

By Christopher Reardon,  
Hao Zhang, Rachel Wright,  
and Lynne E. Parker

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

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

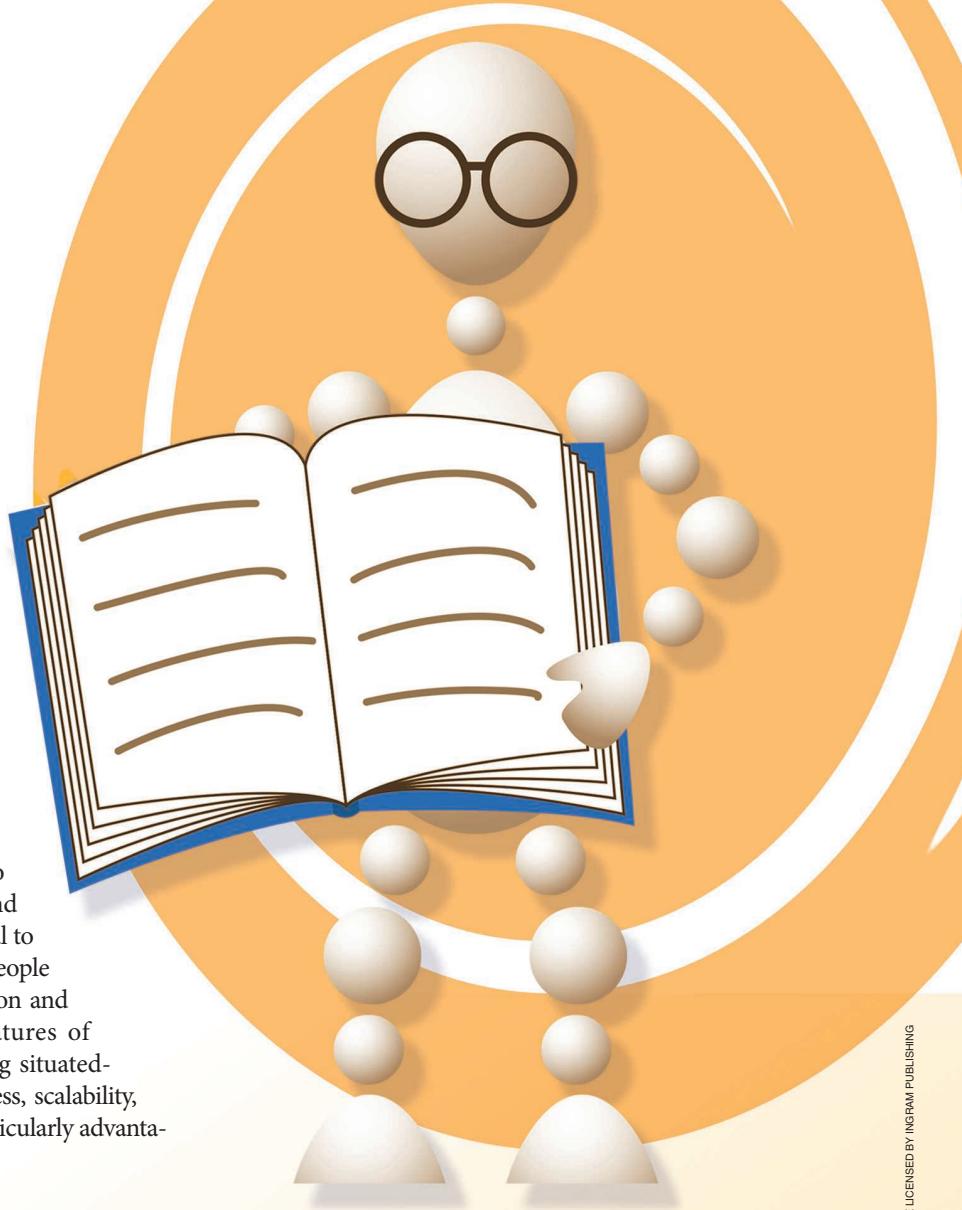


IMAGE LICENSED BY INGRAM PUBLISHING

# Robots Can Teach Students With Intellectual Disabilities

*Educational Benefits of Using Robotic and Augmented Reality Applications*

Digital Object Identifier 10.1109/MRA.2018.2868865

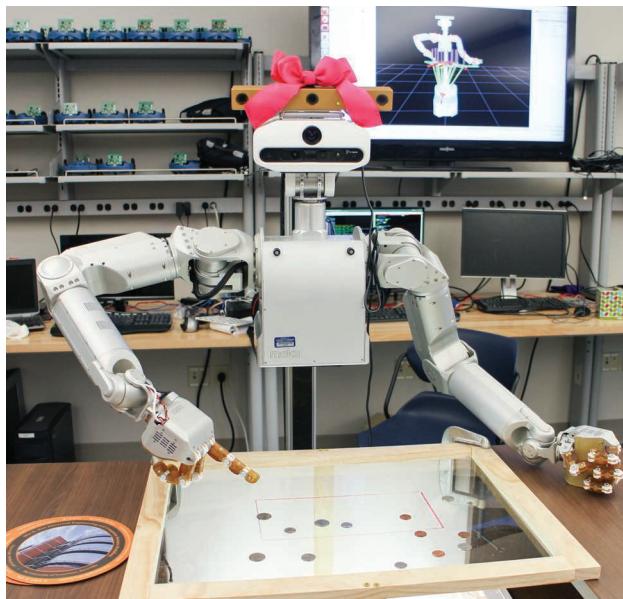
Date of publication: 15 October 2018

Intelligent systems, whether robots, augmented reality (AR) systems, or other technologies, need to make decisions when interacting with humans to facilitate 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, which 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 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 demonstrating the effectiveness of this approach for teaching college-age students with I/DDs. Our results show that proven education methodologies can be leveraged to provide intelligent autonomous instruction to students with I/DDs. We expand on several findings that may be constructive in other efforts to create IRI systems and, finally, discuss open challenge areas.

### 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 Figure 1 were capable of assisting a human teacher by providing instruction to students in a classroom setting, it could offload some of the teacher's tasks, 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]–[4], the ability of an IRI to augment a human instructor's instruction could allow for better use of limited (human) teaching resources.



**Figure 1.** A photo of our IRI.

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 provide sustained attention in assisting students. A robot's precision enables it to provide perfectly timed instruction and avoid issues and mistakes to which human instructors are prone. One robot could potentially observe and instruct large numbers of students simultaneously. Pupils may perceive a robot 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 young people already comfortable with using technology for learning purposes, a robot could represent an embodied and more physically interactive tool 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 that students—and, in particular, those with I/DDs—have significant potential to benefit from IRIs for several reasons. Young people with I/DDs entering adulthood today face harsh realities, such as low employment rates, poor wages and benefits, limited community support, and low rates of independent living [8]. In addition to the advantages of an IRI discussed previously, the types of tasks commonly taught to students with I/DDs lend themselves well to robot demonstration and observation. One area, for example, is the teaching of life skills, i.e., knowledge and/or abilities that increase a person's independence in personal terms, for community engagement, and on the job. Particularly relevant to the educational experiments conducted for this article are vocational and functional academic domain task skills one would encounter in daily life.

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

All of this research shows promise for robotic instructors. The high-level goals of this article 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 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 employed to teach students with intellectual disabilities in a one-on-one setting.

We have already 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 article and the presented findings will help drive further research into the creation of intelligent instruction systems, particularly for students with elevated needs, such as those with disabilities.

### 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 a teacher would use to provide instruction.

Prompt/response strategies allow an IRI to perform in multiple modalities. 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. These structured designs assure that student performance moves toward independence.

Leveraging response-prompting instructional strategies in the cognitive domain represents a novel approach to an intelligent robotic and AR system for instruction. We believe that examining this approach provides valuable and constructive evidence for instruction by autonomous robots.

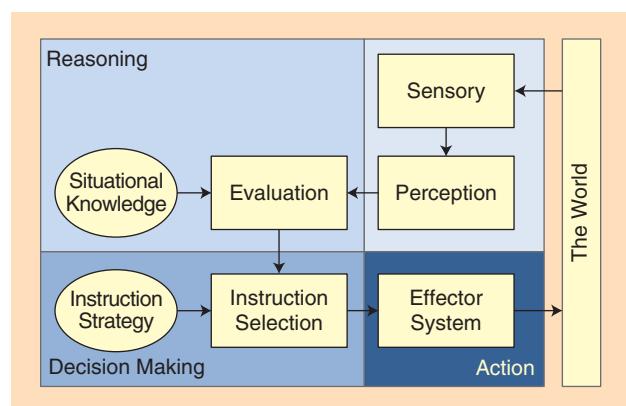
### Cognitive Process Overview

Figure 2, inspired by [18], illustrates the cognitive framework at a high level. In the cognitive process, it is by taking sensory information from the

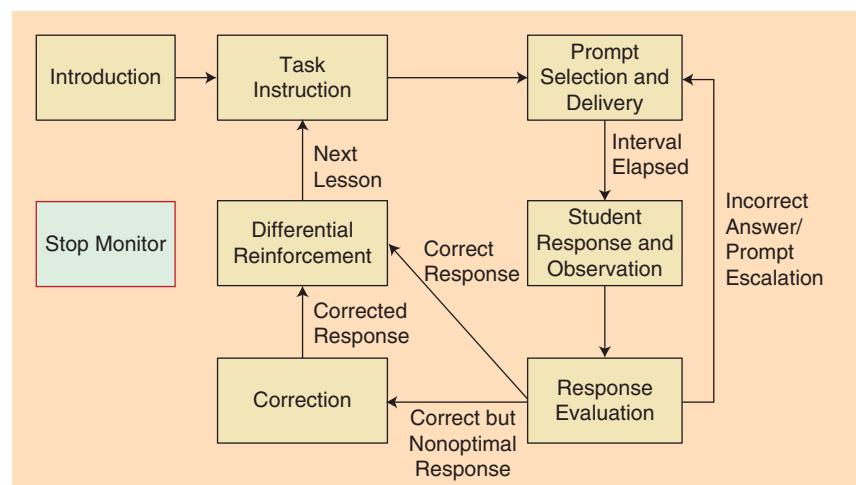
world and perceiving information salient to the task at hand that we create an interpretation of the states and actions of the world. Then, using contextual knowledge, that interpretation is reasoned about to generate a higher-level representation of the situation: i.e., a decision is made. In the intelligent 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 Figure 3. The instructional intervention begins with an introduction and general instructions for the scenario. In task instruction, the task is introduced and the target stimulus presented. Next, the prompt is selected. The intrusiveness of the prompt depends on the methodology being used (see the “Response-Prompting Types” section) 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 nonoptimal answer, a correction occurs before



**Figure 2.** The cognitive process flowchart. The information flows from lighter- to darker-shaded process steps.



**Figure 3.** The overall instruction process based on the response-prompting methodology.

positive reinforcement; 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. This reinforcement is differential, in that the degree of reinforcement is inversely proportional to the intrusiveness of the required prompt, i.e., the lower the level of prompt intrusiveness required, the more positive the reinforcement.

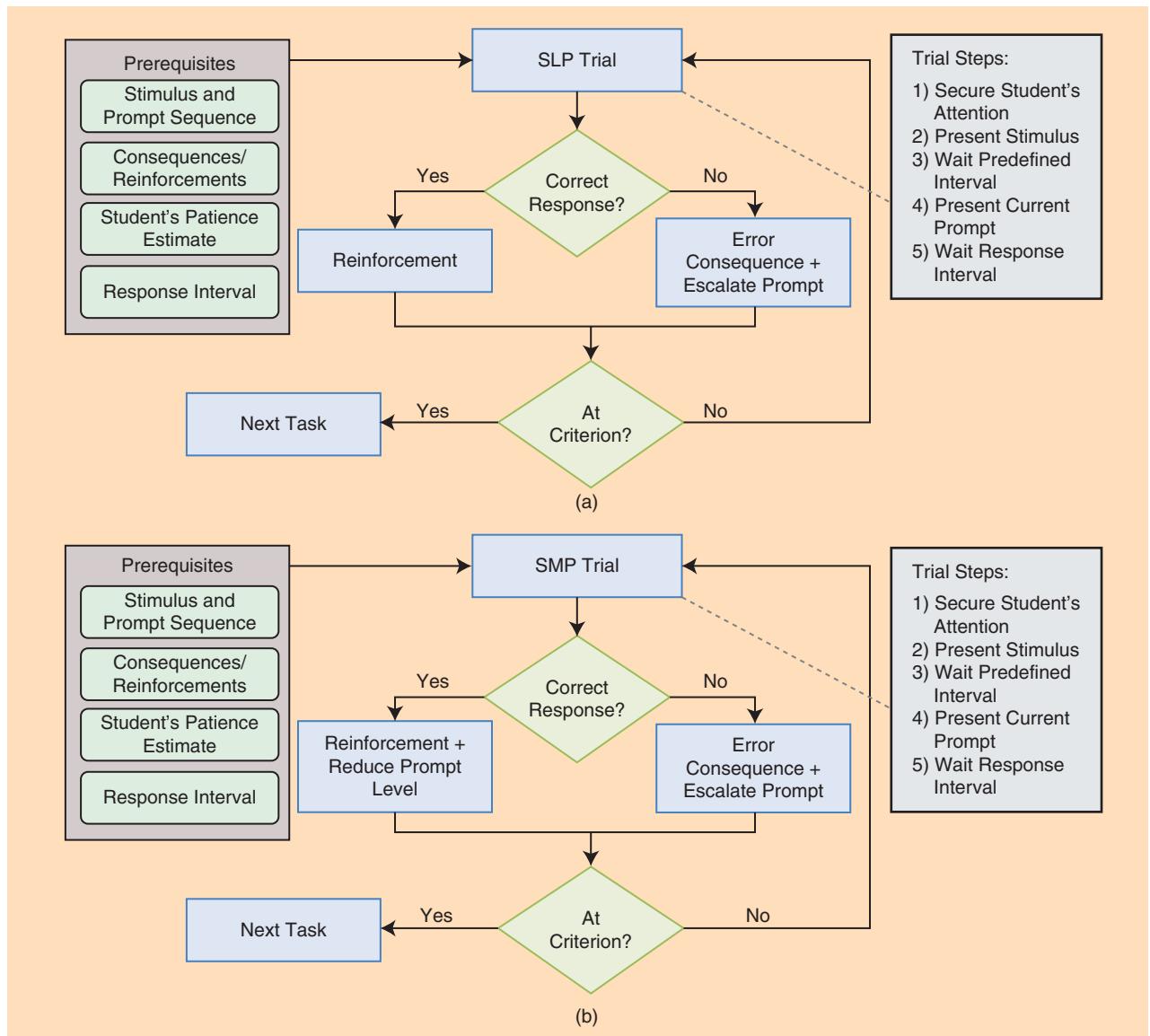
### Response-Prompting Types

The System of Least Prompts (SLP), System of Most Prompts (SMP), and Constant Time Delay (CTD) [17] are popular and well-validated response-prompting strategies for instruction of

chained and discrete tasks. In addition, they have been successfully used to teach students with I/DDs (Figures 4 and 5). Furthermore, these methods are applicable to a wide range of students and have proved to be effective for teaching a wide variety of skills. For this article, we computationally encoded the SLP, SMP, and CTD strategies for use in an IRI system's decision-making process. We also demonstrate generalizability over an AR system.

### SLP

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



**Figure 4.** Flowcharts for (a) SLP and (b) SMP response-prompting strategies, adapted from [17]. Note that the key difference between SLP and SMP is in the prompt sequence input: SLP is least-to-most intrusive (see the "SLP" section), and SMP is most-to-least intrusive (see the "SMP" section). CTD has only one prompt, the controlling prompt, and the time interval is varied (see the "CTD" section).

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, i.e., the 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.

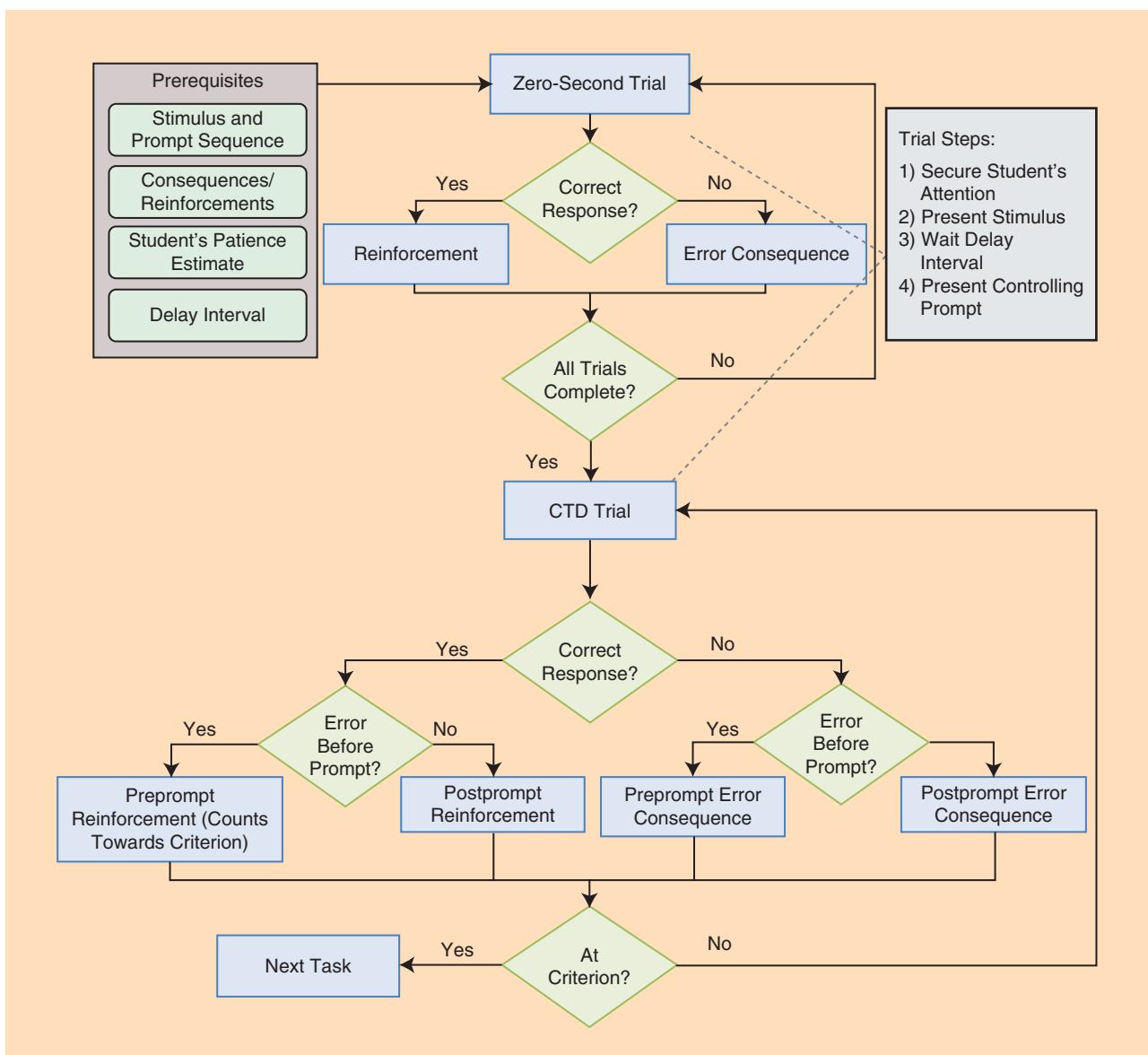
The goal of SLP is that, after sufficient repetitions, students require fewer prompts and eventually respond correctly before any prompt is delivered. As the student answers correctly at lower prompt levels, a process of self-fading occurs, where the student's answers determine the rate at which the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

## SMP

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 underlying idea for the SMP approach is to guarantee that the student first makes a successful response (via the controlling prompt) and then to fade the intrusiveness of the prompt and work toward fully independent behavior. One observed difference between SMP 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



**Figure 5.** A flowchart CTD response-prompts strategy, adapted from [17]. Note that CTD has only one prompt, the controlling prompt, and the time interval is varied (see the “CTD” section).

prompts are arranged from most to least intrusive, errors may be less frequent.

## CTD

In CTD, prompts are delivered after a time delay following a task direction, i.e., 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, known as the *prompt delay interval*, is varied. Initially, the delay between the task direction and the controlling prompt is zero, in what is termed the *zero-second delay trial*. The prompt delay interval is constant for a set of instruction trials until the criterion is met; it is then systematically increased. The goal of CTD is a consistently correct response before the prompt.

## ***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 for which 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 within 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, and so on. Because of their sequential nature, chained tasks can be taught from the beginning of the sequence in what is known as *forward chaining* or, by iterating from the end of the sequence, *backward chaining*. Tradeoffs exist for both. This work uses forward and backward chaining as appropriate to the task and examines how they impact successful learning from intelligent systems.

## ***Creation of Prompting Hierarchies***

The process of creating a prompt hierarchy involves devising a series of prompts and arranging them in a hierarchical structure 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 the 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, our interdisciplinary team of authors collaborated closely to ensure a successful and productive experience for the students involved.

## **Experiments**

### ***Single-Case Experimental Design***

Experiments were developed for this research using a single-case experimental design (SCED), a common design method

in special education research. Rather than comparing groups or subjects, in SCED participants serve as their own control, with performance in at least two experimental phases being compared [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/DDs is infeasible but also because, even if it were possible, the participants' capabilities and diagnoses could be so diverse that drawing even a coarse statistical inference would be challenging. The objective of the SCED is to determine whether a causal or functional relationship exists between the delivery of the independent variable (the intelligent instruction system) and significant increases in the dependent variable (the acquisition and maintenance of the skills taught).

Two types of SCED designs were used for this research: a combined multiple baseline across participants and a combined multiple baseline across skills. These designs allow for the 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 using a multiple baseline across skills, skills are taught one at a time, and instruction is introduced for each skill sequentially after the previous one has been learned. In experiments using a 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 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 that any observed change occurs due to extraneous factors (e.g., practice or history effects) is eliminated. This allows for experimental control and establishing a causal relationship [21].

## ***Student Participant Population***

All studies for this research were performed in accordance with institutional review board protocols and approval. The students who participated in these studies were college age and attendees of a postsecondary education (PSE) program designed for young adults with I/DDs at the University of Tennessee, Knoxville, named FUTURE (more information on the FUTURE program is available at <http://futureut.utk.edu>). All students 1) were ages 18–34 with an intelligence quotient between 57 and 67, 2) received special education services throughout school, and 3) earned modified high school diplomas before participating in the FUTURE program.

To meet the three replication minimum requirements for evaluating the intervention effects, three students per experiment were taught until our criteria were reached, i.e., they mastered the skill being taught. Seven students participated overall, four of whom participated in multiple experiments.

## 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 Figure 6. The IRI autonomously performs the instruction, prompting, observation, evaluation, and feedback (correction or reinforcement) loop shown in Figure 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 the Robot Operating System (ROS; <http://ros.org>) for messaging, inter-process communication, and common robotics libraries.

The custom object-tracking system shown in Figure 7 provided the IRI with the ability to observe objects with which both the student and robot interacted and thus accurately interpret the student's performance. The object-tracking system was implemented as a set of custom ROS nodes using OpenCV to process live image streams from a camera mounted under a transparent tabletop, as illustrated in Figure 8. Details of the vision system performance are reported in [15], where we previously introduced response prompting for instruction with an IRI. Interaction and feedback are provided through synthesized speech, speech recognition, and gestures.

The robotic hardware for this research is a Meka Robotics M3 mobile humanoid robot [Figure 8(a)] with 7 degrees-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 (version 1.09) camera, one universal serial bus camera, a Bluetooth microphone, and stereo speakers. The robot is equipped with two personal computers—one providing real-time functionality of the base, arms, hands, and lift and the second 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 in the following)
- 2) *geometric assembly of larger objects from smaller pieces*, as shown in Figure 9, for which prerequisite skills were taught using AR instruction (see the “Intelligent AR Instruction Experiments” section).

Both skills involved live interaction with objects and demonstration of correct responses.

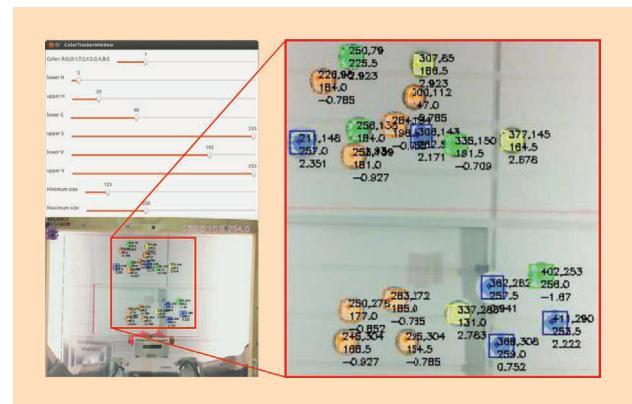
Making change is a life and vocational skill that involves calculating the correct quantities and denominations of currency to be exchanged after a cash transaction. The assembly skills taught in our geometric assembly experiment are useful in job settings,

and the geometric reasoning aspects are useful in both work and personal life.

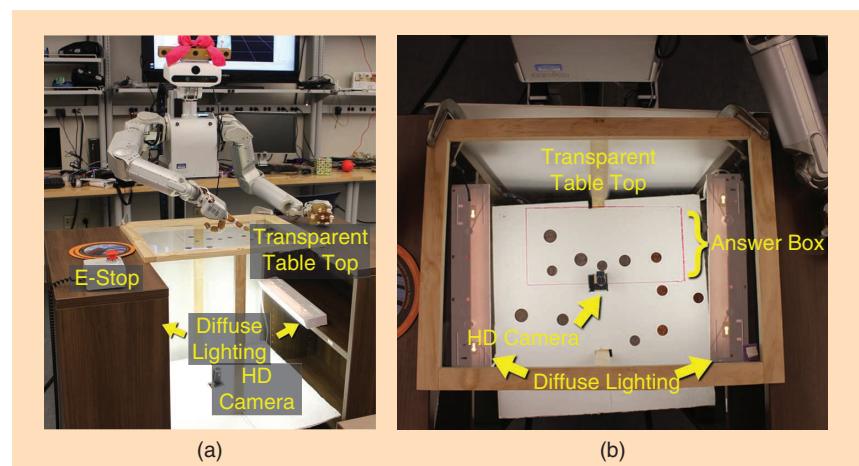
Instruction strategies appropriate to the skill being taught were selected. For the geometric assembly skill instruction, an SMP prompting strategy was used with backward chaining, along with a multiple-baseline-across-skills experimental



**Figure 6.** The interaction setting for instruction.

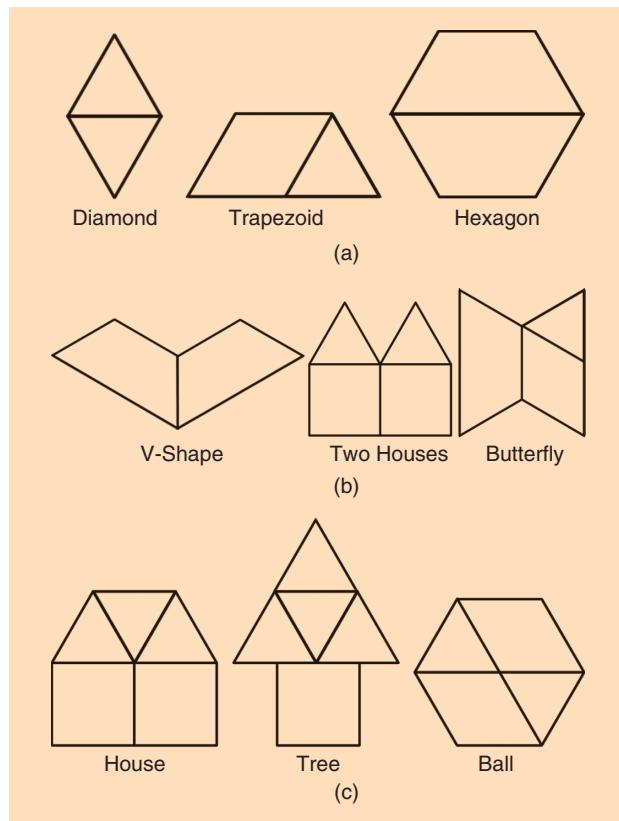


**Figure 7.** The object tracker graphical user interface, with live, adjustable parameters (upper left) and the annotated live image (lower left). The magnified image on the right shows an enlarged view of the annotated image. The annotations include position, orientation, size, centroid location, and bounding box for each object.



**Figure 8.** The table setup for the instructional setting from the (a) student's view and (b) overhead. HD: high-definition.

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 used a



**Figure 9.** The puzzle tasks for the geometric assembly skills: (a) subcomposition, (b) symmetry, and (c) complex assembly.

multiple-baseline-across-participants methodology to control for any outside influences affecting the participants' ability to make change.

An example prompt hierarchy for part of the making change task is shown in Table 1. The SLP strategy was used, with prompts 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 as follows: no response (NR), partially correct (PC), correct, and incorrect (I).

In addition, to evaluate the student volunteers' attitudes about learning from a robot, Likert-type scale statements and open-ended questions were used to collect subjective data before and after participants interacted with the IRI for the making change experiment. The results of this survey are discussed in the "Subjective Acceptability" section.

### Intelligent AR Instruction Experiments

We also implemented response prompting for intelligent instruction on a portable AR device. AR devices share similar features with IRI instruction, such as context awareness, precision, and tirelessness when situated in an environment with the user. Both are capable of delivering to students visual and auditory prompts 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 versus a humanoid robot).

Figure 10 shows an overview of the AR system. When wearing an AR device and learning a new sequential task, the student 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

**Table 1. The prompt hierarchy for the making change task using SLP.**

Prompt Level	Response	Prompt Description
Verbal cue 1	NR	Verbal interaction to determine how much change is due
	PC	Verbal encouragement, verbally providing goal
	C	Differential positive reinforcement
	I	Same as NR
Verbal cue 2	N	Verbal interaction to determine which coin to begin with
	PC	Verbal encouragement, verbally providing goal and shortage between current state and goal
	C	Differential positive reinforcement
	I	Verbal encouragement, providing goal and excess between current state and goal
Direction 1	N	Gesture to correct first coin, verbally providing goal
	PC	Gesture to correct next coin, verbally providing goal and shortage
	C	Differential positive reinforcement
	I	Gesture to coin to remove, verbally providing excess
Direction 2	N	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

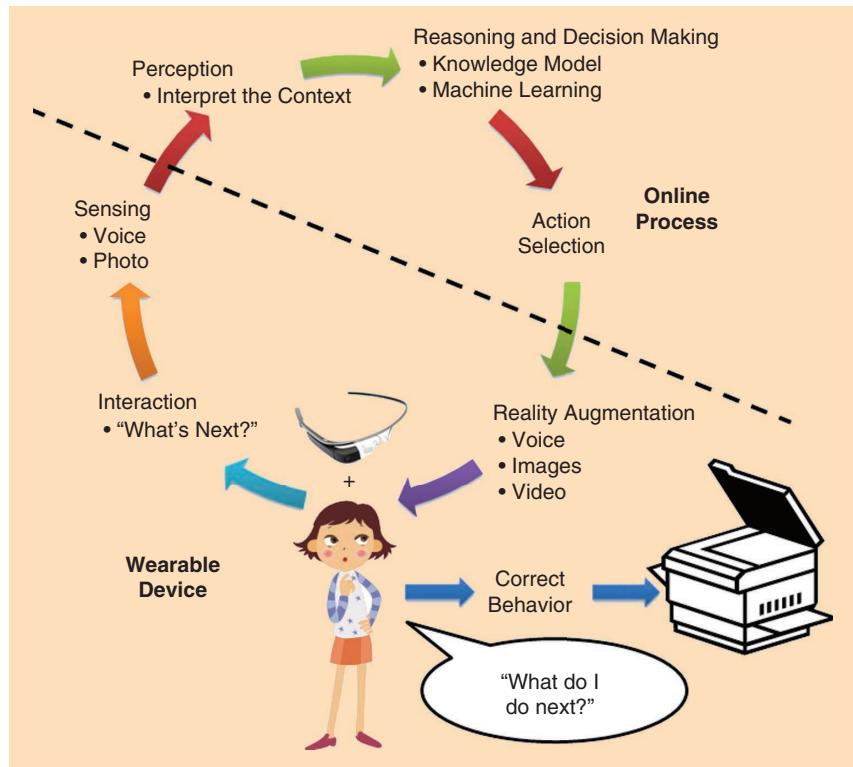
experiment, the prompt modalities are an augmented version of the uploaded image as well as the audio. Live video augmentation is outside the performance abilities of our hardware implementation but could be included in future work.

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 are provided within 5–10 s via the Glass display and built-in speaker. On the cloud side, intelligent instruction is made possible using OpenCV for image processing and support vector machines 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, and 3) performing geometric reasoning tasks with puzzle blocks to acquire the prerequisite ability for the advanced skills taught by the IRI system. Figure 11 shows example annotated image prompts from these experiments. A summary of these experiments appeared in [16].

For the copy machine skill (an office vocational work skill), 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. For the student



**Figure 10.** The AR system overview. The student is instructed to face the device and ask for assistance. A picture taken from the wearable camera is classified, and an augmented image and audio containing the proper instructional prompt for the next step are delivered.

account statement skill (an employment and independent living skill), the student was asked to log in and retrieve a copy of his or her prepaid student account statement. The third experiment, geometric reasoning, taught object manipulation, placement, orientation, and assembly subskills that are highly vocationally relevant.

For these experiments, instructional strategies were carefully designed using the appropriate combination of prompts, prompting strategies, and prompt chaining for each skill taught and the AR technology method (Table 2). 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, in which each subskill or step was mastered before the successive one was introduced. Also, in all three, the skills were taught in forward-chained order because the

**Table 2. The experiment overview.**

Platform	Experiment	Prompt	Design	Relationship
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	Backward chaining

tasks (e.g., navigating through a copier interface) can only progress in a forward direction. The geometric reasoning skill also used discrete subskills at some steps in the chain. For the prompting strategy, the geometric reasoning skill used CTD. The other two AR experiments used a simplified self-directed method that provided controlling prompts only when the students requested assistance from the AR device. This was similar to CTD but with the student controlling the time delay.

## Results and Discussion

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

Results from the IRI-instructed making change skill are shown in Figures 12 and 13. These results demonstrate 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 forward chaining, the IRI was able to provide increasingly supportive prompts to help all students to mastery, as discussed in the “SLP and Forward Chaining” section.

The results from the IRI-instructed geometric assembly skill and the AR-instructed skills are shown in Figures 13 and 14. Performance in the geometric assembly [Figures 13(a) and 14(a)] shows strong success that can be largely attributed to the structure of task instruction using SMP with backward chaining, as discussed in the “SMP and Backward Chaining” section.

The results from the AR instruction experiments [Figures 13(b) and 14(b)] verify the generalizability of the proposed approach to other interaction domains. Indeed, as discussed in the “Efficiency of Self-Directed AR Instruction” section, students were able to make rapid gains with AR instruction, which we attributed to the self-directed prompt control combined with the immediacy of intelligent, context-aware instruction in AR. The results presented here are a significant expansion of preliminary work [15], [16] that introduced the IRI concept.

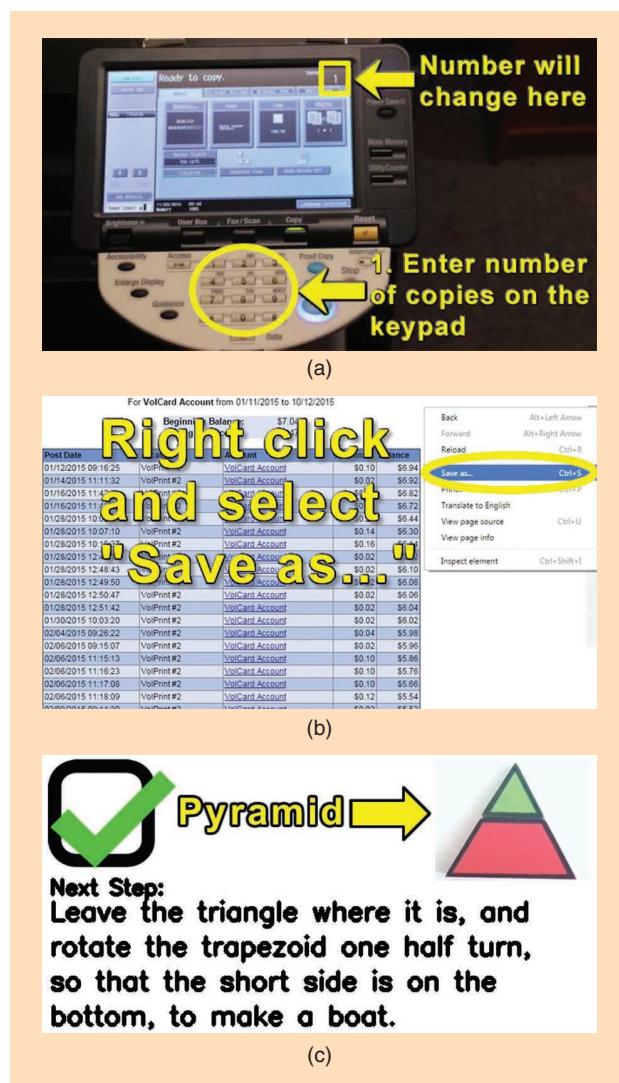
To enable a comparison across the diverse skills and experimental configurations, Table 3 summarizes the total number of errors students made before mastering the corresponding skill or subskills in the making change and geometric assembly experiments (Figure 9). The values in Table 3 represent the total number of errors each student made across all trials for each skill.

Although 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 the copy machine skill, and higher rates occurred with the highly complex making change skill, which was divided into subskills for correct use of a calculator and correct exchange of coins.

In addition, whereas the outcomes were very positive for those students involved, careful decisions were made to combine the best response-prompting strategies, experimental design, and prompt relationship with each skill being taught to achieve the optimal performance, as summarized in Table 2. 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, which we hope can constructively inform future work in designing intelligent robotic and AR instruction technologies.

### SLP and Forward 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



**Figure 11.** The example annotated image prompts from the AR instruction experiments: (a) a step from the copy machine study; (b) a step from the student account statement study; and (c) a step in the geometric reasoning study, where a student is learning the rotate and half-turn subskills. The image prompts are shown to the user via the AR device and accompanied by audio prompts.

clear prompt. We found that, despite this challenge, the strength of the methodology allows the robot to provide increasingly intrusive prompts that can be designed to progressively limit the response space. Then, by concluding with the controlling prompt, which allows for only the correct response, success is guaranteed.

### SMP and Backward Chaining

Importantly, we found that backward 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 and forward-chaining method described previously) and because of a robot's advantages in performing repetitive tasks. Backward chaining and most-to-least-intrusive prompting means that the skill is taught backward 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 having 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 fewer piece in the puzzle until the student was finally asked to assemble the puzzle him or herself from scratch.

It is worth noting that, in the results shown in the last row of Table 3, all errors that occurred in the complex assembly subskill involved assembling the “ball” hexagon-shaped object [Figure 9(c)], which has many possible hexagon-shaped but incorrect sub-compositions; students had no difficulty assembling the other figures in this subskill.

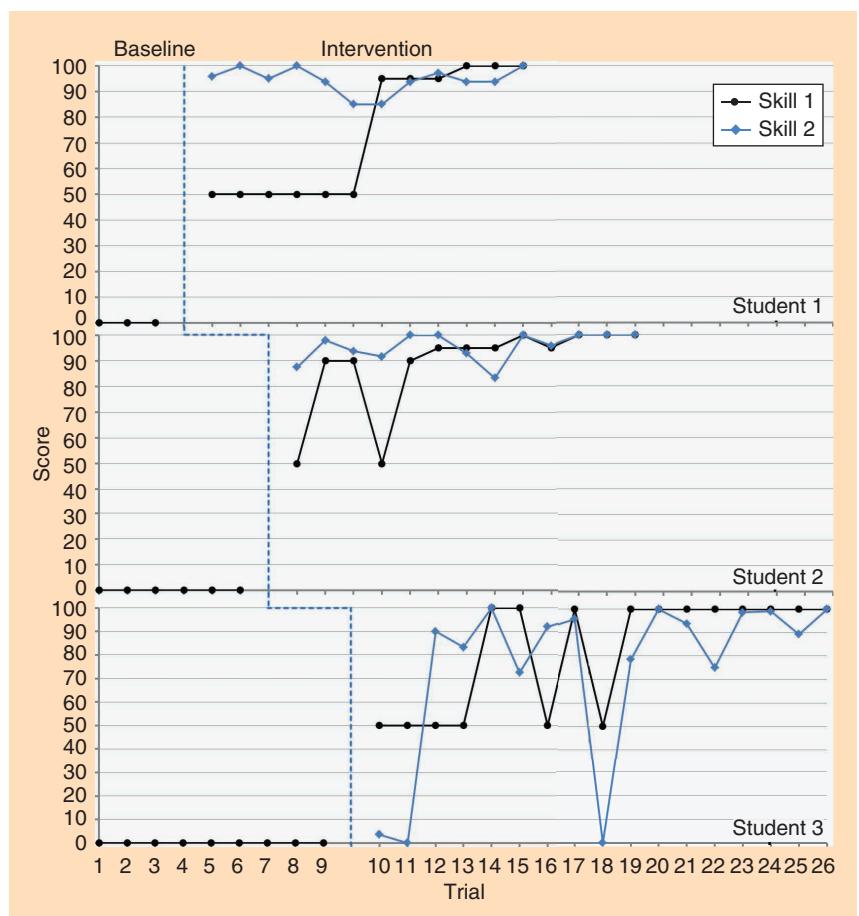
In light of these observations, we believe that the SMP and backward-chaining approach, while very time-consuming and perhaps tedious and more mistake-prone for a human instructor, is a perfect match for the tirelessness and precision of an IRI. Additionally, this high degree of repetition can greatly benefit students with I/DDs.

### Subjective Acceptability

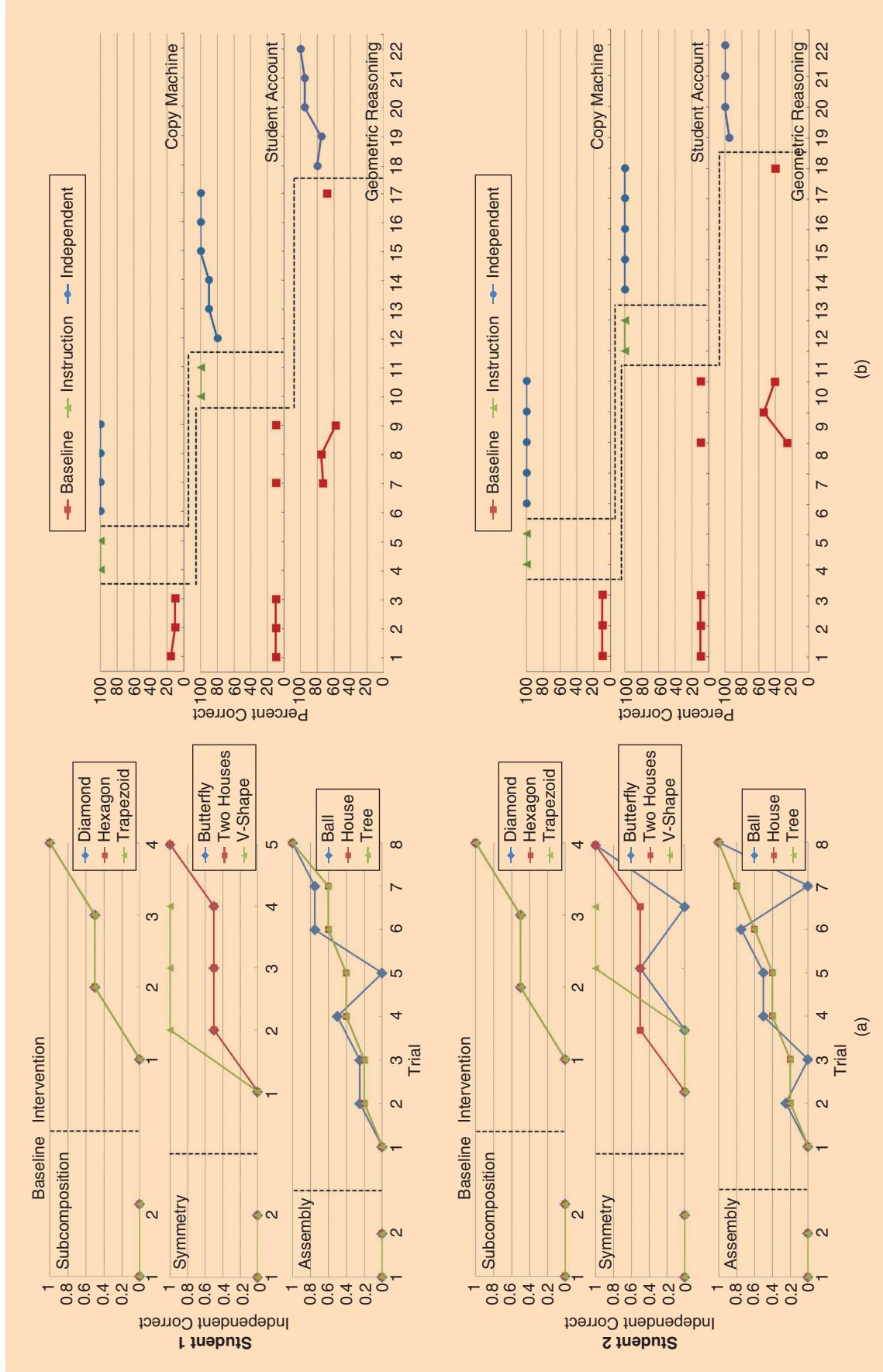
To determine the students' opinions on being instructed by a robot, an acceptability study using a Likert-type scale was conducted before and after the students' successful instruction with the IRI as part of the making change experiment.

A five-point Likert-type scale was used for each statement, and optional open-ended follow-up prompts appropriate to each statement (e.g., “Why or why not? or “Please explain”) were posed. To ensure a uniform understanding of the questions, surveys were performed orally by experts; visual aids were provided for response anchor points, with responses from “strongly disagree” scored at -2 to “strongly agree” scored at +2. The preassessment survey consists of 18 statements divided into eight categories. The postassessment survey consists of 30 statements in 14 categories. There are two to three statements in each category, and a summative analysis was applied. This study found mixed enthusiasm before instruction by an IRI and positive opinions of the overall experience and performance of the IRI after instruction.

In Table 4, we see categories from the preinstruction and postinstruction 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, postinstruction results show a positive opinion of the overall experience and performance of the robot. Compared with the students' lower levels of willingness to work



**Figure 12.** The results from the IRI-instructed making change experiment using a multiple baseline across participants. Data are 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 subskills: 1) identifying the correct amount of change and 2) providing the correct combination of coins. The scoring methodology is detailed in [15].



**Figure 13.** Students 1 and 2: (a) The IRI geometric assembly skill results; here, three puzzles were taught, with three puzzles per skill (see Figure 9). Nonincreasing performance generally represents errors, because, when using the SMP and backward-chaining method, the number of possible correct independent steps increases until the student is asked to assemble the entire figure independently. (b) The 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 they control their own prompts.

with a robot before 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 previously.

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

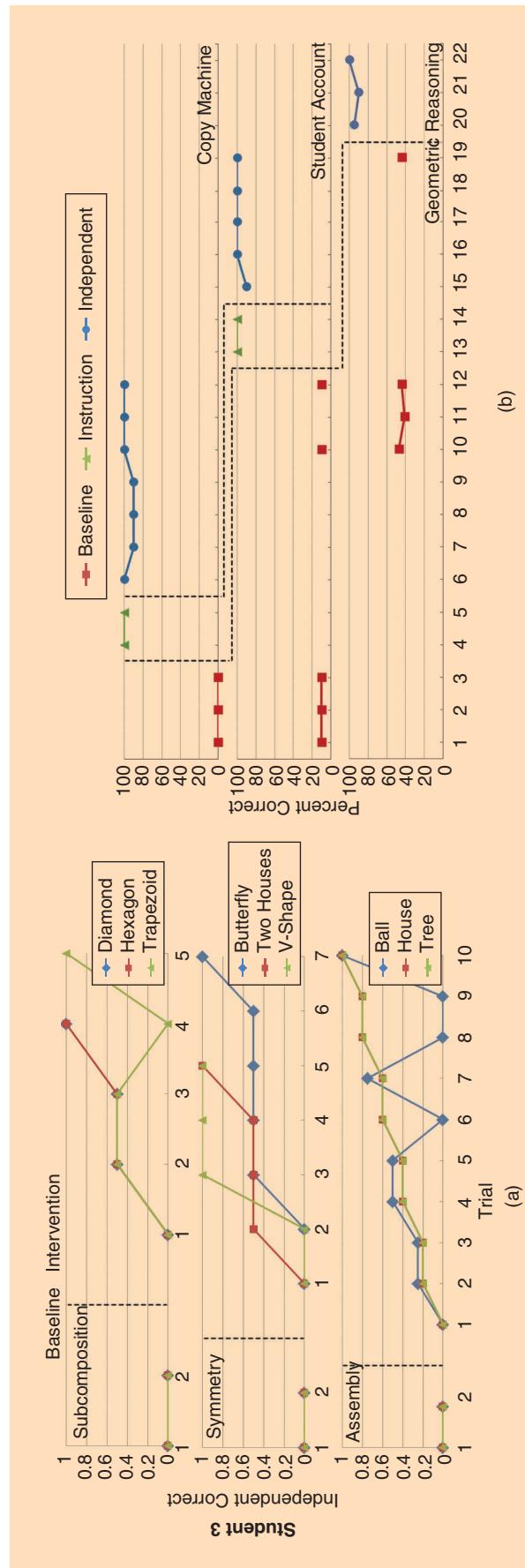
### **Efficiency of Self-Directed AR Instruction**

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.

### **Challenge Areas**

Based on 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 a 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



**Figure 14.** Student 3: (a) The IRI geometric assembly skill results; here, three skills were taught, with three puzzles per skill (see Figure 9). Nonincreasing performance generally represents errors, because when using the SMP and backward-chaining method, the number of possible correct independent steps increases until the student is asked to assemble the entire figure independently. (b) The AR 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 they control their own prompts.

**Table 3. The 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	9	9	6
Calculator	9	6	14
Coin response	0	0	2
Geometric assembly	0	3	4
Subcomposition	2	4	11
Symmetry			
Complex assembly			

**Table 4. The acceptance survey summative results.**

	Student 1	Student 2	Student 3
<b>Category (Preinstruction)</b>			
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
<b>Category (Postinstruction)</b>			
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

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 encourage interact primarily 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 he or she is idle, either due to uncertainty about the correct response or to completion of the response, is a trivial task

for a human teacher. However, this necessary step can be difficult for an IRI; failure in this area can result in loss of trust in the IRI's abilities (reported in an open-ended survey response by one student with whom our IRI initially had difficulty engaging) as well as missing an instruction.

### 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, and verbiage) to create an IRI.

Regarding the first, the population size for this article is not statistical. As noted in the "Experiments" section, 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-sized study is infeasible, additional replications would help strengthen understanding of this relationship. Furthermore, the variability of the population of persons with I/DDs is large. Consequently, future studies should be performed with the understanding that this variability could impact the rate of successful learning. The PSE program (see the "Student Participant Population" section) from which this study's drew participants is exposed to technology as part of the program; this exposure could have influenced performance.

Regarding the design decisions, as noted in the "Challenge Areas" section, an IRI is a complex system. The researchers learned early on that even small changes, e.g., in verbiage, can significantly impact interaction. This is par-

ticularly true for the population studied. Our work sought to achieve a good-performing design through collaboration with experts in the education field. In addition to being based on such collaborative work, future studies should seek to define and categorize the types of design decisions that could affect study outcomes.

### Conclusion

In this article, we described the use of response prompting for cognitive decision making in intelligent instruction, with applications in robotics and AR. We gave an overview of our

interdisciplinary collaborative efforts in applying this approach to the instruction of students with I/DDs 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. Although these results are a preliminary attempt to begin addressing some of the issues facing those with I/DDs, we believe they show that this novel approach has merit in many applications. Most importantly, we believe the findings herein should prove constructive toward future investigations and the development of intelligent instructive technologies, especially those that provide critical assistance to this important population.

## Acknowledgments

We acknowledge the wonderful students, staff, and faculty of the FUTURE PSE program and the College of Education at the University of Tennessee, particularly Prof. Mari Beth Coleman and Prof. David Cihak for 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 *Proc. Tenth Annu. ACM/IEEE Int. Conf. Human-Robot Interaction*, 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," *J. Educ. 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," *J. STEM Teacher Educ.*, vol. 48, no. 1, 2011.
- [4] R. McKeown, "Teacher education 1992 and 2012 reflecting on 20 years," *J. Educ. Sustainable Dev.*, vol. 6, no. 1, pp. 37–41, Apr. 2012.
- [5] D. Leyzberg, S. Spaulding, M. Toneva, and B. Scassellati, "The physical presence of a robot tutor increases cognitive learning gains," in *Proc. Annu. Meeting of the Cognitive Science Society*, 2012, pp. 1882–1887.
- [6] A. Tapus, C. Tapus, and M. Mataric, "The role of physical embodiment of a therapist robot for individuals with cognitive impairments," in *Proc. IEEE Int. Symp. Robot and Human Interactive Communication (RO-MAN)*, 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 *Proc. IEEE Int. Symp. Robot and Human Interactive Communication (RO-MAN)*, 2008, pp. 701–706.
- [8] M. Grigal and D. Hart, *Think College! Postsecondary Education Options for Students with Intellectual Disabilities*. Baltimore, MD: Brookes Publishing Company, 2010.
- [9] A. Meghdari, M. Alemi, M. Ghazisaeid, A. Taheri, A. Karimian, and M. Zandvakili, "Applying robots as teaching assistant in EFL classes at Iranian middle-schools," in *Proc. Int. Conf. Education and Modern Educational Technologies (EMET-2013)*, Venice, Italy, 2013, pp. 67–73.
- [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 *Proc. SIGCHI Conf. Human Factors in Computing Systems*, 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 Robot. Automat. Mag.*, vol. 23, no. 2, pp. 74–80, June 2016.
- [12] M. Fridin, "Storytelling by a kindergarten social assistive robot: A tool for constructive learning in preschool education," *Comput. Educ.*, 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 *Proc. Human Factors and Ergonomics Society Annu. Meeting*, 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 *Proc. Companion of the 2017 ACM/IEEE Int. Conf. Human-Robot Interaction*, 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 *Proc. IEEE Int. Symp. Robot and Human Interactive Communication (RO-MAN)*, 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 *Proc. Int. Florida Artificial Intelligence Research Society Conf. (FLAIRS)*, 2016, pp. 505–509.
- [17] M. Wolery, M. Ault, and P. Doyle, *Teaching Students with Moderate to Severe Disabilities: Use of Response Prompting Strategies*. White Plains, NY: Longman, 1992.
- [18] D. Isla, R. Burke, M. Downie, and B. Blumberg, "A layered brain architecture for synthetic creatures," in *Proc. Int. Joint Conf. Artificial Intelligence*, 2001, pp. 1051–1058.
- [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 Int. Symp. Robot and Human Interactive Communication (RO-MAN)*, 2014, pp. 561–566.
- [20] D. L. Gast and A. D. Spriggs, "Visual analysis of graphic data," *Single Subject Research Methodology in Behavioral Sciences*, D. L. Gast, Ed. New York: Routledge, pp. 199–233.
- [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, Jan. 2005.

**Christopher Reardon**, U.S. Army Research Laboratory, Adelphi, Maryland. E-mail: christopher.m.reardon3.civ@mail.mil.

**Hao Zhang**, Department of Computer Science, Colorado School of Mines, Golden. E-mail: hzhang@mines.edu.

**Rachel Wright**, Common Threads Resource Center, Madison, Wisconsin. E-mail: rachelw@commonthreadsmadison.org.

**Lynne E. Parker**, Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville. E-mail: leparker@utk.edu.

