from the other studies presented here was employed with regards to prompting and chaining.

In this study, the System of Most Prompts strategy was employed, combined with backwards chaining the instruction sequence. Students were therefore taught each puzzle in reverse order. Instruction began with the last step of each puzzle, and proceeded to the first. At each step, a partially completed puzzle assembly or "jig" was placed on the table by the experimenters, and the appropriate prompt was delivered. A two-level prompting strategy was employed, consisting of the controlling prompt and the independent prompt.

At each step, the first prompt delivered was the controlling prompt, consisting of a gesture and verbal instruction using the terminology taught in the prerequisite skills (Section 4.8.3). Then, the next time the same step of the puzzle was presented, the independent prompt was delivered, which was to ask the student to complete the entire puzzle from that step forward.

Each skill's tasks were taught simultaneously; after one step was performed on a task, a different puzzle task was presented. For example, when teaching the assembly skill, a step was taught on the house, then a step on the tree, then a step on the ball puzzle. Each time the student successfully performed at the independent prompt, the step was decremented, until the student was able to perform the entire task independently.

Extra or missing piece errors, or errors in the rotation of a correct piece, were detected and dynamic corrections were presented. Possible responses include NR - No Response, CI - Correct but Incomplete, C - Correct, EE - Extra pieces present

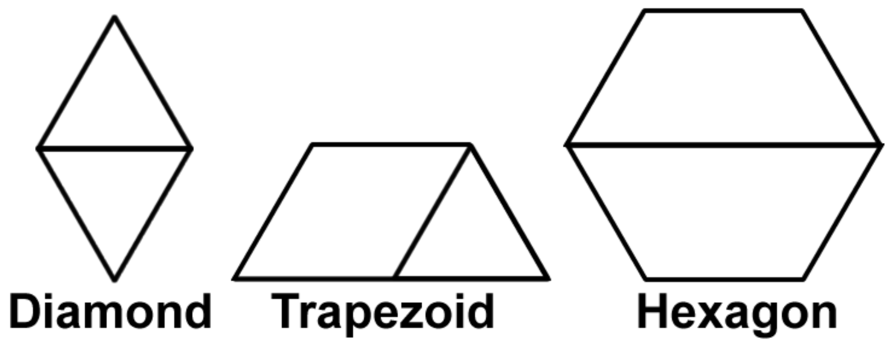
Error, EM - Error pieces Missing, and ER - Rotation Error. A table of the feedback for each response is presented in Table 4.4. Note that Rotation Error was treated as a special case of Correct but Incomplete; the prompt level was not escalated before re-prompting. The script and schedule used for these experiments is located in Appendix A.3

Table 4.4: Response feedback for *geometric reasoning and assembly* experiment.

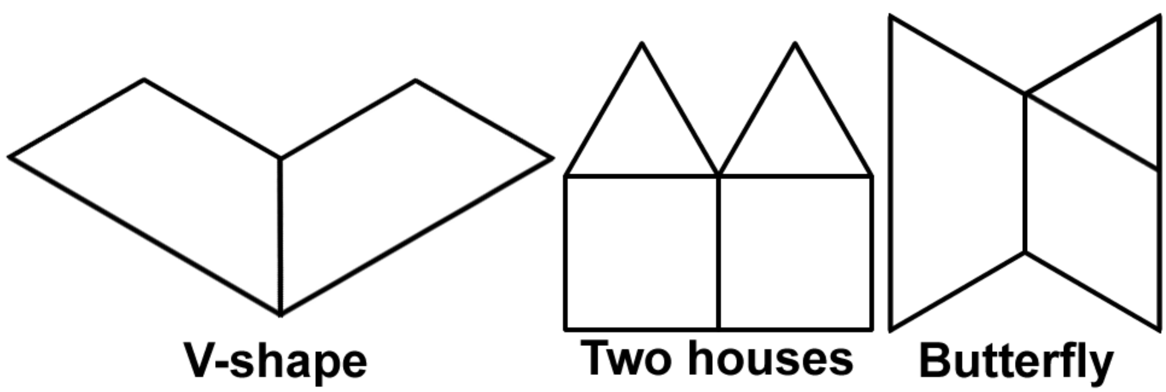
Response	**Feedback Description**
C - Correct	Reinforce skill type.
CI - Correct Incomplete	Affirm partial correctness. Deliver Independent prompt.
NR - No Response	Escalate prompt level, re-prompt.
EE - Error Extra	Request removal of extra piece with gesture. Escalate, re-prompt.
EM - Error Missing	State missing piece with gesture to location. Escalate, re-prompt.
ER - Error Rotation	Request rotation of piece by amount (quarter, half, small) and direction, with gesture. Re-prompt.

4.10 Summary

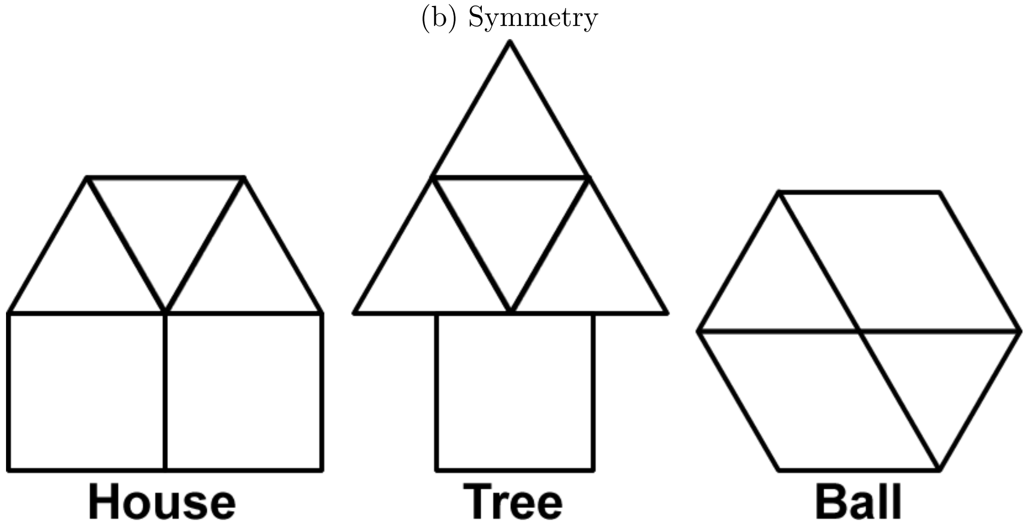
With the successful implementation of a complete IRI and ARI system, including accurate and fast object tracking, easily interpretable gestures for the IRI, and a contextually aware ARI system using supervised learning, experiments teaching five skills to students with I/DD have been conducted. Chapter 5 gives the results of these experiments.



(a) Subcomposition



(b) Symmetry



(c) Assembly

Figure 4.34: Puzzle tasks for geometric reasoning skills.

Chapter 5

Results

This chapter presents the results of experiments detailed in Sections 4.7, 4.8, and 4.9. Results show success in ARI instruction of the *copy machine*, *student account statement*, and *geometric reasoning* skills (Section 5.2) and IRI instruction on the *making change* (Section 5.3) and *geometric reasoning and assembly* (Section 5.4) tasks.*

5.1 Interpreting the graphs

The students' performance results are presented in graph format in the following sections. Figures 5.1-5.7, each contain three sub-graphs. These subgraphs represent skills or students, depending on whether the design was across participants or skills. Each subgraph has dashed lines to separate the phases of the study. The dashed lines are connected across subgraphs to delineate data from each phase. Phases usually include Baseline and Intervention; in the ARI studies, the Intervention phase was split into Instruction and Independent for two of the skills, since first the students were asked to perform the task using the intervention 100% of the time, then to perform the skill independently unless they required the intervention. In all graphs,

*Results from collaborative work shown in Sections 5.2 and 5.3 have been published in Reardon et al. (2015) and Reardon et al. (2016).

the y-axis represents the performance of the student, usually in terms of independent performance of the behavior, or a score such as percent correct responses. The x-axis represents the trial number. In series observations, such as in baseline and intervention, the observation dots are connected with a line to highlight any trend, whereas in probes (i.e., to re-check the baseline performance prior to intervention, or re-check the student's retention of the learned skill) dots are shown as unconnected individual readings.

5.2 Augmented reality instruction

The results show that with the provided intervention all students were able to achieve skill acquisition to mastery for all three skills (*copy machine*, *student account statement*, and *geometric reasoning*), defined as independent performance at 100% correct. As shown in Figures 5.1, 5.2, and 5.3 in the Baseline phase all students were confirmed unable to perform the skills independently prior to intervention. Following the training phase, in which the students were required to use the Glass for all steps, students rapidly acquired the skills. In the *copy machine* and *student account* experiments, none of the three students could perform a single step of the skill prior to instruction. In the instruction phase, students were required to use the ARI to perform each step, and therefore were 100% correct. Students 1 and 2 were able to demonstrate the *copy machine* skill flawlessly in the Independent phase, whereas Student 3 required assistance from the ARI for one step, three out of the first four Independent trials. The *student account* skill was more challenging to Students 1 and 3, perhaps because it was lengthier; these students required minimal assistance from the ARI again before reaching independence. For the *geometric reasoning* skill, each student had some knowledge in Baseline (for example, shape names). No Instruction phase was required in the *geometric reasoning* experiment, as each step's direction was delivered after the completion of the previous step. Student 1 required more instruction than the others, but all were ultimately able to perform the skill.

It is very likely that the rapid gains in skill acquisition can be attributed to the self-directed nature of the intervention system; providing the participants with the ability to control their own prompts, combined with the intelligent AR environment, may have contributed to increased engagement and efficiency in learning the tasks. In this way, because students were able to take control of their own learning experience, they were able to learn more efficiently. A genuine level of satisfaction with the experience on the part of the students was observed, which could also be attributed to the efficiency of the experience.

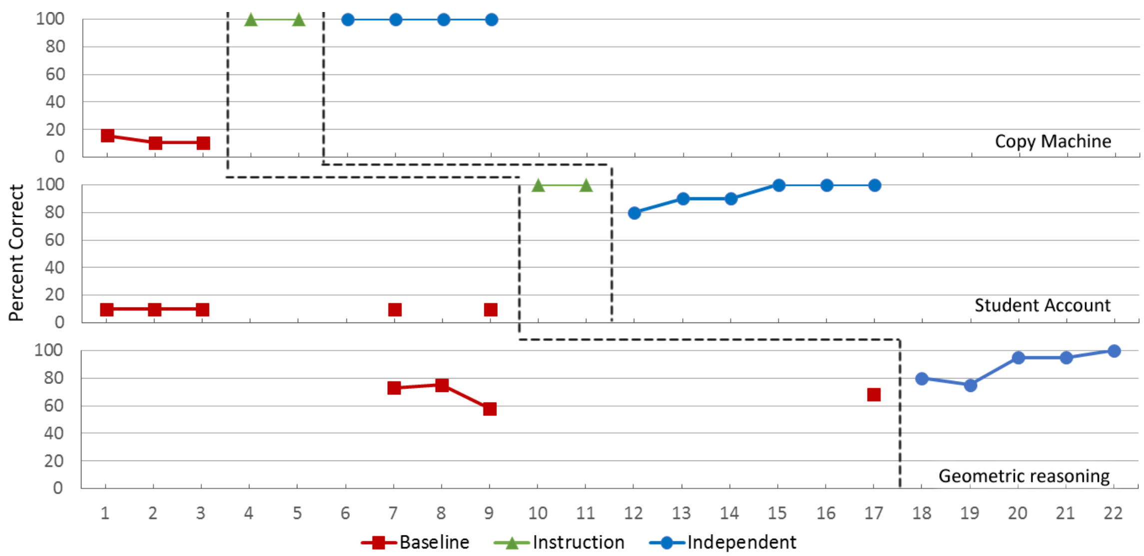


Figure 5.1: Results for ARI experiments for Student 1. Data is collected in three phases, shown separated by dashed lines: baseline, measuring each participant's performance prior to instruction; instruction, in which the ARI system was used to teach each step; and intervention, where students attempted to perform the task independently, only using the system as needed.

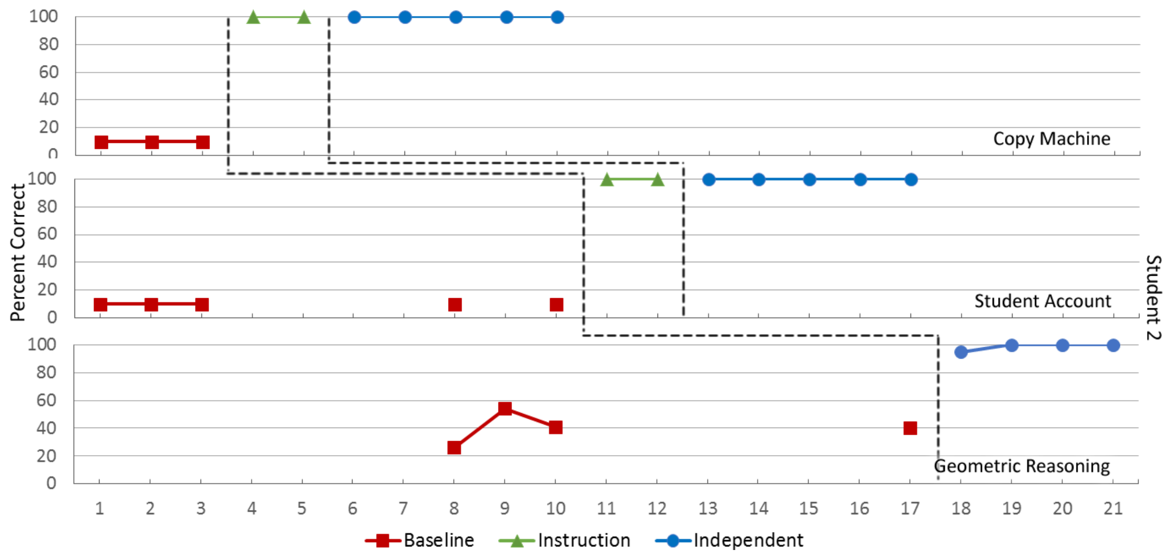


Figure 5.2: Results for ARI experiments for Student 2.

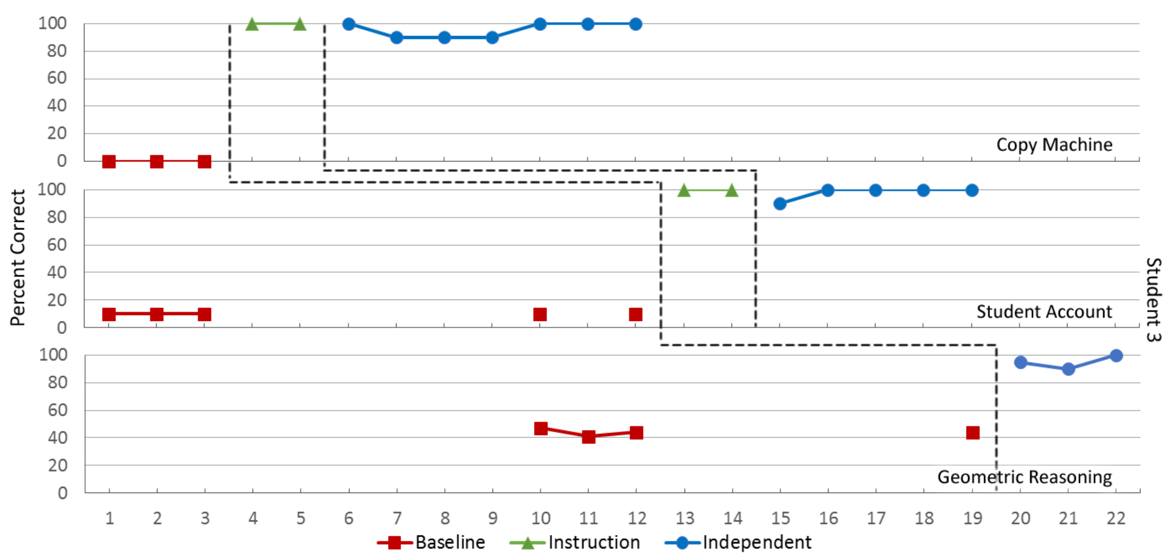


Figure 5.3: Results for ARI experiments for Student 3.

5.3 Making change

5.3.1 Participant performance

The task of *making change* is treated as two sub-skills: Skill 1, the ability to identify the correct amount of change for a given price and Skill 2, the ability to provide the correct amount of change using the least amount of coins possible. Rosie provides instruction for both sub-skills, with calculator instruction for Skill 1.

To score the participant's performance, points are assigned on a weighted scale based on prompt levels, as in Ault and Griffen (2013), using a 100-point scale. If a student responds 100% independently, a score of 100 points is recorded. Because each solution has n coins, for Skill 2 there are at most n possible controlling prompts. So, for each correct coin placement step performed prior to the delivery of the controlling prompt, a score of $100/n$ points is recorded. In the worst case, where a student does nothing and Rosie directs each coin (Direction 2 in Table 4.3), a score of 0 points is recorded. For each less intrusive prompt, points are deducted in proportion to the intrusiveness: $-50/n$ for prompts regarding the value of the coins, $-25/n$ for general prompts, and $-5/n$ for prompts related to an inefficient combination (e.g., 5 pennies vs. a nickel). For Skill 1, an independent correct response was given 100 points, a correct response with prompting/instruction was given 50 points, and a 5 point deduction was applied for each missed response in the calculator instruction sequence. Figure 5.4 shows results for each student's performance across baseline, Skill 1, and Skill 2 conditions.

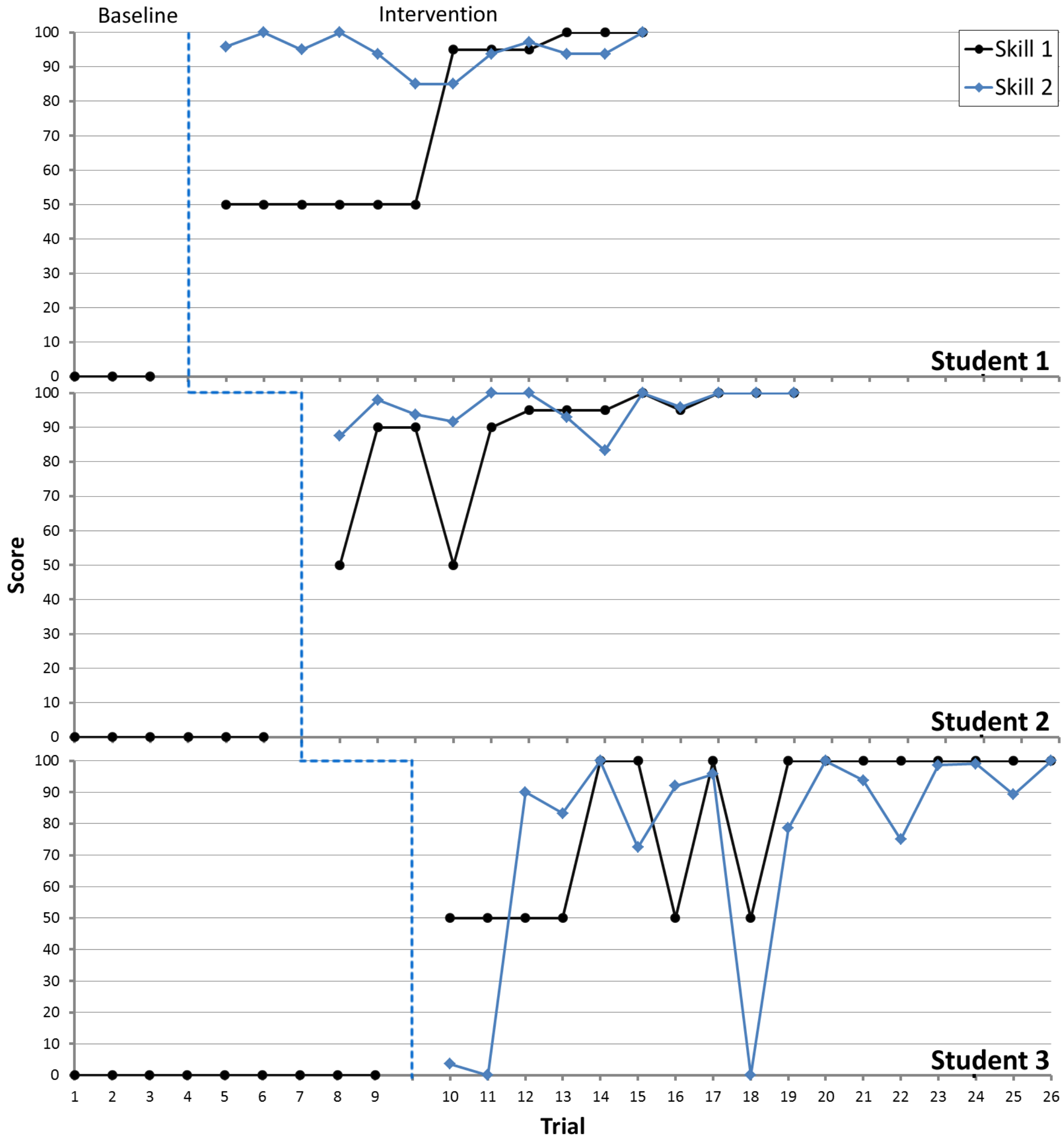


Figure 5.4: Results from the *making change* experiment using multiple baseline across participants. Data is collected in two phases, shown separated by dashed lines: baseline, measuring each participant's performance prior to instruction; and intervention, where students performed the task with instruction as needed. The skill was divided into two sub-skills: 1) identifying the correct amount of change and 2) providing the correct combination of coins.

With the IRI the students were able to achieve skill acquisition to mastery, defined as 100% correct performance of Skill 1 following calculator instruction and 100% independence of Skill 2. While data collected during baseline conditions verified that the participants were not able to perform the making change task, specific prerequisite knowledge was not assessed, such as coin value identification and the ability to follow Rosie's directions. Student 3 demonstrated more limited understanding of these prerequisites than Students 1 and 2, and therefore had a steeper learning curve to achieve criteria for acquisition and mastery. Such considerations were applied to the successive robot interventions for people with I/DD, as discussed in Section 6.2.5.

5.3.2 Subjective acceptability survey

An acceptability study was conducted to examine the opinions of students under Rosie's tutelage. Survey responses were scored from -2 (Strongly Disagree) to 2 (Strongly Agree), and averaged across categories. The full surveys are available in Appendix B. In Table 5.1, we see categories from the pre- and post-instruction Likert-type survey results. The initial results of the assessment of students' opinions prior to working with a robot instructor showed mixed enthusiasm for the experience; however, post-instruction results show a positive opinion of the overall experience and performance of the robot. Compared to the students' lower levels of willingness to work with a robot pre-instruction, the students showed greater willingness to work with Rosie again. They also trusted the robot, were willing to obey the robot's instructions, and found the experience good overall. Interestingly, Student 3, who had the most difficulty, also gave the experience the lowest scores for how easy it was to learn, and how much she trusted the robot. Regarding the mixed ratings of the usefulness of the robot's gestures, one possible explanation for the lower ratings of Students 1 and 3 is that despite giving very positive feedback overall, Student 1's higher affinity meant she had little need for the IRI's gestures, whereas Student 3 felt

Table 5.1: Acceptance survey summative results. Questions bolded for discussion.

Category (Pre-Instruction)	S1	S2	S3
Do you like computers in general?	2	1	0
Do you like robots?	**1**	**2**	**0**
Have you been exposed to robots before?	**-0.33**	**0.67**	**-1**
Are robots useful?	**1.5**	**1.5**	**0.5**
Would you learn from a robot?	**-0.5**	**1**	**1**
How comfortable are you with the skill?	2	1.5	0
How well do you think you perform the skill?	2	0.5	0.5
Category (Post-Instruction)	**S1**	**S2**	**S3**
Was the robot good or bad overall?	**1.5**	**1.5**	**1.5**
Do you view the robot as an embodied intelligence?	0.75	1.25	0.75
Did the robot seem to understand your actions?	0	1	1
Was the robot knowledgeable?	**1.5**	**1**	**2**
Did you trust the robot's instructions?	**1.5**	**1.5**	**0.5**
Did you follow the robot's instructions?	**2**	**1**	**1.5**
Was the robot easy to learn from?	**2**	**1**	**-0.5**
Was the robot's speech clear?	2	1	0.5
Were the gestures the robot made useful?	**-0.33**	**1.6**	**-0.67**
How comfortable are you with the skill?	0	1	1
How well do you perform the skill?	0	1	1.5
Do you feel like the robot did a good job?	**1.5**	**1**	**1.5**
Would you work with the robot again?	**2**	**1.5**	**2**

more frustrated by the longer time to acquire the skill, as reflected in her score on ease of learning question and as discussed above.

5.3.3 Open question responses

Open-ended responses were solicited after each Likert-style survey question. Students were not required to give responses to every open question; Tables 5.2-5.7 show only the questions with responses.

The open-ended question answers highlighted their willingness to work with a robot in the future. One student had several suggestions for additional teaching tasks. Students could also envision the advantages of a robotic instructor, with comments about the ability to "give [human] teachers a break," and the increased ease of learning from a robot, because where teachers move at the pace of the class the "pace could be better" with a robot. Student 3, who had the most difficulty with the task, noted that sometimes the robot "had problems." This perception was perhaps due to a combination of Student 3's lesser experience with prerequisite skills, as noted in Section 5.3.1, and the unusual difficulty the speech recognition system had understanding her speech, discussed in Section 6.2.4. However, Student 3 also enjoyed the experience enough to note that the IRI "helped alot" and that she's "really warming up to her [the robot]."

Table 5.2: Pre-assessment survey open responses for Student 1.

#	**Statement**	**Open-ended response**
1	I like computers.	They allow me to connect with family.
3	I think a robot could make a good teacher.	Sometimes they can, but they can't get all the facts.
4	I have difficulty making change with coins.	I understand how.
6	I don't think I'd enjoy working with a robot.	I'd love to.
8	I think robots can do many things well.	They could help you understand some things you don't.
9	Computers are useless.	They help with homework, let you email teachers.
14	Robots are bad.	They can be good or bad.
15	I am good at making change with coins.	I'm better than with bills.
16	I don't want a robot to teach me.	I'd like to.

Table 5.3: Pre-assessment survey open responses for Student 2.

#	Statement	Open-ended response
1	I like computers.	Computers help, for example searching, finding images, combining things together.
2	I have seen a robot before.	At the science museum in Oak Ridge.
3	I think a robot could make a good teacher.	They might show us what they have learned if they are programmed to learn.
4	I have difficulty making change with coins.	I know how to count, but sometimes I have difficulty.
5	I like robots.	I find them quite amazing, they can be programmed with emotional feelings.
6	I don't think I'd enjoy working with a robot.	I have a strong feeling in the future people will use robots more often.
7	I have worked or played with a robot before.	When I was a little girl I had a toy robot.
10	I think I could work together with a robot.	Good question for me. Sometimes I would have to learn what kind of programs they have. For example, Rosie on the Jetsons would sometimes lose control.
11	I know a lot about robots.	I'm still learning.

Table 5.4: Pre-assessment survey open responses for Student 3.

#	Statement	Open-ended response
1	I like computers.	For searching stuff. Them dying or slowness is a downfall.
2	I have seen a robot before.	Disney.
3	I think a robot could make a good teacher.	It would be cool.
5	I like robots.	They're different, I'd like to get to work with or build them.
6	I don't think I'd enjoy working with a robot.	They could be difficult sometimes. Transformers are bad.
10	I think I could work together with a robot.	Sometimes they are useful, sometimes they give you wrong information.

Table 5.5: Post-assessment survey open responses for Student 1.

#	Statement	Open-ended response
1	I like Rosie the robot.	It's easier to have a robot teach you.
4	Rosie couldn't tell what I was doing.	Sometimes she couldn't see some movements.
5	Rosie knew how to make change with coins.	She was better than me.
6	I didn't always follow Rosie's instructions.	I was able to understand her instruction.
10	Making change with coins makes me feel nervous	I'm afraid to give the wrong change.
11	I am better at making change with coins than I was before working with Rosie.	Rosie showed me how to make change the correct way.
14	Rosie's gestures were very confusing.	I sometimes didn't know what she was gesturing to.
16	Learning from Rosie was difficult.	It was easy to learn.
17	I did not enjoy Rosie the robot.	I did enjoy Rosie the robot.
19	Rosie did not care how I performed.	She did care.
27	After working with Rosie, I'm about the same at making change with coins as I was before.	I'm getting better.
28	Rosie was hard to understand.	She was easy to understand.
30	Rosie's instructions were not very useful.	They were useful.
31	Rosie was easy to learn from.	She was better than my last math teacher.

Table 5.6: Post-assessment survey open responses for Student 2.

#	Statement	Open-ended response
1	I like Rosie the robot.	She helps teach students to count money.
3	Rosie wanted me to learn.	She can count money like most people.
5	Rosie knew how to make change with coins.	She pointed at coins and you can tell she can count money really good.
8	I would like Rosie to teach me a new task in the future.	A robot like Rosie could give teachers a break.
18	Rosie appeared aware of me.	She could listen and be positive.
31	Rosie was easy to learn from.	The pace could be better with a robot than a teacher.

Table 5.7: Post-assessment survey open responses for Student 3.

#	Statement	Open-ended response
1	I like Rosie the robot.	First it was weird working with a robot, then she helped me count change. She helped a lot. I'm warming up to her.
3	Rosie wanted me to learn.	The way she told me wrong or right every time helped me learn to count change.
4	Rosie couldn't tell what I was doing.	Sometimes she would mess up.
5	Rosie knew how to make change with coins.	She will point them out.
7	I trusted Rosie to give me the right instructions.	She helped with the calculator and pointing.
8	I would like Rosie to teach me a new task in the future.	It's cool working with a robot.
14	Rosie's gestures were very confusing.	Sometimes the arm moved wrong or she repeated things.
16	Learning from Rosie was difficult.	When I first started, I was confused.
19	Rosie did not care how I performed.	Sometimes she had problems and was hard to understand.

5.4 Geometric reasoning and assembly

As described in Section 4.9.2, three skills were taught, with three puzzles for each skill. Scoring was defined as the proportion of correct independent steps with respect to the total number of steps, as in Libby et al. (2008). The results for Students 1-3 are presented in Figures 5.5-5.7.

As the task instruction was backwards-chained, the opportunity for independent responses increases after each successful trial. Initially, a puzzle with only one missing piece is presented with a controlling prompt, so no independent response is possible. Then, on the next trial involving the same puzzle, the student is asked to complete the puzzle (with one missing piece) independently, giving them one opportunity for an independent response. This proceeds until all puzzles are performed independently.

The results for this study show strong success, which can partially be attributed to the structure of the task instruction, SMP with backwards chaining. Further discussion on the methodology's strengths is conducted in Section 6.1. As shown in Figures 5.5-5.7, all students had difficulty with the "ball" puzzle (Figure 4.34), as the arrangement of pieces in the ball shape was part of the solution, not only the overall shape. This is representative of many assembly tasks, where there can be multiple ways to assemble the pieces, but only one correct way. All students were able to master this puzzle after instruction. Table 5.8 shows errors broken down by puzzle. Both Students 2 and 3 had initial difficulty with the "butterfly" puzzle for reasons similar to the ball. The V-shape, while the among the simplest tasks by number of steps, was also challenging, perhaps because the shape is non-intuitive until the concept of mirroring/symmetry is established.

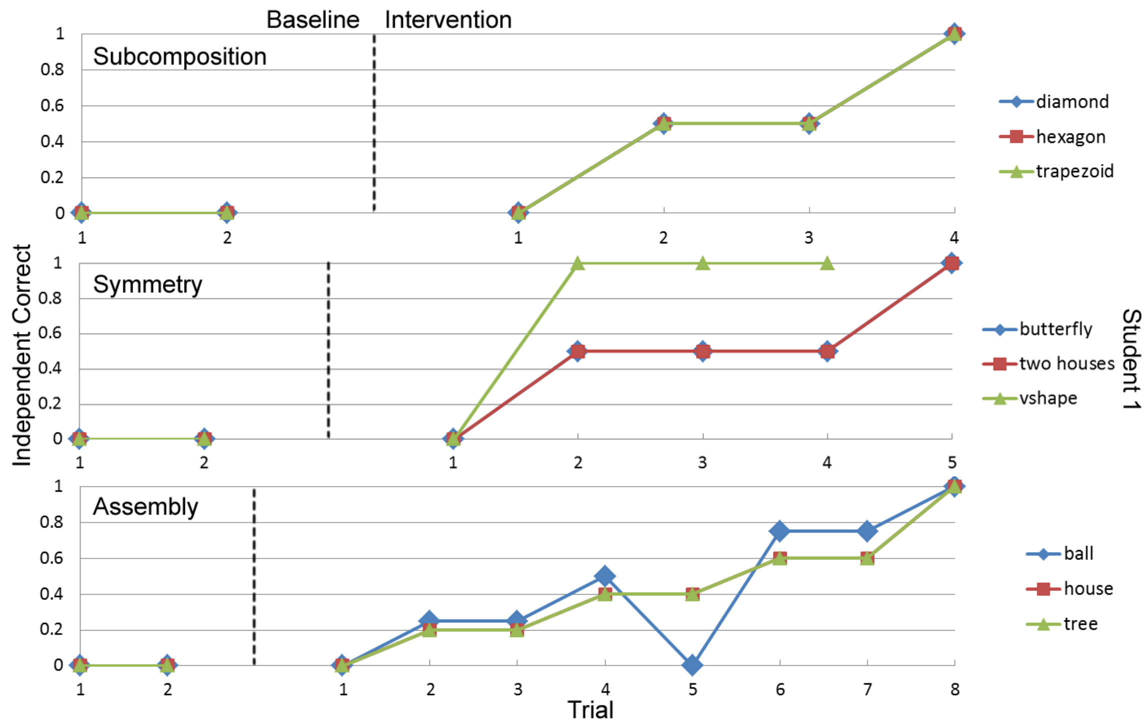


Figure 5.5: Results for *geometric reasoning and assembly* experiments for Student 1.

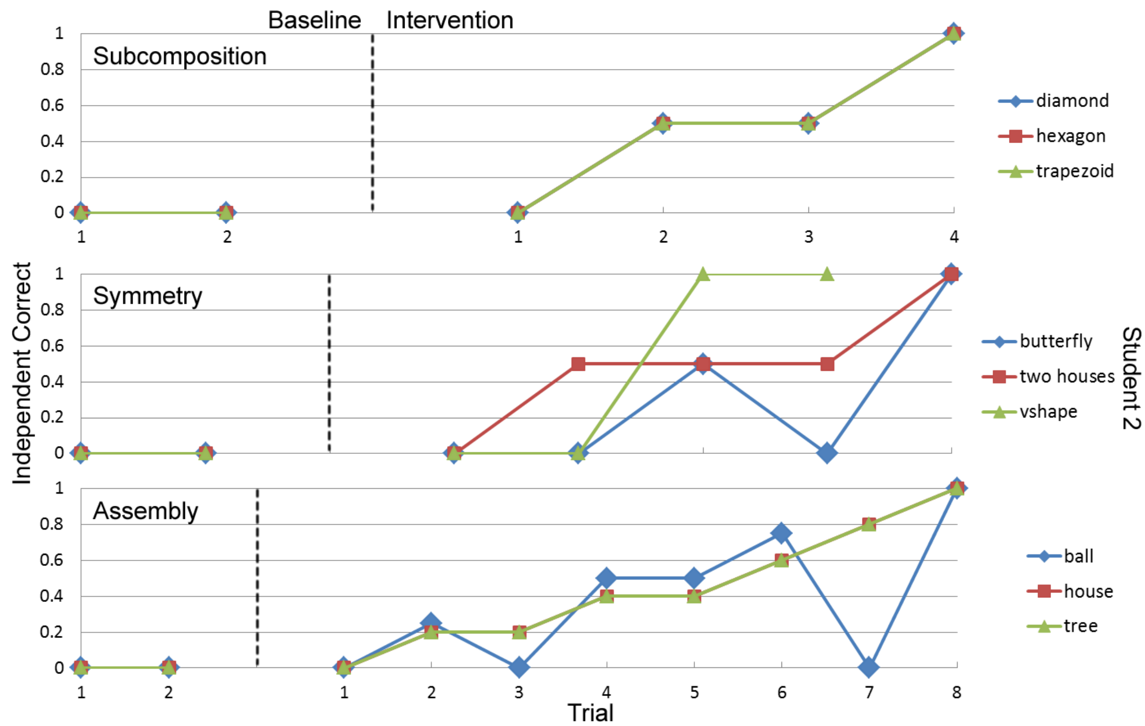


Figure 5.6: Results for *geometric reasoning and assembly* experiments for Student 2.

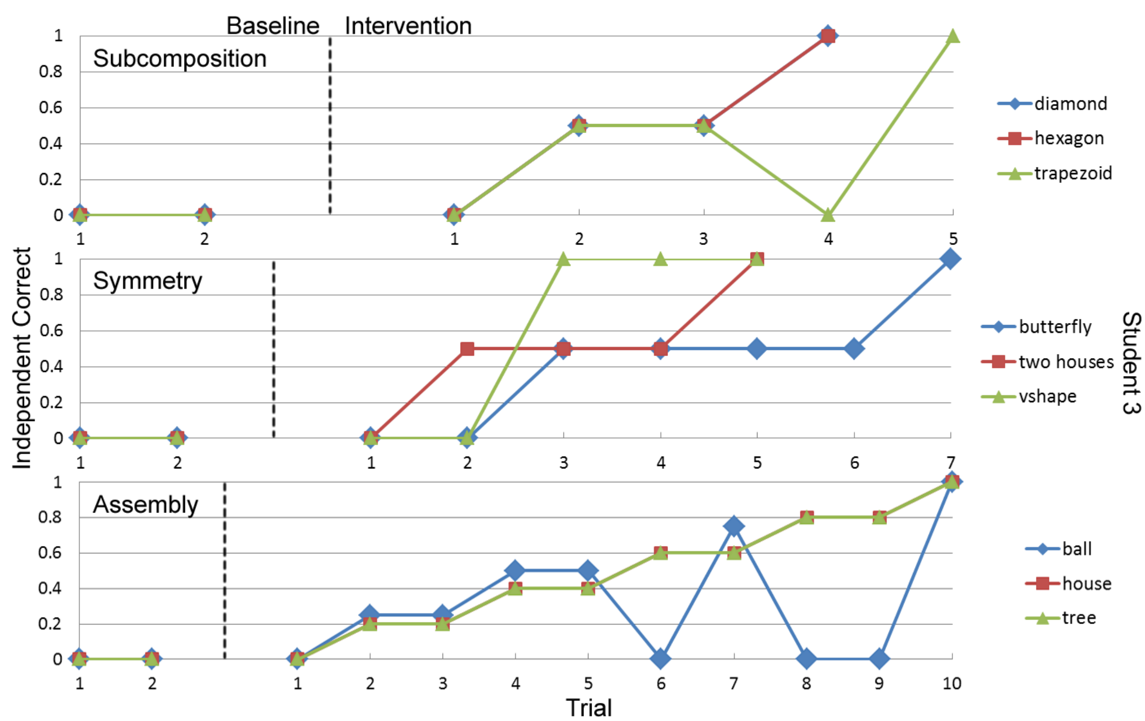


Figure 5.7: Results for *geometric reasoning and assembly* experiments for Student 3.

Table 5.8: Error counts and (rates) for *geometric reasoning and assembly.*

Puzzle	**Student 1**	**Student 2**	**Student 3**
butterfly	0 (0.0)	2 (0.5)	3 (0.5)
two houses	0 (0.0)	0 (0.0)	0 (0.0)
vshape	0 (0.0)	1 (0.33)	1 (0.33)
diamond	0 (0.0)	0 (0.0)	0 (0.0)
hexagon	0 (0.0)	0 (0.0)	0 (0.0)
trapezoid	0 (0.0)	0 (0.0)	2 (0.5)
ball	2 (0.29)	4 (0.57)	11 (1.22)
house	0 (0.0)	0 (0.0)	0 (0.0)
tree	0 (0.0)	0 (0.0)	0 (0.0)

5.5 Summary

In all five experiments conducted, students demonstrated successful acquisition of the skills when instructed by the IRI and ARI system. Constructive observations from these experiments and the construction of this system, as well as future directions for further study are discussed in Chapter 6.

Chapter 6

Discussion

To successfully create this intelligent instruction system, many important decisions concerning experimental design, human-robot interaction, and the "undefined" aspects of the formulation of response prompting for this purpose (Section 3.3.8) were made. The following are important insights into issues of prompting and cognition (Section 6.1) and those that involve the design of an intelligent instruction system (Section 6.2), as well as open questions for future directions (Section 6.3).

6.1 Prompting and cognition

The adoption of response prompting for cognition and the use of multiple versions of response prompting allowed for an interesting comparison between methodologies, particularly when combined with forwards and backwards chaining.

6.1.1 Forwards chaining and System of Least Prompts

Observation: Using SLP with chained tasks allows for an open-ended response; however, the power of the methodology helpfully limits the response space.

The SLP methodology was selected for the *making change* task because it allowed the student to present a discrete response to the question, using increasingly intrusive prompts. This was ultimately successful, although it was a more open form implementation than the method employed in the *geometric reasoning and assembly* task, since the students could respond with any combination of coins, especially if they are uncertain. This led to longer than expected experiences for the students, particularly Student 3, as a result of lacking prerequisite skills (this issue is discussed more in Section 6.2.5). However, despite this issue it is interesting to note the methodology's strength is that a correct answer should always be arrived at, given certain assumptions.

A key element of response prompting is the concept of the *controlling prompt*. The controlling prompt is the most intrusive prompt, providing the highest level of assistance to achieve the task. In the *making change* study using SLP, the controlling prompt is the explicit coin that should be removed or added to move the student towards the goal, along with a gesture to that coin.

Because the controlling prompt is selected to be appropriate not only for the task but for the student receiving instruction, the student should always be capable of following the controlling prompt. By following the controlling prompt, a student should always achieve the correct answer. Furthermore, the number of possible steps is bounded by the number of possible actions in the task. When no repetitive actions are possible, as in the case of a discrete response involving object manipulation, this becomes a lower bound that is entirely within control of the developer. For example, if the student follows the controlling prompt in the *making change* scenario, the maximum number of prompts is the number of coins available to the student, n.

In the worst case, if the correct answer contains some subset of coins, s, where $s \leq n$, at the point where the controlling prompt is reached there the correct s coins are not in the box, and the incorrect coins $n - s$ are. In that case, $n - s$ prompts would be required to remove all incorrect coins from the box, and s steps would be required to add the correct coins. Since $s + (n - s) = n$, there would be at most n steps to reach a correct answer. In our configuration, this would be 20 steps.

To examine a representative low performance, we present as a case study a single trial by a student with I/DD, the results of which are shown in Table 6.1. In this case, the student has difficulty both calculating the first step, i.e., subtracting the price of the item (53 cents) from the payment given (one dollar), as well as counting the value of coins to make a desired sum. In this trial, the student hesitated to begin, remained idle, and triggered a No Response prompt. At prompt level (PL) 1 Rosie asked him to perform the subtraction step to determine how much total change is due for the target price. The student responded, "I don't know," so Rosie gave the student the correct total change number for the price. The student then gave a response of one quarter (Q) and waited for the next prompt. Rosie then gave the Partially Correct prompt for PL 2, which was verbally telling him the quantity he has in his response, the goal, and the difference (shortage) amount. The student responded by placing a larger number of coins than necessary in the response box. Rosie, at PL 3, informed him of the quantity in the response, the goal, and the excess amount, and instructed him to remove a quarter. The student complied, and then waited until Rosie executed PL 4. At PL 4, Rosie walked the student through removing and adding the appropriate coins, until the correct response (C) was provided. Then Rosie positively reinforced the student verbally. From this example case, it can be observed from even this low performance that in just seven steps the student was able to converge to the correct response, and most trials would take much fewer steps. Once the controlling prompt is reached, it is the number of objects available that determines the maximum number steps until a correct response; this quality can be leveraged by developers to control the dynamics and the expected success rate of systems.

Table 6.1: Case study of a low performance trial.

Target:P=53 cents		**Response**					
Step	**PL**	**Q**	**D**	**N**	**P**	**Total**	**Resp.**
1	0	0	0	0	0	0	NR
2	1	1	0	0	0	25	PC
3	2	2	1	1	3	68	I
4	3	1	1	1	3	43	I
5	4	1	1	0	3	38	I
6	4	1	1	0	2	37	PC
7	4	1	2	0	2	47	C

6.1.2 Backwards chaining and System of Most Prompts

Observation: Backwards chaining combined with SMP lends itself very well to instruction from a robot.

I believe that the rapid success the students had in the IRI *geometric reasoning and assembly* study can be directly attributed to the methodology employed, specifically, using backwards chaining with SMP. This approach, while somewhat time-consuming, is entirely appropriate for a robot instructor. By backwards chaining, i.e., starting with an almost complete puzzle then working backwards, the overall state space was reduced, decreasing the chance of errors significantly. Then, using SMP, the students were asked to solve the puzzle independently after each step. By the time the students were performing the entire puzzle independently, there was little to no chance of large mistakes. Using terminology taught in the ARI *geometric reasoning* study, combined with precise gestures to deliver controlling prompts, created a very clear, straightforward experience, and genuinely set the students up for success. Further, while it was time-consuming for the student, this is exactly the sort of scenario where the tirelessness and precision of a robot could be leveraged, especially if the speed of the instruction is streamlined for efficiency. I believe that these qualities that make backwards chaining and SMP on robot instructors so successful is an extremely important characteristic.

6.2 Designing an intelligent instruction system

Designing an intelligent instruction system requires more than a robot that is programmed to teach. Issues of successful human-robot interaction for this domain, ideas that are not fully conceptualized in the prompting methodologies, and other pitfalls came to light when programmed on an instruction system; these issues are discussed here.

6.2.1 Designing prompts

Observation: Prompt design is a critical determining factor in success or failure of an instruction system.

Prompts, when presented for instruction by an intelligent system, particularly a robot, are the primary aspect of instruction. The decision of what verbiage to use and even how to gesture is therefore critical for successful instruction. Verbiage that is confusing or unclear will result in a failed prompt and uncertainty for the participant. For example, when designing the first IRI study, *making change*, a volunteer with I/DD who was not part of the experimental participants aided in testing the system. It was quickly discovered that in the event that a person does not possess the skill to solve a math problem in their head, having the robot verbally set up a math problem (e.g., "If there are 100 cents in a dollar, and the price is 52 cents, how much change is that?") is not constructive to instruction. To solve this problem, the separate calculator sub-skill was introduced, as discussed in Section 5.3.1. Gestures that are ambiguous have a similar effect, and even gestures that are slow and predictable can create undesired anticipation. Reviewing these interactions with experts in education is highly recommended. The decision to forgo using direct object manipulation and instead use gestures for the IRI's controlling prompts was made with the approval of experts. Successful instruction was possible using gestures for the controlling prompt because the students were mentally capable of extrapolating the meaning of the gesture; for instruction of other populations (e.g., younger or with more severe disabilities), this may not be possible.

6.2.2 Levels of interaction

Observation: Incorporating multiple levels of interaction detail increases efficiency and affects trust.

The importance of trust in human-robot interaction is a well-studied topic (Hancock et al. (2011)). When performing something as repetitive as instruction, it is both expedient and enhances the user experience to design the interaction with multiple modes, or levels, of interaction. The motivation is to both decrease the overall cognitive load for interaction and give the intelligent system the appearance of intelligence, thereby increasing the student's trust in the robot's instruction. For this reason, dialog, especially repeated text, should be minimized, and quick short verbiage for sections that must be repeated should be used, thus mimicking the efficiency of a human instructor while exhibiting the consistency of a robot. Something as simple as having the robot introduce itself more than once to the same participant provides a less efficient experience and could affect the student's perception of the intelligence of the robot, thereby reducing trust in the robot's abilities.

6.2.3 The importance of accurate perception and reasoning

Observation: Accurate perception and reasoning is critical.

The accurate evaluation of human responses is a critical part of an intelligent instruction system. A failure to perform this evaluation correctly that results in an incorrect feedback from the instructor could be considered inherently *harmful* to the student, which raises ethical concerns as well. This important ability is built upon the perception capabilities of the system. However, the ability to reason about that which was perceived adds an additional layer of complexity to the problem. For example, in the work presented here, the manner in which the tangram puzzles were reasoned upon was a complex problem. The system had to evaluate the correctness of the puzzle and determine what feedback to deliver. It had to be designed in such a way as to accommodate for slight variations in the location and orientation of the puzzle: the smallest puzzle piece has a radius of less than 1cm, which is an unreasonable margin of error within which to expect a puzzle to be constructed in the same spot repeatedly. For this reason, the first piece of each puzzle was treated as an "anchor," against which the position and orientation of which all other pieces were checked for correctness. Further, the order of evaluation was a decision point: e.g., if you have the solution, but there are extra pieces present, what should the feedback be? The decision was made to classify errors into extra, missing, and misalignment errors, in order, and only present the correction for the first error in that order. This prioritization was similar to what a human instructor might do and simplified the interaction.

6.2.4 Speech recognition

Observation: Accurate speech recognition is essential for an intelligent system.

The speech recognition implementation, while not perfect, was tested and appeared robust, particularly on the relatively small corpus used for these experiments (common affirmative/negative and start/stop words, plus numbers for the *making change* study). However, the speech recognition system made frequent mistakes when interpreting Student 3 in the *making change* study. Student 3 is not diagnosed with any speech articulation difficulties; the errors were possibly due to the student's accent and cadence. Re-training the speech recognition on the student's voice only yielded a scored accuracy rate of 85%; empirically, performance was observed lower still. To correct for this, a manual override was added to allow the experiment operator to override any verbal response with typed text. Further, to prevent accidentally enabling the verbal e-stop described in Section 4.5, when the operator override is enabled, verbal e-stop commands are disabled. This was another example of a design decision to accommodate for the need for the intelligent system to appear intelligent or risk eroding the student's trust.

6.2.5 Teach prerequisite skills

Observation: Assessing and teaching prerequisite skills is essential for success.

As discussed in Section 5.3.1, an important lesson was learned when Student 3 was subject to intervention. Although her ability to perform the skill independently was assessed in baseline, her knowledge of prerequisite skills was not assessed. (In particular, the student was not completely clear on the value of each coin; this issue was addressed by affixing stickers with the coin values next to each pile of coins.) During intervention, therefore, she was indirectly gaining experience with a concept simultaneously. The lesson learned here was that prerequisite skills should be assessed and taught, if necessary. This lesson was applied successfully through the use of the ARI system to teach *geometric reasoning* prerequisite skills prior to the IRI teaching more advanced *geometric reasoning and assembly.*

6.2.6 Securing attention

Observation: The "secure attention" step of the response prompting methodologies can be performed through a combination of speech, gaze, and gestures.

Securing attention is part of the instructional methodology that, while intuitive to a human instructor, needed to be implemented simply and effectively on a robot instructor. Using speech and speech recognition, the IRI asks the student if he or she is ready to begin. The use of human skeleton tracking via an RGB-D camera gave the robot the impression of maintaining eye contact. After the Task Instruction and the Prompt Delivery stages (Figure 4.13), the beginning of the Student Response and

Observation stage is signaled by the robot looking down to the table. When the IRI is ready to give feedback, it looks up at the student to reestablish contact. Anecdotally, it was observed that adding this simple feature established a clear “turn-taking” relationship between the robot and the student.

6.2.7 Idle detection

Observation: An intelligent system must use observations of the human or environment to determine when a response has been given, particularly when interacting with objects.

A second aspect of adopting the instructional methodologies from Section 3.3 was quickly realized. When interacting with objects, there is an essential need for the ability to detect when the student has presented their response and when they are still working. While a prompt interval is specified by the methodology, a second “idle” interval was added to determine how long the students should be allowed to pause before the robot thinks they have responded. Initially, using skeleton tracking information to determine the location of the participant’s hands was attempted to evaluate their state; however, noisy and inaccurate information made this prohibitive without significant investment. Instead, the state of the objects was used; by determining if the sum of the absolute changes in position and orientation of the objects exceeds an empirically defined threshold, an idle/working state of the participant is established. When that state is maintained for a period greater than the idle interval, the IRI assumes the student has responded.

6.3 Future Directions

This work is only a initial foray into the realm of using robots, augmented reality, and other intelligent systems to teach. Possible future directions include exploring several open questions.

How can accurate and generalized perception be achieved? The problem of perception persists in any robotic system, and when interacting with humans, failures in perception are often unacceptable. When failures can impact students' learning, this is even more of an issue. Generalizing perception of people and objects is also a compelling direction, as the breadth of important skills that can be taught is enormous. Being able to teach many skills in many settings would require a robust approach to perception.

Can the audience for teaching be expanded through an adept manipulation system? In this research, using gestures for controlling prompts was sufficient given the target population's abilities and the skills being taught; however, to teach certain skills, or to teach certain populations (e.g., those with less ability), controlling prompts that involve direct manipulation of objects would be required. A robust manipulation system for this purpose would therefore be desired.

How could learning a model of the students improve the system? The ability to model the students' abilities would allow the system to be more adaptive in several directions. First, knowledge of the student's ability level would allow for dynamically changing the prompt hierarchy, for example, changing the number of prompt levels, the intrusiveness of the controlling prompt, and the rate at which the intrusiveness changes. It could also allow for changing certain experiment parameters, such as the prompt interval, task interval, and the idle interval introduced in this research, to better fit the performance speed of the student. Finally, the task difficulty could be adapted to the student's abilities, perhaps by using simpler or more challenging versions of the same task, as necessary.

What is necessary to scale the approach to multiple students, for example in a classroom setting? There are two ways to scale this approach to groups: multiple students, each with their own technological device (e.g., robot or AR), and multiple students interacting with the same device. Multiple students interacting with the same robot, for example, would require multi-human perception, and would resemble a classroom setting. If the students each had their own device and were performing a cooperative or competitive task, a distributed approach to perception and/or reasoning might be employed. Both scenarios represent an interesting opportunity to grow this research beyond teaching a single student at a time.

What other HRI design aspects can improve the user experience and learning efficiency? This research made many decisions that were observed to be practical with regards to HRI. Informed by other research, it would be interesting to explore those decisions, for example, social cues as in Hegel et al. (2011b) in the specialized context of the student-instructor roles. Decisions informed by education methods or observing teachers could also be informative. These explorations should be driven by the questions: what interactions can a teacher employ to improve teaching, and can those be applied to a robot?

Can AR and robots be combined in the same instruction task for better instruction? This work used AR and a robot to teach related skills, but did not combine them for instruction on a single skill. Identifying a skill or a domain of skills that would benefit from and lend itself to instruction using both simultaneous technologies for instruction as a complete system would be informative.

What does a knowledge model that can handle uncertainty look like, and how can that be combined with the cognitive approach to reasoning? In this research, assembly tasks were treated as having only one correct solution, as in many job settings following the instructions is an important skill. Using a knowledge model that allows for multiple solutions to a problem, more than one sequence to arrive at those solution, or allows for uncertainty in responses, would increase the power of this approach in significant ways.

Can a human teach a robot that then teaches a human? Can LfD approaches be used to teach a robot a skill, then from that knowledge could response prompting be used to teach a human? This knowledge transference problem is extremely challenging and spans several domains. For example, the knowledge representation used would be key: the robot must not only be able to know what to do, but why it must be done as well. Then, using that knowledge, a method of generating prompts would be required as well. If achieved, this would represent a large step forward in understanding of knowledge and learning.

Chapter 7

Conclusion

This dissertation presents the creation of an intelligent instruction system, using intelligent augmented reality instruction (ARI) and an intelligent robot instructor (IRI) to teach socially valid life skills to students with intellectual and developmental disabilities.

To accomplish this, the novel use of response prompting for real-world, automated intelligent systems, both in augmented reality and in AI robotics, is introduced and this approach is adopted for cognitive decision making. Success with both ARI and IRI systems shows the generalizability of the response prompting approach. The benefits of this combination are demonstrated by using the ARI system to teach essential prerequisite skills prior to instruction with an IRI. Multiple varieties of prompting with forwards and backwards chained as well as non-sequential tasks are implemented, and the applicability of these approaches to AR and robotic instruction are compared in the context of skills being taught. For this work I have built a system from the ground up, including creating custom object detection and tracking for instructional interaction, and employing machine learning and computer vision methods for augmented reality instruction.

Experiments designed and evaluated using SCED methodology from the education field have been conducted, with results demonstrating that the system successfully teaches students with I/DD these important skills. Subjective results from students in the participant group are presented that indicate enthusiasm for this learning experience.

Several useful observations can be drawn from this research: For the important contribution of response prompting, we observe that, in instruction involving objects, despite the large number of possible responses available (relative to the number of objects), the SLP methodology still converges to a successful instructional experience. The SLP approach is contrasted with the comparatively rapid success of instruction with SMP and backwards chained tasks, which while having a higher overhead of number of instructional steps that would perhaps be tedious for a human instructor to perform would of course be not so for a robot. Of course, not all tasks can be taught with either approach, so the difficulties, tradeoffs, and the abilities of the students should be considered. In the context of designing intelligent instructional systems, we observe the extreme importance of designing prompts and the mechanics of interaction for successful learning. Prompt design is a critical determining factor in the success or failure of an instruction system, particularly with regards to verbiage and interaction. Incorporation of levels of verbal and gesture interaction that vary (i.e., become more streamlined) in the context of continued instruction, accurately perceiving and reasoning about the interaction medium, and accurate speech recognition are all essential for a system that appears intelligent. This both prevents loss of trust in the instruction given, and ensures correct feedback, thus causing no confusion or harm to the student's learning. The need to assess prerequisite skills prior to instruction is important, as an automated system might not be as flexible as a human instructor in situations where the student does not possess such knowledge. When designing response prompting steps, the importance of performing the "securing attention" step is identified, which is addressed with a verbal queue and head gestures, and "idle detection" is incorporated to the instructional process, which is the ability to determine when a response has been given and that the student is ready for feedback.

There are many promising future directions for extending research in intelligent instruction by embodied robotic instructors and augmented reality systems. Perception will always be a critical component; as the interaction medium changes, the ability to quickly and accurately perceive responses is critical. Providing instruction through manipulation of objects would broaden the scope of both what can be taught as well as who can be taught, as some populations would require more concrete controlling prompts (i.e., through physical demonstration). The ability to teach tasks that lend themselves to group instruction to multiple students simultaneously through the incorporation of multi-human perception and/or distributed systems would begin to show the scaling potential of this approach to a classroom setting. Exploring design choices and improvements, particularly with regards to HRI, offers a great number of directions to attempt to improve the efficiency of the system and the experience for the student. Identifying a domain of skills for which student learning would benefit from a combination of AR with an IRI in the same instruction would be a novel and powerful application. Examining skills that have uncertainty in responses, problems with more than one solution, or more than one way to arrive at a solution, and creating systems that can be adaptive to those scenarios would broaden the scope and could challenge the design approach in interesting ways. Finally, combining robot learning with robot teaching through transfer of knowledge would be extremely challenging but would represent a powerful opportunity to explore understanding of knowledge and learning.

Ultimately, this research shows that through the advancement of intelligent, autonomous robotic and augmented reality instruction there is excellent potential to empower people, particularly those with I/DD, to lead more independent lives.

Bibliography

Addo, I. D., Ahamed, S. I., and Chu, W. C. (2013). Toward Collective Intelligence for Fighting Obesity. In *2013 IEEE 37th Annual Computer Software and Applications Conference*, pages 690–695. IEEE. 2, 19

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist*, 51:355–365. 15

Argall, B. D., Chernova, S., Veloso, M., and Browning, B. (2009). A survey of robot learning from demonstration. *Robotics and autonomous systems*, 57(5):469–483. 13

Ault, M. J. and Griffen, A. K. (2013). Teaching with the system of least prompts an easy method for monitoring progress. *TEACHING Exceptional Children*, 45(3):46–53. 91

Aurasma (2016). HP Aurasma https://www.aurasma.com/. 8

Bainbridge, W. A., Hart, J., Kim, E. S., and Scassellati, B. (2008). The effect of presence on human-robot interaction. In *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on*, pages 701–706. IEEE. 2, 16

Begoli, E., Ogle, C. L., Cihak, D. F., and MacLennan, B. J. (2013). Towards an integrative computational foundation for applied behavior analysis in early autism interventions. In *Artificial Intelligence in Education*, pages 888–891. Springer. 20, 21

Benjamin, D. P., Lyons, D., and Lonsdale, D. (2004). ADAPT: A cognitive architecture for robotics. In *International Conference of Cognitive Modeling*. 15

Billard, A., Robins, B., Nadel, J., and Dautenhahn, K. (2007). Building robota, a mini-humanoid robot for the rehabilitation of children with autism. *Assistive Technology*, 19(1):37–49. 1, 19

Boe, E. E. (2014). Teacher demand, supply, and shortage in special education. *Handbook of Research on Special Education Teacher Preparation*, page 67. 4

Bohning, G. and Althouse, J. K. (1997). Using tangrams to teach geometry to young children. *Early childhood education journal*, 24(4):239–242. 81

Bohren, J. and Cousins, S. (2010). The smach high-level executive [ros news]. *Robotics & Automation Magazine, IEEE*, 17(4):18–20. 58

Bower, M., Howe, C., McCredie, N., Robinson, A., and Grover, D. (2014). Augmented reality in education–cases, places and potentials. *Educational Media International*, 51(1):1–15. 7, 21

Bradski, G. (2000). *Dr. Dobb's Journal of Software Tools*. 43, 45, 53

Brewster, S., Lumsden, J., Bell, M., Hall, M., and Tasker, S. (2003). Multimodal'eyes-free'interaction techniques for wearable devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 473–480. ACM. 20

BuildAR (2016). Layar https://www.buildar.com/. 8

Burghart, C., Mikut, R., Stiefelhagen, R., Asfour, T., Holzapfel, H., Steinhaus, P., and Dillmann, R. (2005). A cognitive architecture for a humanoid robot: a first approach. In *IEEE-RAS International Conference on Humanoid Robots*. 15

Buxton, H. (2002). Generative models for learning and understanding dynamic scene activity. In *ECCV Workshop on Generative Model Based Vision*. 15

Cakmak, M., Chao, C., and Thomaz, A. L. (2010). Designing interactions for robot active learners. *Autonomous Mental Development, IEEE Transactions on*, 2(2):108–118. 16

Cakmak, M., Lopes, M., et al. (2012). Algorithmic and human teaching of sequential decision tasks. In *AAAI Conference on Artificial Intelligence (AAAI-12)*. 14

Carnegie Mellon University (2016). CMU Sphinx http://cmusphinx.sourceforge.net/. 42

Chang, Y.-J., Kang, Y.-S., and Huang, P.-C. (2013). An augmented reality (ar)-based vocational task prompting system for people with cognitive impairments. *Research in developmental disabilities*, 34(10):3049–3056. 21

Chang, Y.-J. and Wang, T.-Y. (2010). Comparing picture and video prompting in autonomous indoor wayfinding for individuals with cognitive impairments. *Personal and Ubiquitous Computing*, 14(8):737–747. 21

Chao, C., Cakmak, M., and Thomaz, A. L. (2010). Transparent active learning for robots. In *Proceeding of the 5th ACM/IEEE international conference on Human-robot interaction - HRI '10*, page 317, New York, New York, USA. ACM Press. 82

Chater, N., Tenenbaum, J. B., and Yuille, A. (2006). Probabilistic models of cognition: where next? *Trends in Cognitive Sciences*, 10(7):292–293. 16

Chin, K.-Y., Wu, C.-H., and Hong, Z.-W. (2011). A Humanoid Robot as a Teaching Assistant for Primary Education. *2011 Fifth International Conference on Genetic and Evolutionary Computing*, pages 21–24. 18

Clements, D. H., Sarama, J., and DiBiase, A.-M. (2004). *Engaging young children in mathematics: Standards for early childhood mathematics education*. Routledge. 81

Crowley, J. L. (2006). Things that see: Context-aware multi-modal interaction. In *Cognitive Vision Systems*, volume 3948, pages 183–198. 15

Cypher, A. and Halbert, D. C. (1993). *Watch what I do: programming by demonstration*. MIT press. 13

Dautenhahn, K. and Werry, I. (2004). Towards interactive robots in autism therapy: Background, motivation and challenges. *Pragmatics & Cognition*, 12(1):1–35. 1, 19

Dollar, P., Wojek, C., Schiele, B., and Perona, P. (2012). Pedestrian detection: An evaluation of the state of the art. *Proc. Pattern Analysis and Machine Intelligence*. 46

Duric, Z., Gray, W., Heishman, R., Li, F., Rosenfeld, A., Schoelles, M., Schunn, C., and Wechsler, H. (2002). Integrating perceptual and cognitive modeling for adaptive and intelligent human-computer interaction. *Proceedings of the IEEE*, 90(7):1272–1289. 15

eSpeak (2016). eSpeak text to speech http://espeak.sourceforge.net/. 42

Everingham, M., Van Gool, L., Williams, C. K., Winn, J., and Zisserman, A. (2010). The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88(2):303–338. 46

Fasola, J. and Mataric, M. (2013). A socially assistive robot exercise coach for the elderly. *Journal of Human-Robot Interaction*, 2(2):3–32. 17, 20

Feil-Seifer, D. and Mataric, M. (2011). Automated detection and classification of positive vs. negative robot interactions with children with autism using distance-based features. In *Proceedings of the 6th International Conference on Human-Robot Interaction*, pages 323–330. ACM. 1, 19

Feil-Seifer, D. and Mataric, M. J. (2008). B3IA: A control architecture for autonomous robot-assisted behavior intervention for children with autism spectrum disorders. In *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on*, pages 328–333. IEEE. 16, 19

Foote, T. (2013). tf: The transform library. In *Proc. Technologies for Practical Robot Applications*. 45

Foster, S. L. and Mash, E. J. (1999). Assessing social validity in clinical treatment research: issues and procedures. *Journal of consulting and clinical psychology*, 67(3):308–19. 3

Freina, L. and Ott, M. (2015). A literature review on immersive virtual reality in education: State of the art and perspectives. *Proceedings of eLearning and Software for Education (eLSE)(Bucharest, Romania, April 23–24, 2015)*. 21

Fridin, M. (2014a). Kindergarten social assistive robot: First meeting and ethical issues. *Computers in Human Behavior*, 30:262–272. 18

Fridin, M. (2014b). Storytelling by a kindergarten social assistive robot: A tool for constructive learning in preschool education. *Computers & Education*, 70:53–64. 2, 18

FUTURE (2016). FUTURE Postsecondary Education Program http://futureut.utk.edu/. 37

Gast, D. L. and Spriggs, A. D. (2010). Visual analysis of graphic data. *Single subject research methodology in behavioral sciences*, pages 199–233. 60, 63, 78

Gillespie, A., Jones, J., Clarke, D., Bundy, D., and Lee, S. (2003). Skills for health. skills-based health education including life skills: an important component of a child-friendly/health-promoting school. *World Health Organization*. 3

Gómez, J., Alamán, X., Montoro, G., Torrado, J. C., and Plaza, A. (2014). Amicog–mobile technologies to assist people with cognitive disabilities in the work place. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, 2(7):9–17. 21

Gómez, J., Montoro, G., Haya, P. A., García-Herranz, M., and Alamán, X. (2011). Distributed schema-based middleware for ambient intelligence environments. *Ubiquitous Developments in Ambient Computing and Intelligence: Human-Centered Applications*, pages 205–2018. 21

Gray, W. D., Young, R. M., and Kirschenbaum, S. S. (1997). Introduction to this special issue on cognitive architectures and human-computer interaction. *Human-Computer Interaction*, 12(4):301–309. 15

Greczek, J., Kaszubski, E., Atrash, A., and Mataric, M. (2014). Graded cueing feedback in robot-mediated imitation practice for children with autism spectrum disorders. In *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 561–566. IEEE. 19

Grigal, M. and Hart, D. (2010). *Think College! Postsecondary Education Options for Students with Intellectual Disabilities.* Brookes Publishing Company. 7

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., and Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 53(5):517–527. 108

Hegel, F., Gieselmann, S., Peters, A., Holthaus, P., and Wrede, B. (2011a). Towards a typology of meaningful signals and cues in social robotics. In *Robot and Human Interactive Communication, 2011. RO-MAN 2011. The 20th IEEE International Symposium on*, pages 72–78. IEEE. 17

Hegel, F., Gieselmann, S., Peters, A., Holthaus, P., and Wrede, B. (2011b). Towards a typology of meaningful signals and cues in social robotics. In *RO-MAN, 2011 IEEE*, pages 72–78. IEEE. 113

Herberg, J. (2013). Eliciting ideal tutor trait perception in robots Pinpointing effective robot design space elements for smooth tutor interactions. *Human-Robot Interaction*, pages 137–138. 18

Horner, R. and Spaulding, S. (2010). Single-case research designs. *Encyclopedia of research design*, pages 1386–1394. 60

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., and Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, 71(2):165–179. 60, 63, 78

Howard, A. M., Roberts, L., Garcia, S., and Quarells, R. (2012). Using mixed reality to map human exercise demonstrations to a robot exercise coach. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 291–292. IEEE. 2, 19

Howard, L. A., Larison, I. W., and James Kaplan, P. (2013). Addressing the special education teacher shortage in west virginia via online learning: A case study. In *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, volume 2013, pages 531–539. 4

iRobot (2016). SMACH http://www.irobot.com/For-Business/Ava-500.aspx. 20

Isla, D., Burke, R., Downie, M., and Blumberg, B. (2001). A layered brain architecture for synthetic creatures. In *International Joint Conferences on Artificial Intelligence*. 15, 16, 26

Jacko, J. A. (2012). *Human Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications*. CRC press. 20

Jaimes, A. and Sebe, N. (2007). Multimodal human–computer interaction: A survey. *Computer vision and image understanding*, 108(1):116–134. 20

Junaio (2016). Junaio https://my.metaio.com/dev/junaio/. 8

Khan, F., Zhu, X. J., and Mutlu, B. (2011). How Do Humans Teach: On Curriculum Learning and Teaching Dimension. *NIPS*, pages 1–9. 14

Kidd, C. D. and Breazeal, C. (2004). Effect of a robot on user perceptions. In *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, volume 4, pages 3559–3564. IEEE. 2, 17

Kiesler, S., Powers, A., Fussell, S. R., and Torrey, C. (2008). Anthropomorphic interactions with a robot and robot-like agent. *Social Cognition*, 26(2):169–181. 2, 17

Kose, H. and Yorganci, R. (2011). Tale of a robot: humanoid robot assisted sign language tutoring. *Humanoid Robots (Humanoids)*, pages 105–111. 2, 18

Krämer, N. C. and Bente, G. (2010). Personalizing e-Learning. The Social Effects of Pedagogical Agents. *Educational Psychology Review*, 22(1):71–87. 2, 17

Kuno, Y., Shimada, N., and Shirai, Y. (2003). A robotic wheelchair based on the integration of human and environmental observations- look where you're going. *IEEE Robotics & Automation Magazine*, 10(1):26–34. 20

Laird, J. E., Newell, A., and Rosenbloom, P. S. (1987). SOAR: an architecture for general intelligence. *Artificial Intelligence*, 33(1):1–64. 15

Lancioni, G. E. and O'Reilly, M. F. (2001). Self-management of instruction cues for occupation: Review of studies with people with severe and profound developmental disabilities. *Research in Developmental Disabilities*, 22(1):41–65. 7

Layar (2016). Layar https://www.layar.com/. 8

Lee, D.-H. and Kim, J.-H. (2010). A framework for an interactive robot-based tutoring system and its application to ball-passing training. In *2010 IEEE International Conference on Robotics and Biomimetics*, pages 573–578. IEEE. 2, 19

Leyzberg, D., Spaulding, S., Toneva, M., and Scassellati, B. (2012). The physical presence of a robot tutor increases cognitive learning gains. In *34th Annual Conference of the Cognitive Science Society (Cog. Sci.12)*. 2, 16

Libby, M. E., Weiss, J. S., Bancroft, S., and Ahearn, W. H. (2008). A comparison of most-to-least and least-to-most prompting on the acquisition of solitary play skills. *Behavior Analysis in Practice*, 1(1):37. 99

Mckeown, R. (2012). Teacher education 1992 and 2012 reflecting on 20 years. *Journal of Education for Sustainable Development*, 6(1):37–41. 2, 4

McMahon, D., Cihak, D. F., and Wright, R. (2015). Augmented reality as a navigation tool to employment opportunities for postsecondary education students with intellectual disabilities and autism. *Journal of Research on Technology in Education*, 47(3):157–172. 21

McMahon, D. D., Cihak, D. F., Gibbons, M. M., Fussell, L., and Mathison, S. (2013). Using a mobile app to teach individuals with intellectual disabilities to identify potential food allergens. *Journal of Special Education Technology*, 28(3):21–32. 21

Michaud, F. and Théberge-Turmel, C. (2002). Mobile robotic toys and autism. In *Socially Intelligent Agents*, pages 125–132. Springer. 1, 19

Morford, M., Nutter, M., Vogel, D., Bower, H., Hillbrant-johnson, J., Edwards, B., Lynch, V., and Schmidt, G. (2006). Puget Sound Educational Service District Life Skills Curriculum. 4

Mori, M. (1970). The uncanny valley. *Energy*, 7(4):33–35. 17

Nagel, H.-H. (2004). Steps toward a cognitive vision system. *AI Magazine*, 25(2):31–50. 15

OpenNI Consortium (2014). OpenNI http://openni.org/. 42

Quigley, M., Conley, K., Gerkey, B. P., Faust, J., Foote, T., Leibs, J., Wheeler, R., and Ng, A. Y. (2009). Ros: an open-source robot operating system. In *ICRA Workshop on Open Source Software*. 41

Reardon, C., Wright, R., Cihak, D., and Parker, L. E. (2016). Intelligent context-aware augmented reality to teach students with intellectual and developmental disabilities. In *Proceedings of 29th International Florida Artificial Intelligence Research Society Conference (FLAIRS)*. 23, 87

Reardon, C., Zhang, H., Wright, R., and Parker, L. E. (2015). Response prompting for intelligent robot instruction of students with intellectual disabilities. In *Robot and Human Interactive Communication (RO-MAN), 2015 24th IEEE International Symposium on*, pages 784–790. IEEE. 23, 87

Rector, K., Bennett, C. L., and Kientz, J. A. (2013). Eyes-free yoga: an exergame using depth cameras for blind & low vision exercise. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, page 12. ACM. 20

Richard, E., Billaudeau, V., Richard, P., and Gaudin, G. (2007). Augmented reality for rehabilitation of cognitive disabled children: A preliminary study. In *Virtual Rehabilitation, 2007*, pages 102–108. IEEE. 21

Robins, B., Dautenhahn, K., and Dickerson, P. (2009). From isolation to communication: a case study evaluation of robot assisted play for children with autism with a minimally expressive humanoid robot. In *Advances in Computer-Human Interactions, 2009. ACHI'09. Second International Conferences on*, pages 205–211. IEEE. 1, 19

Roth, P. and Pun, T. (2003). Design and evaluation of multimodal system for the non-visual exploration of digital pictures. In *INTERACT*. 20

Sapargaliyev, D. (2015). Learning with wearable technologies: A case of google glass. In Brown, T. and van der Merwe, H. J., editors, *The Mobile Learning Voyage From Small Ripples to Massive Open Waters*, volume 560 of *Communications in Computer and Information Science*, pages 343–350. Springer International Publishing. 21

Scassellati, B., Admoni, H., and Mataric, M. (2012). Robots for use in autism research. *Annual Review of Biomedical Engineering*, 14:275–294. 1, 19

Sidman, M. (1960). *Tactics of scientific research*. Basic Books, Incorporated. 60

Simpson, R., LoPresti, E., Hayashi, S., Nourbakhsh, I., and Miller, D. (2004). The smart wheelchair component system. *Journal of Rehabilitation Research and Development*, 41(3B):429. 20

Smith, C. C. (2013). *Using Mobile Technology to Improve Autonomy in Students with Intellectual Disabilities in Postsecondary Education Programs*. PhD thesis, The University of Tennessee. 21

Sucan, I. A. and Chitta, S. (2016). MoveIt! http://moveit.ros.org/. 42

Tapus, A., Tapus, C., and Mataric, M. (2009). The role of physical embodiment of a therapist robot for individuals with cognitive impairments. In *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, pages 103–107. IEEE. 2, 17

Town, C. and Sinclair, D. (2003). A self-referential perceptual inference framework for video interpretation. In *International Conference on Vision Systems*. 15

UNICEF (2016). Life skills UNICEF Webiste http://www.unicef.org/lifeskills/. 3

Varela, F. J. and Dupuy, J. (1992). *Understanding Origins – Contemporary Views on the Origin of Life, Mind and Society*, chapter Whence perceptual meaning? A cartography of current ideas, pages 235–263. Kluwer Academic Publishers. 15

Vernon, D., Metta, G., and Sandini, G. (2007). A survey of artificial cognitive systems: Implications for the autonomous development of mental capabilities in computational agents. *IEEE Transactions on Evolutionary Computation*, 11(2):151–180. 15

Wainer, J., Feil-Seifer, D. J., Shell, D. A., and Mataric, M. J. (2007). Embodiment and human-robot interaction: A task-based perspective. In *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on*, pages 872–877. IEEE. 2, 17

Watlington, E., Shockley, R., Guglielmino, P., and Felsher, R. (2010). The high cost of leaving: An analysis of the cost of teacher turnover. *Journal of Education Finance*, 36(1):22–37. 2, 4

Weir, S. and Emanuel, R. (1976). *Using LOGO to catalyse communication in an autistic child*. Department of Artificial Intelligence, University of Edinburgh. 1, 19

Westeyn, T. L., Kientz, J. A., Starner, T. E., and Abowd, G. D. (2008). Designing toys with automatic play characterization for supporting the assessment of a child's development. In *Proceedings of the 7th international conference on Interaction design and children*, pages 89–92. ACM. 20

Wilkin, T. and Nwoke, G. (2011). Career and technical education teacher shortage: A successful model for recruitment and retention. *Journal of sTEm Teacher Education*. 2, 4

Wolery, M., Ault, M. J., and Doyle, P. M. (1992). *Teaching students with moderate to severe disabilities: Use of response prompting strategies*. Longman Publishing Group. 25, 28, 29, 30, 31, 32

Xu, F. and Griffiths, T. L. (2011). Probabilistic models of cognitive development: Towards a rational constructivist approach to the study of learning and development. *Cognition*, 120:299–301. 15

Yang, B., Zhou, L., and Deng, Z. (2013). C-HMAX: Artificial cognitive model inspired by the color vision mechanism of the human brain. *Tsinghua Science and Technology*, 18(1):51–56. 15

Zhang, C., Zhang, H., and Parker, L. E. (2015). Feature space decomposition for effective robot adaptation. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pages 441–448. IEEE. 5

Zhang, Y. and Rau, P.-L. P. (2015). Playing with multiple wearable devices: Exploring the influence of display, motion and gender. *Computers in Human Behavior*, 50:148–158. 20

Appendix

Appendix A

Experiment Outlines

A.1 ARI: Geometric reasoning with tangrams

The following is a script outlining the experimental interaction between the ARI and student for the *geometric reasoning* experiment described in Section 4.8.3.

1. Start
2. Teach placing on marker
   (a) Prompt: "Place a shape on the marker, covering the center line."
   (b) "When you're ready, tap to take a picture and check your answer." Take picture
      i. Correct:
         A. "Correct. That is the marker."
         B. Advance to next prompt (Go To "III.")
      ii. Incorrect:
         A. Show image of solution
         B. Repeat prompt (Go To "a.")
3. Teach shapes

(a) Prompt: "Place a *square*, *diamond*, *trapezoid*, *triangle*, *hexagon* on the marker"

(b) Take picture

  i. Correct:

    A. "Correct. That is a *shape*. This *shape* has *number* sides. (The trapezoid has a short top and along bottom.)"

    B. Advance to next prompt.

  ii. Incorrect:

    A. Show image of solution

    B. Repeat prompt

4. Teach relative direction and adjacent

(a) Prompt: "Place a square on the marker"

(b) Take picture

  i. Correct:

    A. "Correct. That is a *shape*. This *shape* has *number* sides. (The trapezoid has a short top and along bottom.)"

    B. Advance to next prompt.

  ii. Incorrect:

    A. Show image of solution

    B. Repeat prompt

(c) Prompt:

  i. "Now place a triangle above it so that the sides line up and it looks like an orange house with a green roof." OR

  ii. "Place a square on the marker, then place a triangle *below/ to the left of/ to the right of* it so that the sides line up."

iii. Take picture

A. Correct: "Correct. This is called adjacent. The triangle is *above/below/left of/right of* the square." Advance to next prompt.

B. Incorrect: Show image of solution. Repeat prompt

5. Teach sub-composition

(a) Prompt: "Place one triangle on the marker, point up"

i. (standard correct/incorrect)

(b) Prompt: "Now place another triangle right-adjacent to it to make a diamond shape."

i. (standard)

6. Teach rotate

(a) Prompt: "Place a trapezoid on the marker with the long side down"

(b) Take picture

i. Correct:

A. "Correct. You have placed the trapezoid with the long side down"

B. Advance to next prompt

ii. Incorrect:

A. Show image of solution

B. Repeat prompt

(c) Prompt: "Now, rotate the trapezoid one quarter turn to the right"

i. Take picture

ii. Correct:

A. "Correct. That is one quarter turn to the right"

B. Advance to next prompt

iii. Incorrect once:

A. Show image of previous step with turn arrow

B. Repeat prompt

iv. Incorrect twice:

A. Show image of solution

B. Repeat prompt

(d) Prompt: “Now, rotate the trapezoid one half turn”

i. Take picture

ii. Correct:

A. “Correct. That is one half turn”

B. Advance to next prompt

iii. Incorrect once:

A. Show image of previous step with turn arrow

B. Repeat prompt

iv. Incorrect twice:

A. Show image of solution

B. Repeat prompt

(e) Prompt: “Now, rotate the trapezoid one quarter turn to the left”

i. Take picture

ii. Correct:

A. “Correct. That is one quarter turn to the left”

B. Advance to next prompt

iii. Incorrect once:

A. Show image of previous step with turn arrow

B. Repeat prompt

iv. Incorrect twice:

        A. Show image of solution
        B. Repeat prompt

   (f) Prompt: "Place a triangle above and adjacent to the trapezoid, in the middle, so that it looks like a boat."
      i. Take picture
      ii. Correct:
         A. "Correct. You have made a boat."
         B. Advance to next prompt
      iii. Incorrect:
         A. Show image of solution
         B. Repeat prompt

   (g) Prompt: "Now, rotate the trapezoid one half turn, so that the long side is on the bottom, to make a pyramid"
      i. Take picture
      ii. Correct:
         A. "Correct. That is one half turn. You have made a pyramid"
         B. Advance to next prompt
      iii. Incorrect once:
         A. Show image of previous step with turn arrow
         B. Repeat prompt
      iv. Incorrect twice:
         A. Show image of solution
         B. Repeat prompt

7. Stop
   (a) Reset Step number for next trial.

A.2 SLP: Making change

The following is a script outlining the experimental interaction between the IRI and student for the *making change* experiment described in Section 4.9.1 using the System of Least Prompts methodology.

A.2.1 Baseline data collection

1. Investigator **introduces task**.
   (a) "I am going to see how well you make change. You will be the cashier. I will tell you the prices, and you will give me the correct change by placing it on the table."
2. Investigator **verifies the attention** of the student.
   (a) "Are you ready to begin?"
3. Investigator **presents target stimulus**, i.e., the change-making scenario. E.g., "The price of the item is fifty-seven cents. If I give you a dollar, show me how much change I get back."
4. Participant performs task as he or she is able, by selecting correct coins.
5. Investigator announces the completion of the trial and thanks the participant for his/her help.

A.2.2 IRI instruction using SLP

1. Rosie the robot **introduces task**.
   (a) "Hello, my name is Rosie, and I'm learning how to be a teacher. Today, I'm going to help you learn how to make change. You will be the cashier. I will tell you the prices, and you will give me the correct change by placing it on the table."

2. Rosie **verifies the attention** of the student.
    (a) "Are you ready to begin?" using speech recognition.
        i. Negative answer loops, waits for "ready"-type response
        ii. Positive answer continues.
3. Rosie **presents target stimulus**, i.e., the change-making scenario. E.g., "The price of the item is fifty-seven cents. If I give you a dollar, show me how much change I get back. Use the coins on the table and place the right answer in the box in the middle of the table."and the prompt, either:
    (a) **No prompt** (at the beginning).
    (b) **Verbal Cue 1**:
        i. **Student does not initiate independently or initiates with incorrect response**: Rosie initiates by asking question: "Okay, let's start by figuring out how much change. One hundred cents subtracted by *price* cents is how much change?" Rosie gives feedback:
            A. Participant answers correctly: "That's right! 100 minus *price* is *answer cents*"
            B. Participant answers incorrectly: "Good try, but 100 minus *price* is *answer cents.*"
        ii. **Student has a partially correct answer** (some of the correct coins and no incorrect coins): Rosie gives verbal cue: "So far you're doing good. You have *current total* cents. You need *goal* cents."
    (c) **Verbal Cue 2**:
        i. **Student does not initiate independently**: Rosie gives verbal cue: "You need to make *goal* cents in change. Which coin should you start with?"
            A. Participant answers correctly: "That's right! Start with a *coin*!"

B. Participant answers incorrectly: "Good try, but try starting with a *coin*."

ii. **Student initiates incorrect response**: Rosie implements verbal feedback and gives a verbal cue: "Good try, but you have too many *incorrect coins*. Try removing *number of incorrect coins* + *incorrect coin names* from the box."

iii. **Student has a partially correct answer** (some of the correct coins and no incorrect coins): Rosie gives verbal cue: "So far you're doing good. You have *current total* cents. You need *goal* cents. That's *difference* more cents."

(d) **Verbal Direction 1**: Give the next coin:

i. **Student does not initiate independently**: Rosie gives verbal directions: "Try starting with a *coin*," (gesture to coin). "A *coin* is *value* cents. *value* from *goal* leaves *difference* cents. Then make change for *difference* cents."

ii. **If the student has a partially correct answer** (some of the coins and no incorrect coins): Rosie gives verbal direction: "So far you're doing good. You have *current total* cents. You need *goal* cents. That's *difference* more cents. Next try a *coin*." (gesture to *coin*). "A *coin* is *value* cents. *value* from *current total* leaves *difference* cents."

iii. **If the student has an incorrect answer** (wrong coins): Ask the student to remove the wrong coins "Please remove *number coins*..." then ask to add the correct coins "Please add *number coins*...".

4. Either:

(a) Rosie escalates prompt if response interval time elapses (repeat step 4) OR

(b) Correct answer is on the board. Rosie provides feedback: "The change from a dollar for *price* cents is *correct answer*." + ("Good job!" or "Way to go!" or "Well done!").

5. Until trial is complete, Rosie **re-verifies attention**:

   (a) "Would you like to try again?"

   (b) "Are you ready to begin?" using speech recognition.

      i. Positive answer repeats instruction.

      ii. Negative answer continues.

A.2.3 Notes

- Experiments take place in Distributed Intelligence Laboratory, Min Kao 629.
- Experiments are recorded via camera.
- Participants wear Bluetooth headphone/microphone (only the microphone will be used) to ensure high accuracy of speech recognition.
- Participants either stand at a 1m high table or sit in a high bench chair.
- After each trial, investigator determine the appropriate **prompt delay interval** for next trial.
- The table has an "answer zone," which is marked on the table, in which the participant will place the coins he or she believes answers the question (Figure A.1).

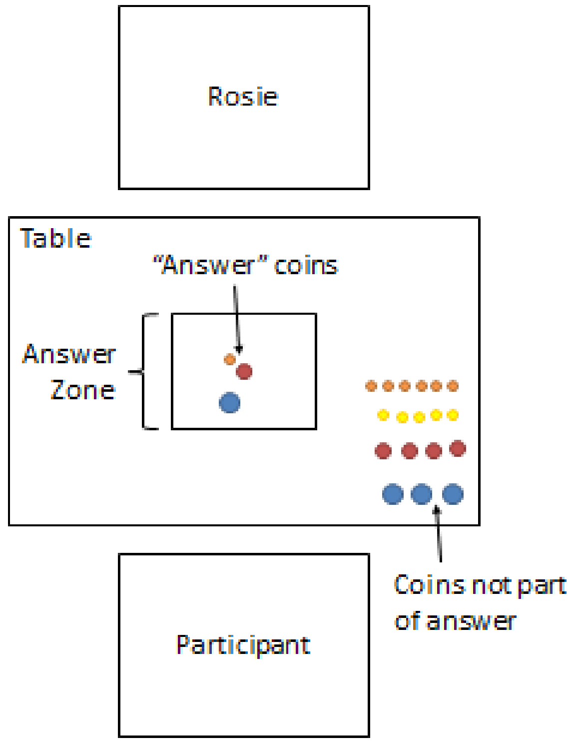


Figure A.1: Table layout and Answer Zone.

A.3 SMP: Geometric reasoning and assembly

The following is a script outlining the experimental interaction between the IRI and student for the *geometric reasoning and assembly* experiment described in Section 4.9.2 using the System of Most Prompts methodology.

A.3.1 Baseline data collection

1. Investigator **introduces task**.

   (a) "We are going to make some puzzles."

2. Investigator **verifies the attention** of the student.

   (a) "Are you ready to begin?"

3. Investigator **presents target stimulus**, i.e., description of puzzle, along with visual representation of outline.

4. Participant performs task as he or she is able.

5. Investigator announces the completion of the trial and thanks the participant for his/her help.

A.3.2 IRI instruction using SMP and backwards chaining

1. Rosie the robot **introduces task**.

    (a) The first time Rosie interacts with the student, she is more verbose:

    i. "Hello, my name is Rosie. I'm going to help you make a puzzle. During the lesson, I will stop you to give you instructions. The first piece goes on the marker, and the rest of the pieces go around it. We'll work backwards so you get to know where all the pieces go."

2. Rosie **verifies the attention** of the student.

    (a) "Are you ready to begin?" using speech recognition.

    i. Negative answer loops, waits for "ready"-type response

    ii. Positive answer continues.

3. Rosie **presents target stimulus**,

    (a) "We are making a *puzzlename*."

4. Rosie presents **task direction**:

    (a) At instruction prompt level 2:

    i. Rosie delivers gesture to the location for the next piece and instruction using learned terminology.

(b) At independent prompt level 1: Rosie asks the student to complete the puzzle independently.

5. After **prompt delay interval** (0 seconds in first), Rosie evaluates the puzzle.

   (a) Errors:

      i. Extra: "Remove the" (point to piece). Repeat prompt.

      ii. Missing: "Add a *shape* here" (point to location). Repeat prompt.

      iii. Misaligned: "Turn the *shape*" (point to piece) "a *[half turn, quarter turn to the left/right, little bit to the left/right]*"

   (b) Correct but puzzle not complete:

      i. Rosie asks to complete the puzzle independently.

   (c) Repeat until no errors

A.3.3 Notes

- Experiments take place in Distributed Intelligence Laboratory, Min Kao 629.
- Experiments are recorded via camera.
- Participants wear Bluetooth headphone/microphone (only the microphone will be used) to ensure high accuracy of speech recognition.
- Participants either stand or sit (if necessary).
- After each trial, investigator will determine the appropriate **prompt delay interval** for next trial.

A.3.4 Schedule

Tables A.1 and A.2 show the schedule of experiments for the *geometric reasoning and assembly* skills. Because backwards chaining is used, steps are decremented.

Two prompt levels are used: control, the controlling prompt, and independent, the prompt for the student to perform the skill (i.e., finish the puzzle) independently.

Table A.1: Experiment schedule for *geometric reasoning and assembly* for the mirroring and subcomposition skills.

Skill	**Session**	**Puzzle**	**Step**	**Prompt Lvl.**
Mirroring	1	two houses	3	Control
		butterfly	2	Control
		v-shape	1	Control
	2	two houses	3	Independent
		butterfly	2	Independent
		v-shape	1	Independent
	3	two houses	2	Control
		butterfly	1	Control
		v-shape	1	Independent
	4	two houses	2	Independent
		butterfly	1	Independent
Subcomposition	1	diamond	1	Control
		trapezoid	1	Control
		hexagon	1	Control
	2	diamond	1	Independent
		trapezoid	1	Independent
		hexagon	1	Independent
	3	diamond	0	Control
		trapezoid	0	Control
		hexagon	0	Control
	4	diamond	0	Independent
		trapezoid	0	Independent
		hexagon	0	Independent

Table A.2: Experiment schedule for *geometric reasoning and assembly* for the assembly skill.

Skill	**Session**	**Puzzle**	**Step**	**Prompt Lvl.**
Assembly	1	house	4	Control
		tree	4	Control
		ball	3	Control
	2	house	4	Independent
		tree	4	Independent
		ball	3	Independent
	3	house	3	Control
		tree	3	Control
		ball	2	Control
	4	house	3	Independent
		tree	3	Independent
		ball	2	Independent
	5	house	2	Control
		tree	2	Control
		ball	1	Control
	6	house	2	Independent
		tree	2	Independent
		ball	1	Independent
	7	house	1	Control
		tree	1	Control
		ball	0	Control
	8	house	1	Independent
		tree	1	Independent
		ball	0	Independent
	9	house	0	Control
		tree	0	Control
		ball	0	Independent
	10	house	0	Independent
		tree	0	Independent
		ball	0	Independent

Appendix B

Subjective Assessment

The subjective Likert-style assessment surveys for pre- and post-experiment are shown in Sections B.1 and B.2. Each survey was conducted orally by the experimenter. Figure B.1 shows the scale tool that was presented to the students when answering each question. Student responses to open questions are shown in Section 5.3.3.

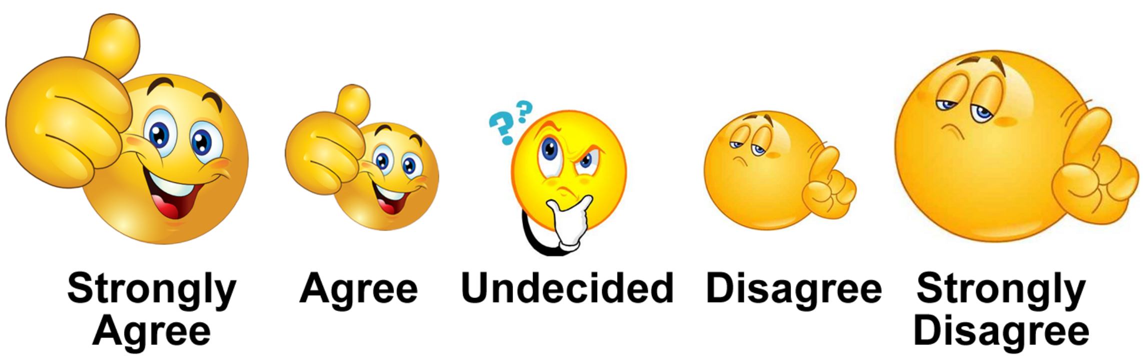


Figure B.1: Subjective assessment visual survey scale tool.

B.1 Pre-assessment survey

Instructions to student: "I'm going to read some statements about working with Rosie the robot. Please listen to each one carefully. Then point to the image that shows how much you agree or disagree with the statement."

Investigator, use the following key:

SA = Strongly Agree

A = Agree

U = Undecided

D = Disagree

SD = Strongly Disagree

Present the student with the following example, and ask them to gesture to the image that represents their sentiment:

Table B.1: Example survey question.

0.	I think pizza with pepperoni is the best.	SA	A	U	D	SD

- If you are really positive that pepperoni pizza is the best, circle SA (Strongly Agree).
- If you think that it is good but maybe not great, circle A (Agree).
- If you can't decide whether or not it is best, circle U (Undecided).
- If you think that pepperoni pizza is not all that good, circle D (Disagree).
- If you are really positive that pepperoni pizza is not very good, circle SD (Strongly Disagree).

Table B.2: Pre-assessment survey questions.

1.	I like computers. Why or why not?	SA	A	U	D	SD
2.	I have seen a robot before. Where?	SA	A	U	D	SD
3.	I think a robot could make a good teacher. Why or why not?	SA	A	U	D	SD
4.	I have difficulty making change with coins. Why or why not?	SA	A	U	D	SD
5.	I like robots. Why or why not?	SA	A	U	D	SD
6.	I don't think I'd enjoy working with a robot. Why or why not?	SA	A	U	D	SD
7.	I have worked or played with a robot before. Where?	SA	A	U	D	SD
8.	I think robots can do many things well. Why or why not?	SA	A	U	D	SD
9.	Computers are useless. Why or why not?	SA	A	U	D	SD
10.	I think I could work together with a robot. Why or why not?	SA	A	U	D	SD
11.	I know a lot about robots. Where did you learn?	SA	A	U	D	SD
12.	I think I could learn from a robot. Why or why not?	SA	A	U	D	SD
13.	Making change with coins is easy for me. Why or why not?	SA	A	U	D	SD
14.	Robots are bad. Why or why not?	SA	A	U	D	SD
15.	I am good at making change with coins. Why or why not?	SA	A	U	D	SD
16.	I don't want a robot to teach me. Why or why not?	SA	A	U	D	SD
17.	I think robots are not very useful. Why or why not?	SA	A	U	D	SD
18.	Making change with coins makes me feel uncomfortable. Why or why not?	SA	A	U	D	SD

B.2 Post-assessment survey

Instructions to student: "I'm going to read some statements about working with Rosie the robot. Please listen to each one carefully. Then point to the image that shows how much you agree or disagree with the statement."

Investigator, use the following key:

SA = Strongly Agree

A = Agree

U = Undecided

D = Disagree

SD = Strongly Disagree

Present the student with the following example, and ask them to gesture to the image that represents their sentiment:

Table B.3: Example survey question.

0.	I think pizza with pepperoni is the best.	SA	A	U	D	SD

- If you are really positive that pepperoni pizza is the best, circle SA (Strongly Agree).
- If you think that it is good but maybe not great, circle A (Agree).
- If you can't decide whether or not it is best, circle U (Undecided).
- If you think that pepperoni pizza is not all that good, circle D (Disagree).
- If you are really positive that pepperoni pizza is not very good, circle SD (Strongly Disagree).

Table B.4: Post-assessment survey questions (1).

1.	I like Rosie the robot. Why or why not?	SA	A	U	D	SD
2.	Rosie did not seem to be aware of me. How could you tell?	SA	A	U	D	SD
3.	Rosie wanted me to learn. How could you tell?	SA	A	U	D	SD
4.	Rosie couldn't tell what I was doing. How could you tell?	SA	A	U	D	SD
5.	Rosie knew how to make change with coins. How could you tell?	SA	A	U	D	SD
6.	I didn't always follow Rosie's instructions. Why or why not?	SA	A	U	D	SD
7.	I trusted Rosie to give me the right instructions. Why or why not?	SA	A	U	D	SD
8.	I would like Rosie to teach me a new task in the future. Why or why not?	SA	A	U	D	SD
9.	I am good at making change with coins now. Why or why not?	SA	A	U	D	SD
10.	Making change with coins makes me feel nervous. Why or why not?	SA	A	U	D	SD
11.	I am better at making change with coins than I was before working with Rosie. Why or why not?	SA	A	U	D	SD
12.	I could understand everything Rosie said when she spoke. Why or why not?	SA	A	U	D	SD
13.	Rosie's movements helped me learn. Why or why not?	SA	A	U	D	SD
14.	Rosie's gestures were very confusing. Why or why not?	SA	A	U	D	SD
15.	Rosie gave good instructions. Why or why not?	SA	A	U	D	SD
16.	Learning from Rosie was difficult. Why or why not?	SA	A	U	D	SD
17.	I did not enjoy Rosie the robot. Why or why not?	SA	A	U	D	SD

Table B.5: Post-assessment survey questions (2).

18.	Rosie appeared aware of me. How could you tell?	SA	A	U	D	SD
19.	Rosie did not care how I performed. How could you tell?	SA	A	U	D	SD
20.	Rosie understood what I was doing. How could you tell?	SA	A	U	D	SD
21.	Rosie was not very good at making change with coins. How could you tell?	SA	A	U	D	SD
22.	I did what Rosie told me to do. Why or why not?	SA	A	U	D	SD
23.	I don't think Rosie always gave me the right instructions. Why or why not?	SA	A	U	D	SD
24.	I do not want to learn from Rosie again. Why or why not?	SA	A	U	D	SD
25.	I still have difficulty making change with coins. Why or why not?	SA	A	U	D	SD
26.	I am comfortable making change with coins. Why or why not?	SA	A	U	D	SD
27.	After working with Rosie, I'm about the same at making change with coins as I was before. Why or why not?	SA	A	U	D	SD
28.	Rosie was hard to understand. Why or why not?	SA	A	U	D	SD
29.	Rosie's arms got in my way. Why or why not?	SA	A	U	D	SD
30.	Rosie's instructions were not very useful. Why or why not?	SA	A	U	D	SD
31.	Rosie was easy to learn from. Why or why not?	SA	A	U	D	SD

Appendix C

Speech

The speech system implemented scales to allow any number of processes to interpret the audio stream. Multiple ROS nodes were run to recognize speech in different contexts, which the behavior software could interpret (or ignore) as appropriate for the situation. The following are the corpuses for each listener node.

C.1 Yes/no corpus

This corpus was used in all IRI experiments for questions that expected a yes or no response.

Corpus: yes, affirmative, fine, good, okay, true, yea, of course, yeah, yep, yup, got it, sure, no, nope, negative, false, never, no way.

C.2 Coin corpus

This corpus was used the *making change* experiment.

Corpus: penny, pennies, dime, dimes, quarter, quarters, nickel, nickels.

C.3 Number corpus

This corpus was used the *making change* experiment.

Corpus: one, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, fourteen, fifteen, sixteen, seventeen, eighteen, nineteen, twenty, thirty, forty, fifty, sixty, seventy, eighty, ninety, hundred, thousand, million, billion, trillion, and, point.

C.4 Start/stop corpus

This corpus was used in all IRI experiments for interaction questions that expected a yes/no or ready/not ready-type response.

Corpus: yes, affirmative, amen, fine, good, okay, true, yea, all right, aye, by all means, certainly, definitely, even so, exactly, gladly, granted, indeed, indubitably, just so, naturally, of course, absolutely, positively, precisely, sure thing, surely, thumbs up, undoubtedly, unquestionably, very well, willingly, yeah, yep, yup, start, go, begin, ready, stop, end, exit, quit, no, nope, negative, nix, absolutely not, false, by no means, never, no way, not at all, thumbs down, rosie.

C.5 Control corpus

This corpus was used in all IRI experiments for the purpose of monitoring for control commands as discussed in Section 4.5.

Corpus: stop, emergency, exit, pause, wait, hold, quit.

Vita

Christopher Reardon received the B.S. degree in computer science from Berry College, Georgia, in 2002, and the M.S. degree in computer science from the University of Tennessee in 2008, where he completed his Ph.D. degree in May, 2016. Previously, Christopher was employed as a Programmer Analyst by the University of Tennessee. Christopher's research interests include intelligent systems and human-robot interaction. Christopher was awarded the Chancellor's Fellowship by the University of Tennessee and the Chancellor's Award for Extraordinary Professional Promise. During his time at the University of Tennessee, Christopher worked as a research intern at both GE Global Research and the U.S. Army Research Laboratory. After completing his Ph.D., he intends to continue research in robotics and intelligent systems.