# Robots Can Teach Students With Intellectual Disabilities: Educational Benefits of Using Robotic and Augmented Reality Applications

Christopher Reardon[1], Hao Zhang[2], Rachel Wright[3], and Lynne E. Parker[4]

In the face of a worldwide teacher shortage, and a critical shortage of special education teachers in the United States, there is an urgent demand for educational resources. For people with intellectual and developmental disabilities (I/DD) in particular, there is a compelling need to develop educational tools and strategies to facilitate independence and self-sufficiency, and to address poor employment outcomes in adulthood.

When provided with the capabilities to make intelligent decisions, robots and other assistive devices have a significant potential to address this problem and empower people with disabilities by providing instruction and educational support. Robots and assistive devices possess important features including situatedness, embodiment, precision, tirelessness, scalability, and context awareness that make them particularly advantageous in the instructional role.

Intelligent systems, be they robots, augmented reality systems, or other technologies, need to make decisions when interacting with humans for instruction. In this article, we introduce the use of response prompting as a basis for this decision making and discuss the construction of a complete system around this cognitive approach that includes perception and interaction. Our goal is to provide cognitive capabilities for an intelligent robot instructor (IRI), and also demonstrate its generalizability across technologies and modalities. Specifically, we teach participants advanced skills with an IRI. We also explore similar instruction strategies on a different type of hardware through the use of a portable augmented reality (AR) assistive device to teach three additional skill sets, including the simpler prerequisite skills for the tasks taught by the IRI. We present the results of a formal study showing the effectiveness of this approach for teaching college-age students with I/DD. Our results show that proven education methodologies can be leveraged to provide intelligent autonomous instruction to students with I/DD. We expand upon several findings that may be constructive towards other efforts to create IRI systems, and conclude with open challenge areas.

[1]U.S. Army Research Laboratory, Adelphi, MD 20783, USA christopher.m.reardon3.civ@mail.mil
[2]Department of Computer Science, Colorado School of Mines, Golden, CO 80401, USA hzhang@mines.edu
[3]Common Threads Resource Center, Madison, WI 53558, USA rachelw@commonthreadsmadison.org
[4]Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN 37996, USA leparker@utk.edu

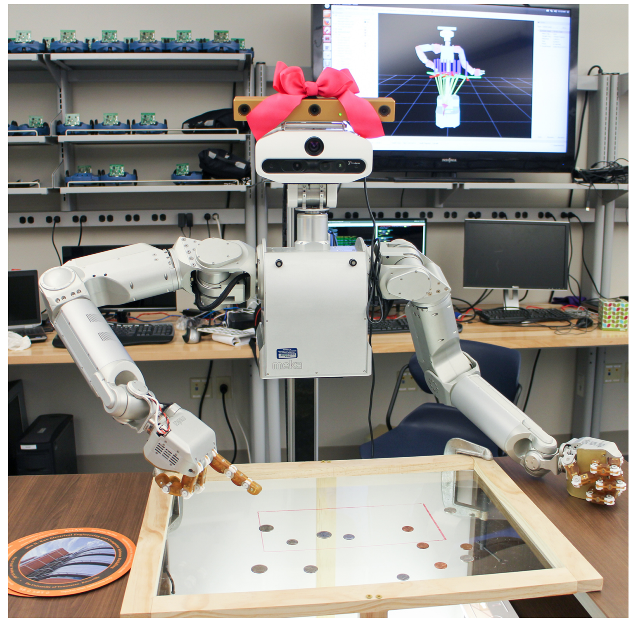


Fig. 1. Illustration of our Intelligent Robot Instructor.

## I. Intelligent Robots for Instruction

A strong case exists for investigating the use of IRIs to teach human pupils [1]. The more time a teacher spends with a student, the better the student learns. If an IRI such as the one in Fig. 1 were capable of assisting a human teacher by providing instruction to students in a classroom setting, 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 [2], [3], [4] 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 for assisting students. The precision of a robot would enable it to provide perfectly timely instructions, and would avoid issues and mistakes that human instructors face. 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 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, for purposes such as robot tutoring [5], engaging cognitively impaired and/or Alzheimer's patients for treatment [6] and obeying instructions and affording robots aspects of human interaction such as personal space [7].

We believe Students with I/DD in particular have a high potential benefit from IRIs for several reasons. Young people with I/DD entering adulthood today face harsh realities, such as low employment rates, poor wages and benefits, limited community supports, and low rates of independent living [8]. In addition to the advantages of an IRI discussed above, the types of tasks commonly taught to students with I/DD lend themselves well to robot demonstration and observation. One area, for example, is the teaching of *life skills*, which refers to knowledge or skills that increase a person's independence in personal, community, or job life. Particularly relevant to the educational experiments conducted in this work are vocational and functional academic domain task skills that one would encounter in daily life.

Robots have successfully been used to instruct students. Many such studies focused on using the robot as an instructional tool, employing Wizard-of-Oz (WoZ) approaches [9], [10] or partially-WoZ approaches [11], or using the robot as an embodied recording of speech and gestures following a script [12]. A Nao IRI was used a decision-tree-type framework to autonomously teach students in [13], [14]; however, whereas this instruction was for math skills using multiple-choice response types, our work is more ambitious in that it employs an IRI teach a variety of skills involving math, geometry, and object manipulation, and further shows that our approach generalizes to other intelligent systems (i.e., an AR device).

All of this research shows promise for robotic instructors. The high-level goals of this work are to demonstrate the following:

- Established educational methodologies can be adapted to help intelligent systems autonomously make instructional decisions
- Knowledge and representational gaps between what is defined in modern educational methodologies and a workable implementation for an intelligent, embodied system such as a robot, can be identified and addressed
- Unique aspects of robotic and intelligent systems would strongly benefit populations in great need of educational support

This work attempts to address these open issues by adapting proven methods used for instruction in the education field to the decision making process of an IRI. The approach is then used to teach students with intellectual disabilities in a one-on-one setting.

We introduced the response prompting concept for instruction in preliminary work with an IRI [15] and an AR device [16]. Here we present new prompting strategies, additional experiments, expanded results and discussion, and observations and lessons learned over this broader study. We hope that this work and the presented findings will help drive further research into the creation of intelligent instruction systems, particularly for students in elevated need, such as those with disabilities.

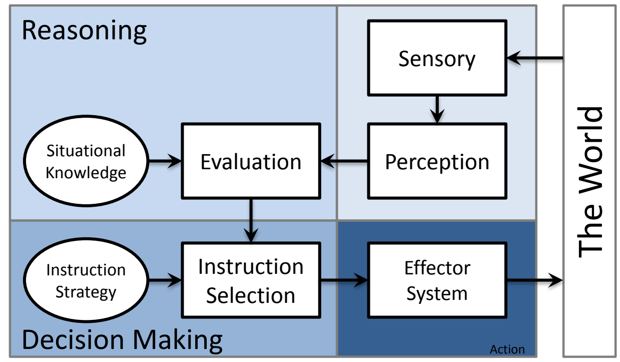



Fig. 2. Cognitive process flowchart. Information flows from lighter to darker shaded process steps.

## II. Response Prompting for Decision Making

To make intelligent decisions in the context of providing instruction, a cognitive framework is proposed that makes use of *response prompting* methodologies borrowed from the education domain [17]. Response prompting is a well-defined, evidence-based collection of teaching practices that involve assessing the environment and acting or providing assistance to stimulate a targeted behavior. By incorporating these proven methodologies into our approach, an intelligent system is able to use the same methods to provide instruction that a teacher would use.

Prompt response strategies allow for multiple modalities that can be performed by an IRI. Common prompt modalities are vocal, visual, gestural, modeling (i.e., demonstrations), or physical prompts. Another useful feature of prompting is the ability to modify the intrusiveness of prompts, called "fading"; more intrusive prompts provide more support, whereas less intrusive prompts allow students to perform the desired behavior independently.

Importantly, these methods go far beyond simple instructional systems: they involve a complex hierarchy of prompts, with multiple paradigms for arranging, traversing, and fading those prompts based on instructional needs and student performance. Through these structured designs, they assure student performance moves toward independence.

Leveraging the response prompting instructional strategies in the cognitive system represents a novel approach to an intelligent robotic and augmented reality system for instruction. We believe examination of this approach provides valuable and constructive evidence for the instruction of students by autonomous robots.

### A. Cognitive Process Overview

Figure 2, inspired by [18], illustrates the cognitive framework at a high level. In the cognitive process, by taking sensory information from the world and *perceiving* information salient to the task at hand, an interpretation of the states

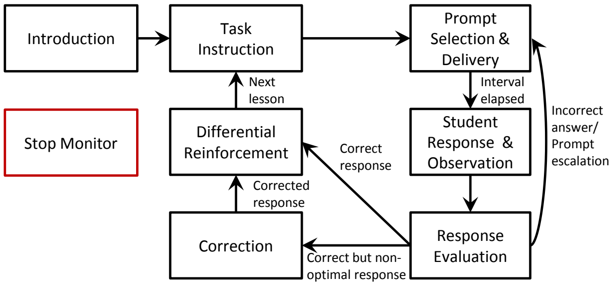


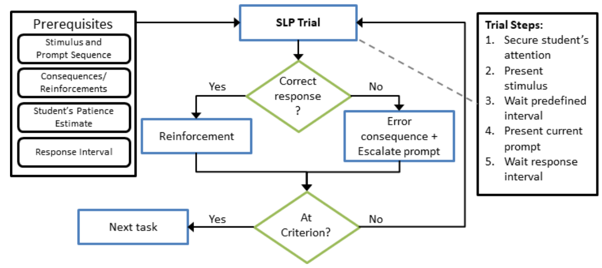

Fig. 3. The overall instruction process based on response prompting methodology.

and actions of the world is created. Then, using contextual knowledge that interpretation is *reasoned* upon to generate a higher-level representation of the situation, and a decision is made. In the intelligent instruction application, this involves first using the response prompting methodology to select the correct instruction response (e.g., present stimulus, prompt, consequence, reinforcement) and then performing the selected action.

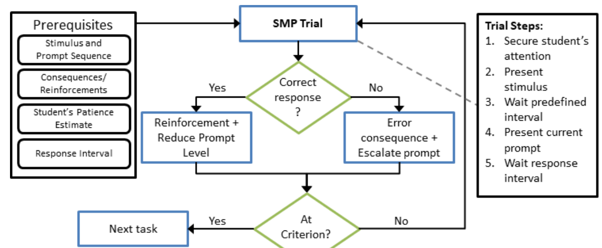
The general instruction process is shown in Fig. 3. The instructional intervention begins by giving an introduction and general instructions for the scenario. In Task Instruction the task is introduced and the target stimulus is presented. Next, the prompt is selected. The intrusiveness of the prompt depends on the methodology being employed (Sec. II-B) and the previous prompt/response history (e.g., prior incorrect responses). Next is the Student Response period, which ends when the student is idle, the task is complete, or the response interval has elapsed. The Response Evaluation determines the outcome of the interaction: a correct answer results in positive reinforcement; for a correct but non-optimal answer a correction occurs before positive reinforcement; and in the case of an incorrect answer the evaluated result information is used as part of the process to select the appropriate feedback. The type of response (correct, incorrect, partially correct, or no response) combined with the known information about the previous prompt and student's state (active or inactive), ultimately determines the appropriate feedback. Only positive reinforcement is used. The reinforcement used is *differential*, in that the degree of reinforcement is inversely proportional intrusiveness of the required prompt; i.e., the lower the level of prompt intrusiveness required, the more positive the reinforcement.

### *B. Response Prompting Types*

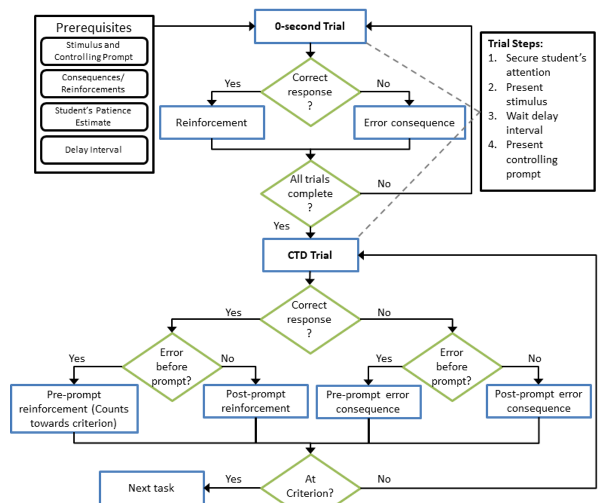
System of Least Prompts (SLP), System of Most Prompts (SMP), and Constant Time Delay (CTD) [17] are response prompting strategies that are popular and well-validated in education for instruction of chained and discrete tasks, and have been successfully used to teach students with I/DD. Further, these methods are applicable to a wide range of students and have been shown effective at teaching a large variety of skills. For this work we computationally encoded SLP, SMP, and CTD for use in an IRI's system's decision making process. We also demonstrate generalizability on an AR system.




(a) System of Least Prompts




(b) System of Most Prompts




(c) Constant Time Delay

Fig. 4. Flowcharts for System of Least Prompts (4a), System of Most Prompts (4b), and Constant Type Delay (4c) response prompting strategies, adapted from [17]. Note a key difference between SLP and SMP is in the prompt sequence input: SLP is least to most intrusive (Sec. II-B.1); SMP is most to least intrusive (Sec. II-B.2). CTD has only one prompt, the controlling prompt, and the time interval is varied (Sec. II-B.3).

*1) System of Least Prompts:* In the SLP method of instruction, a hierarchy of prompts is arranged from least to most intrusive. At the least intrusive end, no prompt is given. At the most intrusive end, the *controlling prompt* is given. The controlling prompt is designed to guarantee that the task is successfully performed. The prompt hierarchy is traversed iteratively to provide more support as needed. At each iteration, the *target stimulus* (e.g., the question) is given with the prompt for the current level. The response is evaluated after the "response interval" – a constant amount of time before and after each prompt – elapses. A correct response is reinforced; an incorrect response results in an escalation of the prompt level and another iteration.

The goal of SLP is that, after sufficient repetitions, students require fewer prompts and eventually respond correctly before any prompt is delivered. As the student answers correctly at lower prompt levels, a process of "self-fading" occurs, where the student's answers determine the rate at which the intrusiveness is decreased.

We note here that the occupational therapy technique of *graded cueing* has been previously used successfully for imitation-games in therapy for people with autism [19] and is similar to the SLP subtype.

*2) System of Most Prompts:* System of Most Prompts (SMP) is very similar to SLP, except that in SMP the hierarchy of prompts is arranged from most to least intrusive. 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.

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 between this and SLP is that with SMP, it is highly likely that the entire prompt hierarchy is traversed for each instruction. This 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.

*3) Constant Time Delay:* In Constant Time Delay (CTD), prompts are delivered after a time delay following a task direction, which is a cue or question for the student. Unlike SMP or SLP, there is only one prompt: the controlling prompt. Only the time between the cue and the prompt, or "prompt delay interval," is varied. 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. Consistently correct response before the prompt is the goal of CTD.

### C. Chained and Discrete Tasks

In addition to prompting strategies, the manner in which the steps of the task can be taught is also a consideration. 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, such as 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 how they impact successful learning from intelligent systems.

### D. Creation of Prompting Hierarchies

The process of creating a prompt hierarchy involves creating a series of prompts and arranging them in a hierarchical structure that is appropriate for the instruction strategy (e.g., SLP or SMP). The dimensions of this structure are dictated by the student's response space (e.g., correct, partially correct, incorrect, or no response), the discretization of task steps, and the modalities and intrusiveness of the response prompts.

To ensure successful instruction, the prompt content and modalities, as well as level of intrusiveness, should be appropriate for the student's capabilities and diagnoses. For these reasons, we strongly recommend collaboration with an education domain expert for this process. For the following experiments, the interdisciplinary team of authors collaborated closely to ensure a successful and productive experience for the involved students.

## III. Experiments

### A. Single-Case Experimental Design

Experiments were developed for this research using single case experimental design (SCED), which is a common design method in special education research. As opposed to comparison between groups or subjects, participants serve as their own control data for the purpose of comparing performances between at least two experimental phases [20]. SCED methods are used in place of statistical methods for large groups. This is not only because recruiting a large number of participants with I/DD is infeasible, but also because even if it were possible their capabilities and diagnoses could be so diverse that drawing even a coarse statistical inference would be challenging.

The objective of SCED is to determine if a causal or functional relation exists between the delivery of the independent variable (IV) – the intelligent instruction system – and significant increases in the dependent variable (DV) – the acquisition and maintenance of the skills taught.

Two types of SCED designs were used for this research: combined multiple baseline across participants and combined multiple baseline across skills. These designs allow 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 alternate designs are not feasible, such as those that would require withdrawal of skill knowledge.

In experiments utilizing a multiple baseline across skills design, skills are taught one at a time, and instruction is introduced for each skill sequentially after learning the previous skill. In experiments using multiple baseline across participants, each student is taught one at a time, and instruction is introduced to each successive student after the previous student finishes learning the skill. In all cases, baselines are taken before instruction, up to the point where

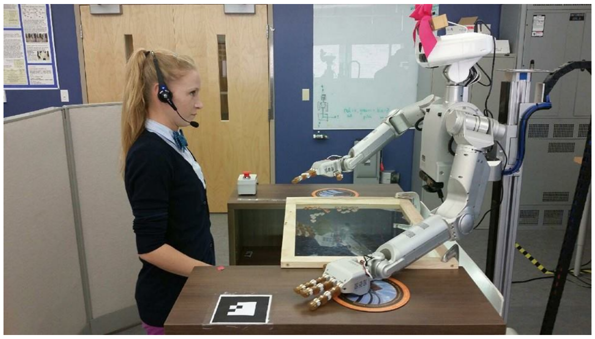


Fig. 5. The interaction setting for instruction.

Fig. 6. The object tracker GUI, with live, adjustable parameters on top left and the annotated live image on the bottom left. Right zoom shows an enlarged view of the annotated image. Annotations include position, orientation, size, centroid location, and bounding box for each object.

the instruction begins, and probes are made after successful demonstration of the skills to measure retention.

By introducing the intervention subsequently across a minimum of three replications, 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 [21].

### B. Student Participant Population

The students[1] who participated in these studies were all college-age and attendees of a post-secondary education (PSE) program designed for young adults with I/DD at the University of Tennessee, Knoxville named FUTURE.[2] All students were aged 18 to 34 with IQ between 57 and 67, received special education services throughout school, and earned modified high school diplomas prior to participating in the FUTURE program.

To meet the three replication minimum requirement to evaluate the intervention effects, three students per experiment were taught until criteria was reached, i.e. they mastered the skill being taught. Seven students participated overall, four of which participated in multiple experiments.

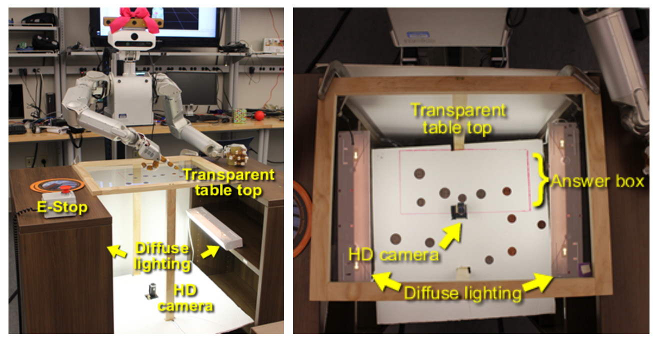
### C. Intelligent Robot Instruction Experiments

Life skills that require object manipulation and discrimination were taught in these experiments. In our setting, the student and the IRI stand across from each other at a table during instruction, as seen in Fig. 5. The IRI autonomously performs the instruction, prompting, observation, evaluation, and feedback (correction or reinforcement) loop shown in Fig. 3. No WoZ techniques were used for these experiments. The complete IRI system was implemented as a suite of C++ and Python software modules, leveraging ROS[3] for messaging, interprocess communication, and common robotics libraries. A custom object-tracking system shown in Fig. 6 provided the IRI with the ability to observe objects being interacted with by both the student and the robot and to accurately interpret the student's performance. The object tracking system was implemented as a set of custom ROS

[1]All studies for this research were performed in accordance with IRB protocols and approval.

[2]More information on the FUTURE program is available here: `http://futureut.utk.edu`.

[3]`http://ros.org`



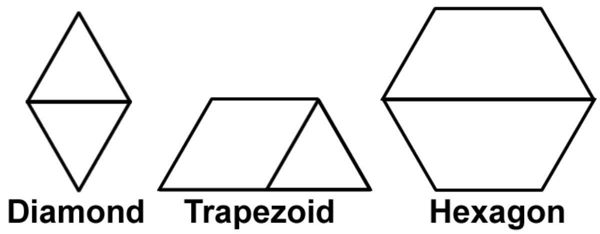

(a) (b)

Fig. 7. The table setup for the instructional setting from the student view (a) and overhead (b).

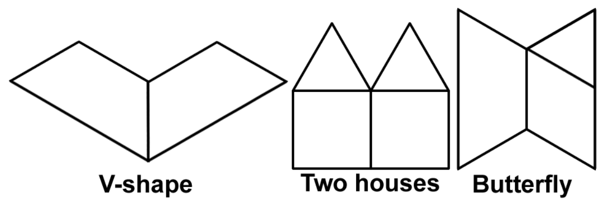
nodes using OpenCV to process live image streams from a camera mounted under a transparent tabletop, as illustrated in Fig. 7. Details of the vision system performance are reported in [15], where we previously introduced response prompting for instruction with an IRI. Interaction and feedback is provided through synthesized speech, speech recognition, and gestures.

The robotic hardware for this research is a Meka Robotics M3 mobile humanoid robot (Fig. 7a) with 7 degree-of-freedom (DOF) arms, 5-DOF hands, and a sensor head with 2-DOF movement. For this research, the IRI makes use of one PrimeSense short-range (v1.09) camera, one USB camera, a Bluetooth microphone, and stereo speakers. The robot is equipped with two PCs – one providing real-time functionality of the base, arms, hands, and lift; the second was dedicated to the vision and audio components.

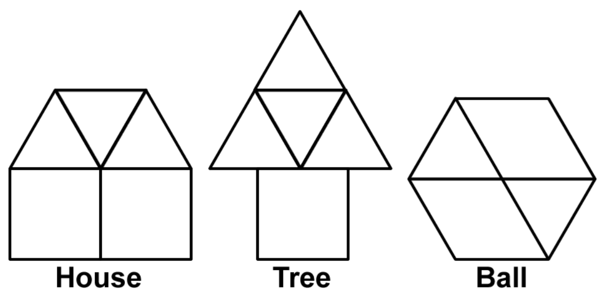
Two skills were taught by the IRI: 1) *making change* i.e., given a dollar and a purchase price, first use a calculator to determine how much change is due, then present the IRI with the correct change in coins (originally presented in [15] and summarized below), and 2) *geometric assembly* of larger objects from smaller pieces, as shown in Fig. 8, for which prerequisite skills were taught using the AR instruction (Sec. III-D). Both skills involved live interaction with objects and




(a) Subcomposition




(b) Symmetry




(c) Complex assembly

Fig. 8. Puzzle tasks for *geometric assembly* skill.

TABLE I
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 |
| | C | Differential positive reinforcement |
| | 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 |
| | C | Differential positive reinforcement |
| | 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 |
| | C | Differential positive reinforcement |
| | 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 |
| | C | Differential positive reinforcement |
| | I | Gesture to each coin to remove, wait until removed, then same as NR |

demonstration of correct responses.

The *making change* skill is a life and vocational skill that involves calculating the correct quantities and denominations of currency to be exchanged after a cash transaction. The assembly skills taught in our *geometric assembly* experiment are useful in job settings, and the geometric reasoning aspects are useful in both work and personal life.

Instruction strategies were selected that are appropriate to the skill being taught. For the *geometric assembly* skill instruction, a SMP prompting strategy was used with backwards chaining, along with a multiple baseline across skills experimental design. For the *making change* experiment, a combination of SLP and discrete and forward chaining was used. Unlike the *geometric assembly* and AR experiments, the experimental design for the *making change* experiment was multiple baseline across participants, to control for any outside influences on participants' ability to make change.

An example prompt hierarchy for part of the *making change* task is shown in Table I. The SLP strategy was employed, 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*. The student responses shown are: NR - No Response, PC - Partially Correct response, C - Correct response, I - Incorrect response.

In addition, 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 before and after interaction with the IRI for the *making change* experiment. Results of this survey are discussed in Sec. IV-C.

### *D. Intelligent Augmented Reality Instruction Experiments*

We also implemented the response prompting for intelligent instruction concept on a portable augmented reality device. AR devices share similar features to an IRI instruction, such as context-awareness, precision, and tirelessness of an intelligent system situated in the environment with the user. Both are capable of delivering visual and auditory prompts to the student that interact with the real environment. Using an AR device for instruction, we demonstrate that our approach generalizes to systems with these overlapping features but a different physical implementation (i.e., a head-mounted AR device vs. a humanoid robot).

Figure 9 shows an overview of the AR system. From the student's perspective, when wearing an AR device and learning a new sequential task he or she can ask for help with the next step in the sequence at any time. The AR device captures an image from the student's point of view, processes it, and presents an appropriate instructional prompt to the student via the AR device. In this experiment, the prompt modalities are an augmented version of the uploaded image as well as audio. Live video augmentation is outside the performance abilities of our hardware implementation, but

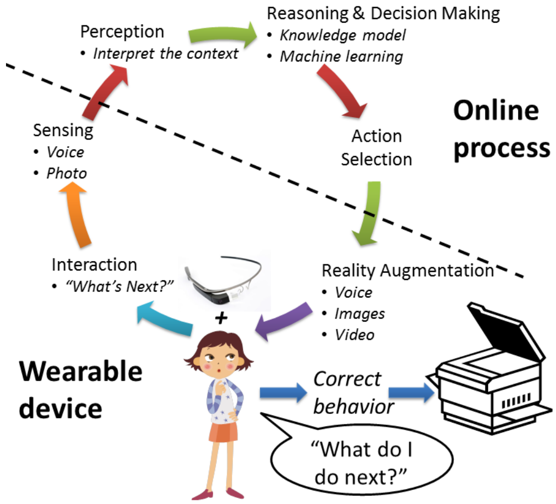



Fig. 9. AR system overview. Student is instructed to face the device and ask for assistance. A picture taken from the wearable's camera is classified and an augmented image and audio containing the proper instructional prompt for the next step is delivered.

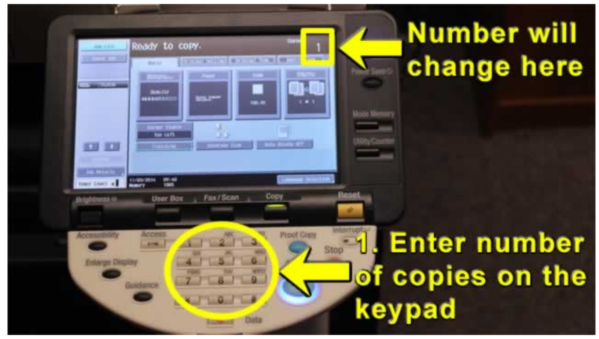
could be included in future work.

To provide this instruction, the system parses the image for relevant information before applying supervised learning to solve the problem of identifying the correct context of the image. Classifier output combined with the task's knowledge model allows the selection of the correct prompt for the next step in the task, which is delivered seamlessly through the AR interface to the student.

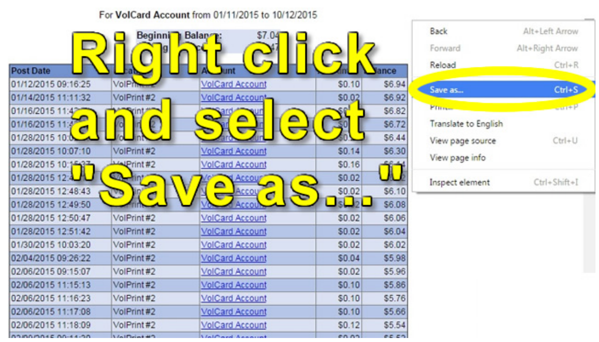
Our AR system implementation consisted of a Google Glass wearable AR device running a simple Android application with a streamlined interface. The simple audio command, "Okay Glass, what's next?" triggers the app. The user clicks to take a picture, which is uploaded to the cloud server, and an image and audio instructional prompt is provided via the Glass display and built-in speaker within 5-10 seconds. On the cloud side, intelligent instruction is made possible using OpenCV for image processing and Support Vector Machines (SVMs) trained on extracted visual features for image classification. In the event of a failed or low-probability classification, the user is presented with a prompt to try again.

Three skills were taught using the AR device: 1) using a *copy machine*, 2) accessing one's *student account statement* online, 3) performing *geometric reasoning* tasks with puzzle blocks to acquire necessary prerequisite skills for advanced skills taught by the IRI system. Fig. 10 shows example annotated image prompts from these experiments. A summary of these experiments appeared in [16].

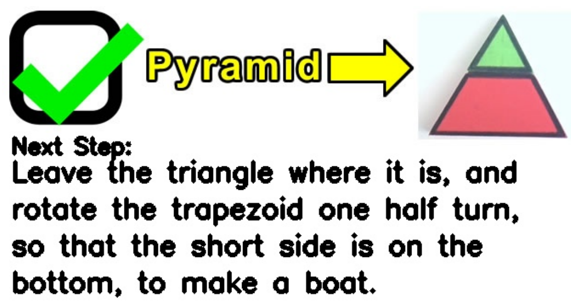
The *copy machine* skill is an office vocational work skill where students were asked to make a specific number of double-sided copies of a document on a commercial copy machine with a complex, less-than-intuitive user interface. The *student account statement* skill is an employment and independent living skill where the student was asked to login and retrieve a copy of their prepaid student account




(a)




(b)




(c)

Fig. 10. Example annotated image prompts from the AR instruction experiments: 10a shows a step from the *copy machine* study; 10b shows a step from the *student account statement* study; and 10c shows a step in the *geometric reasoning* study, where a student is learning the *rotate* and *half turn* subskills. Image prompts are shown to the user via the AR device and accompanied by audio prompts.

statement. The third experiment, *geometric reasoning*, taught object manipulation, placement, orientation, and assembly sub-skills that are highly vocationally relevant.

For these experiments, instructional strategies were carefully designed using the appropriate combination of prompts, prompting strategy, and prompt chaining, for each skill taught *and* the AR technology method (Table II). These combinations of methods were selected with expert consultation to match the appropriate prompt strategy and chaining direction to the structure of the skill. All three AR experiments used a multiple baseline across skills design, where each sub-skill or step was mastered before the successive one was introduced. Also in all three, the skills were taught in forward-chained order since the tasks (e.g., navigating through a copier interface) can only progress in a forward direction. The *geometric reasoning* skill also

TABLE II

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 participants | Discrete and forward chaining |
| | Geometric assembly | SMP | Multiple baseline across skills | Backwards chaining |

employed discrete sub-skills at some steps in the chain. For prompting strategy, *geometric reasoning* employed the Constant Time Delay (CTD). The other two AR experiments used a simplified "self-directed" method that only provided controlling prompts when the students requested assistance from the AR device. This was similar to CTD but with the student controlling the time delay.

## IV. RESULTS AND DISCUSSION

Results from each experiment showed that in all cases and for all subjects, using intelligent robot and augmented reality instruction, the students were able to learn the skills to mastery.

Results from the IRI-instructed *making change* skill are shown in Fig. 11. These results show that Student 3 had a steeper learning curve due to a more limited understanding of the prerequisite coin value identification and direction-following abilities than Students 1 and 2. Despite this, using the SLP methodology combined with forwards chaining, the IRI was able to provide increasingly supportive prompts to help all students to mastery, as discussed in Sec. IV-A.

Results from the IRI-instructed *geometric assembly* skill and the ARI-instructed skills are shown in Fig. 12. Performance in the *geometric assembly* (left column) shows strong success that can be largely attributed to the structure of task instruction using SMP with backwards chaining, as discussed in Sec. IV-B.

The results from the ARI experiments (right column) verify the generalizability of the proposed approach to other interaction domains. Indeed, as discussed in Sec. IV-D, students were able to make rapid gains with ARI instruction, which we attributed to the self-directed prompt control combined with the immediacy of intelligent, context-aware instruction in augmented reality. The results presented here are a significant expansion of preliminary work [15], [16] that introduced the IRI concept.

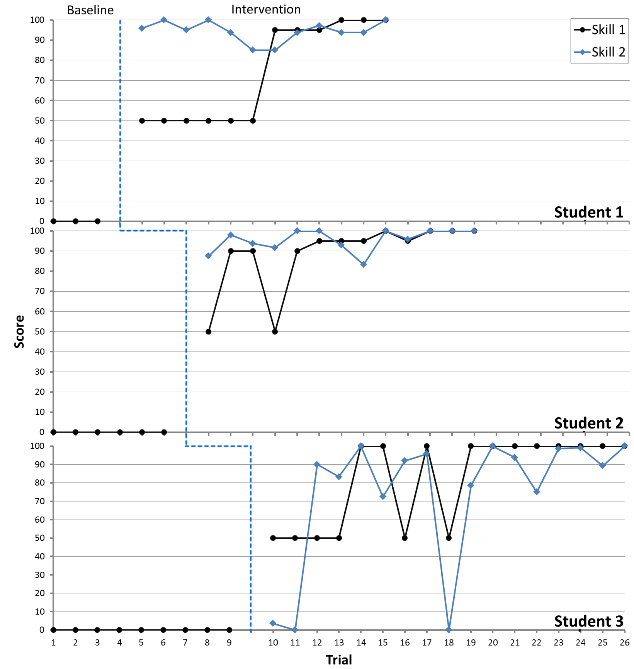



Fig. 11. Results from the IRI-instructed *making change* experiment using multiple baseline across participants. Data is collected in two phases, shown separated by dashed lines: 1) baseline, measuring each participant's performance prior to instruction and 2) intervention, where students performed the 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. Scoring methodology is detailed in [15].

TABLE III

EXPERIMENT RESULTS - TOTAL ERRORS.

| Experiment | | Total Errors Student 1 | Student 2 | Student 3 |
|---|---|---|---|---|
| Copy machine | | 0 | 0 | 3 |
| Student account | | 3 | 0 | 1 |
| Geometric reasoning | | 4 | 1 | 2 |
| Making change | Calculator | 9 | 9 | 6 |
| | Coin response | 9 | 6 | 14 |
| Geometric assembly | Subcomposition | 0 | 0 | 2 |
| | Symmetry | 0 | 3 | 4 |
| | Complex assembly | 2 | 4 | 11 |

To enable a comparison across the diverse skills and experimental configurations, Table III summarizes the total number of errors the students made before mastering the corresponding skill, or sub-skills in the *making change* and *geometric assembly* experiments (Fig. 8). The values in Table III represent the total number of errors each student made in across all trials for each skill.

While it is not possible to determine a superior strategy that generalizes to all cases, we can observe that error rates are influenced by the complexity of the skill. Lower error rates occurred with simpler skills, such as *copy machine*, and higher rates occurred with the highly complex *making change* skill, which was divided into sub-skills for correct use of a calculator and correct response with coins.

In addition, while the outcomes were very positive for those students involved, careful decisions were made to

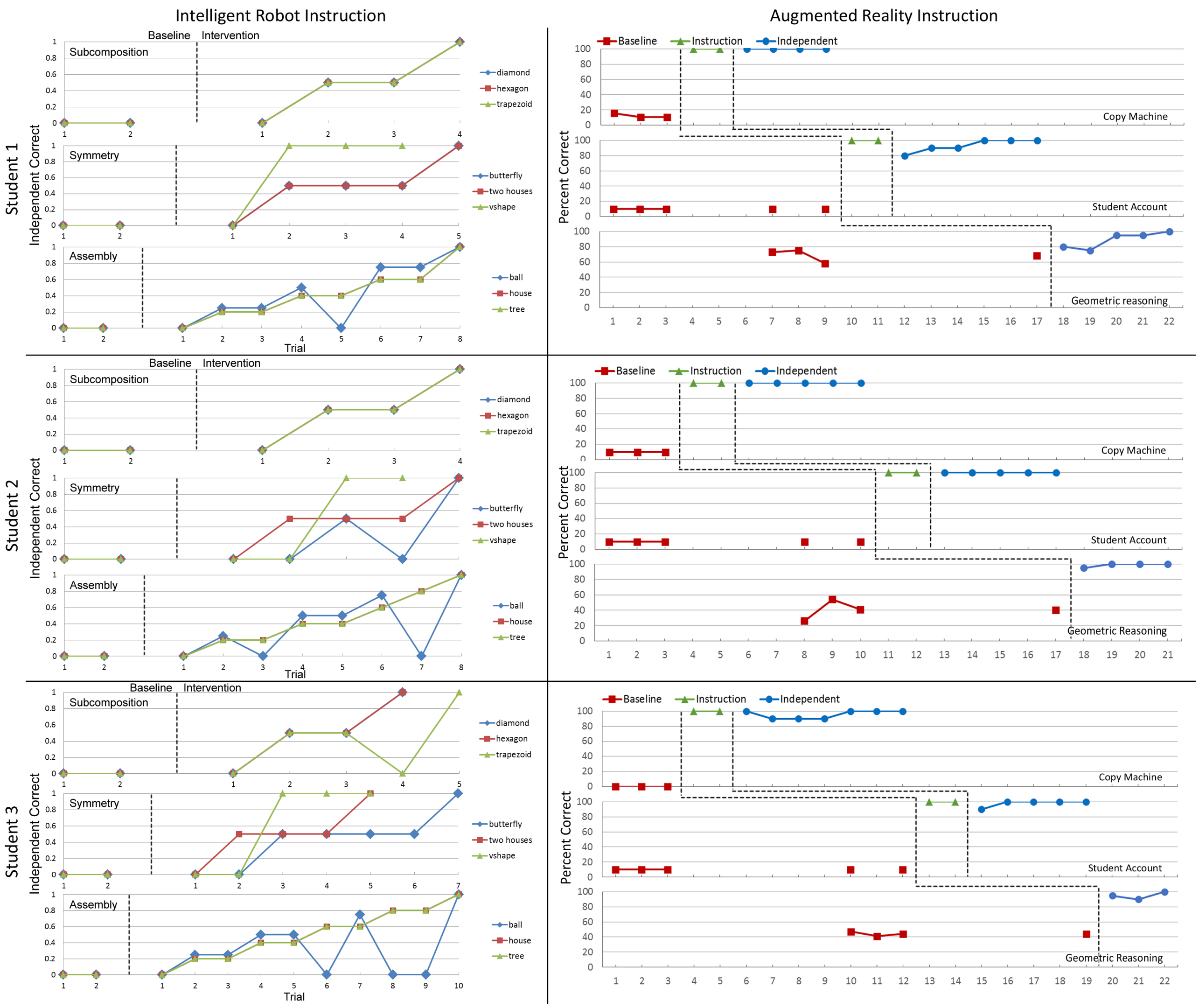



Fig. 12. Left column: IRI *geometric assembly* skill results, where three skills were taught, with three puzzles per skill (see Fig. 8). Non-increasing performance generally represents errors, because when using the SMP + backwards chaining method, the number of possible correct independent steps increases until the student is asked to assemble the entire figure independently. Right column: ARI *copy machine*, *student account*, and *geometric reasoning* skill results. Students are given an Instruction phase after Baseline, where they must perform the skill perfectly, followed by an Independent phase, where the students control their own prompts.

combine the best response prompting strategies, experimental design, and prompt relationship with each skill being taught to achieve the best performance, as summarized in Table II. Therefore, beyond validating our approach, we believe that an additional valuable contribution of these experiments can be found in key findings discussed in the following subsections that we hope can constructively inform future work in designing intelligent robotic and AR instruction technologies.

### A. SLP + Forwards Chaining

The first key finding is that SLP with forward chaining allows students to present an open-ended response, which can be challenging to accommodate in the IRI's perception system. In the *making change* experiment, this manifested as allowing the student to present any combination of coins (however incorrect, in some cases), for which the robot had to have a clear prompt. We found that despite this challenge, the strength of the methodology allows for the robot to provide increasingly intrusive prompts that can be designed to progressively limit the response space. Then, by concluding in the controlling prompt that only allows for the correct response, success is guaranteed.

### B. SMP + Backwards Chaining

Importantly, we found that backwards chaining, combined with SMP, is highly appropriate for instruction from a robot. This is because the state space of responses is restricted by design (i.e. it avoids the open-ended response challenge of the SLP + Forwards Chaining method described above), and because of a robot's advantages in performing repetitive tasks. Backwards chaining and most-to-least intrusive prompting means that the skill is taught backwards from a nearly complete example to total independence. The prompts delivered at each step begin with the controlling prompt and decrease in intrusiveness from there. For example, in the *geometric assembly* experiment, students were first presented with a puzzle with one piece missing, and told explicitly where to put which piece. Then, the prompt was decreased for that step down to an independent prompt (e.g., "try it yourself"). This was repeated with one less piece in the puzzle until the student was finally asked to assemble the puzzle themselves from scratch. It is worthwhile to note that in the results shown in the last row of Table III, all of the errors that occurred in *complex assembly* sub-skill involved assembling the "ball" hexagon-shaped object (Fig. 8c), which has many possible hexagon-shaped but incorrect sub-compositions; students had no difficulty assembling the other figures in this sub-skill. In light of these observations, we believe the SMP + backwards chaining approach, while very time-consuming and perhaps tedious and more mistake-prone for a human instructor, is a perfect match for the tirelessness and precision of an IRI. Additionally, this high degree of repetition can be greatly beneficial for students with I/DD.

### C. Subjective Acceptability

To determine the students' opinions on being instructed by a robot, an acceptability study using a Likert-type scale was conducted before and after the students' successful instruction with the IRI as part of the *making change* experiment.

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 by experts, with visual aids provided for response anchor points, with responses from "strongly disagree" scored at -2 to "strongly agree" scored at +2. The pre-assessment survey consists of 18 statements divided into 8 categories. The post-assessment survey consists of 30 statements in 14 categories. There are 2-3 statements each category, and a summative analysis was applied.

This study found mixed enthusiasm prior to instruction by an IRI, and positive opinions of the overall experience and performance of the IRI following instruction.

In Table IV, 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 the IRI again. They also trusted the robot, were willing to obey the robot's instructions, and found the experience positive overall. Perhaps unsurprisingly, 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 more frustrated by the longer time to acquire the skill, as reflected in her score on the ease of learning question and as discussed above.

TABLE IV

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** |

In open-ended responses, students highlighted their perceived benefits for working with IRIs, including improved pace, patient repetition, and relieving the burden on human teachers.

### D. Efficiency of Self-directed AR Instruction

It is likely that the self-directed nature of the AR instruction system is responsible for the rapid learning in the AR experiments. We found that providing the participants with the ability to control their own prompts, combined with the intelligent AR environment, likely 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.

### E. Challenge Areas

Based upon open-ended responses, student performance, and experimenter observations, several areas that are challenging for designing an intelligent instruction system were also identified. We believe the following lessons-learned observations would be informative toward future work in this area:

- Prompt design is a critical determining factor in successful instruction. Intrusiveness, verbiage, and interaction modalities should be designed with a domain expert.

- Incorporating multiple levels of interaction detail increases efficiency and trust. Different modes of interaction, e.g., controlling verbosity in repetitive interactions, can decrease cognitive load and negate the appearance of lack of intelligence, thereby increasing trust.
- Accurate perception is critical. Vision and speech recognition errors that result in incorrect feedback from the instructor could be harmful to the student's education and must be minimized. Because of variability in speech recognition performance, for example, we chose to minimize the students' verbal interaction with the robot and primarily interact via objects. Robust multimodal perception of student behavior is an open challenge.
- Securing the attention of the student and detecting idle states are essential for intelligent interaction. Determining when a student is or is not paying attention and when they are idle, either because they are uncertain of the correct response or they have completed their response, is a trivial task for a human teacher. However, this necessary step can be difficult for an intelligent robot instructor; failure in this can result in loss of trust in the IRI's abilities (reported in open-ended survey response by one student whom our IRI initially had difficulty engaging) as well as missing an instruction.

### F. Limitations of This Work

The primary limitations of this work can be grouped into issues surrounding the study population (size, variability, and experience) and the design decisions necessary (prompts, interaction modalities, verbiage) to create an IRI.

Regarding the population, the population size for this work is not statistical. As noted in Sec. III, this study uses single-subject experimental design methods instead of statistical methods for large groups because of the small, diverse population involved. In this methodology, three replications are sufficient to establish a causal relationship. Although a large population size study is infeasible, additional replications would help strengthen and understand this relationship. Further, the variability of the population of persons with I/DD is large. Because of this, future studies should be performed with the understanding that this variability could impact the rate of successful learning. The PSE program (Sec. III-B) this study's participants drew upon are exposed to technology as part of the program; this exposure could have influenced their performance.

Regarding the design decisions, as noted in Sec. IV-E, an IRI is a complex system. The researchers in this work learned early on that even small changes, e.g., in verbiage, can have significant impacts on interaction. This is particularly true for this population. This work sought to achieve a good-performing design through collaboration with experts in the education field. In addition to performing such collaborative work, future studies should seek to define and categorize the types of design decisions that could affect study outcomes.

## V. Conclusion

In this article, we described the use of response prompting for cognitive decision-making in intelligent instruction, with applications in robotics and augmented reality. We gave an overview of our interdisciplinary collaborative efforts to apply this approach to the instruction of students with I/DD, to assist with the long-term goal of empowering this population.

We shared the results of studies teaching academic and vocational lessons in life skills, including office vocational, geometric reasoning, and monetary mathematical skills. Our results showed great success using response prompting as part of an overall cognitive framework for intelligent instruction. While these results are a preliminary attempt to begin addressing some of the issues facing those with I/DD, we believe they show that this novel approach has merit in many applications. Most importantly, we believe the findings herein should be constructive towards future investigations and the development of intelligent instructive technologies, especially those that provide critical assistance to this important population.

## Acknowledgment

The authors would like to acknowledge the wonderful students, staff, and faculty of the FUTURE PSE program and the College of Education at the University of Tennessee, particularly Profs. Mari Beth Coleman and David Cihak for their technical advice and support for this project.

## References

[1] J. Kennedy, P. Baxter, and T. Belpaeme, "The robot who tried too hard: Social behaviour of a robot tutor can negatively affect child learning," in *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction*. ACM, 2015, pp. 67–74.

[2] E. Watlington, R. Shockley, P. Guglielmino, and R. Felsher, "The high cost of leaving: An analysis of the cost of teacher turnover," *Journal of Education Finance*, vol. 36, no. 1, pp. 22–37, 2010.

[3] T. Wilkin and G. Nwoke, "Career and technical education teacher shortage: A successful model for recruitment and retention," *Journal of STEM Teacher Education*, 2011.

[4] R. Mckeown, "Teacher education 1992 and 2012 reflecting on 20 years," *Journal of Education for Sustainable Development*, vol. 6, no. 1, pp. 37–41, 2012.

[5] D. Leyzberg, S. Spaulding, M. Toneva, and B. Scassellati, "The physical presence of a robot tutor increases cognitive learning gains," in *Proc. Cognitive Science*, 2012.

[6] A. Tapus, C. Tapus, and M. Mataric, "The role of physical embodiment of a therapist robot for individuals with cognitive impairments," in *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2009, pp. 103–107.

[7] W. A. Bainbridge, J. Hart, E. S. Kim, and B. Scassellati, "The effect of presence on human-robot interaction," in *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2008, pp. 701–706.

[8] M. Grigal and D. Hart, *Think College! Postsecondary Education Options for Students with Intellectual Disabilities*. Brookes Publishing Company, 2010.

[9] A. Meghdari, M. Alemi, M. Ghazisaedy, A. Taheri, A. Karimian, and M. Zandvakili, "Applying robots as teaching assistant in EFL classes at Iranian middle-schools," in *Proc. of the Int. Conf. on Education and Modern Educational Technologies (EMET-2013), Venice, Italy*, 2013.

[10] M. Saerbeck, T. Schut, C. Bartneck, and M. D. Janse, "Expressive robots in education: varying the degree of social supportive behavior of a robotic tutor," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2010, pp. 1613–1622.

[11] I. M. Verner, A. Polishuk, and N. Krayner, "Science class with robothespian: Using a robot teacher to make science fun and engage students," *IEEE Robotics & Automation Magazine*, vol. 23, no. 2, pp. 74–80, 2016.

[12] M. Fridin, "Storytelling by a kindergarten social assistive robot: A tool for constructive learning in preschool education," *Computers & Education*, vol. 70, pp. 53–64, Jan 2014.

[13] K. R. Liles and J. M. Beer, "Rural minority students' perceptions of Ms. An, the robot teaching assistant, as a social teaching tool," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 59, no. 1. SAGE Publications, 2015, pp. 372–376.

[14] K. R. Liles, D. G. Bryant, and J. M. Beer, "How can social robots motivate students to practice math?" in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2017, pp. 353–354.

[15] C. Reardon, H. Zhang, R. Wright, and L. E. Parker, "Response prompting for intelligent robot instruction of students with intellectual disabilities," in *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2015, pp. 784–790.

[16] C. Reardon, R. Wright, D. Cihak, and L. E. Parker, "Intelligent context-aware augmented reality to teach students with intellectual and developmental disabilities," in *International Florida Artificial Intelligence Research Society Conference (FLAIRS)*, 2016.

[17] M. Wolery, M. Ault, and P. Doyle, *Teaching students with moderate to severe disabilities: Use of response prompting strategies*. Longman, 1992.

[18] D. Isla, R. Burke, M. Downie, and B. Blumberg, "A layered brain architecture for synthetic creatures," in *International Joint Conferences on Artificial Intelligence*, 2001.

[19] J. Greczek, E. Kaszubski, A. Atrash, and M. Mataric, "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)*. IEEE, 2014, pp. 561–566.

[20] D. L. Gast and A. D. Spriggs, "Visual analysis of graphic data," *Single Subject Research Methodology in Behavioral Sciences*, pp. 199–233, 2010.

[21] R. H. Horner, E. G. Carr, J. Halle, G. McGee, S. Odom, and M. Wolery, "The use of single-subject research to identify evidence-based practice in special education," *Exceptional Children*, vol. 71, no. 2, pp. 165–179, 2005.