Exploring Task-Level Contingent Mediations for Vocabulary Instruction across Robot, Virtual, and Human Teachers

Wing-Yue Geoffrey Louie[1], *IEEE Member*, Tanya Christ[2], Pourya Shahverdi[1], Katelyn Rousso[1], Evan Dallas[1]
Alexander Tyshka[1], Amanda Wowra[2], Kendra Barnett[2], Iman Bakhoda[2]

***Abstract*— Social robots are being introduced in a variety of educational domains with great success. These social robots are often designed by drawing inspiration from practices held by human teachers. Contingent mediations are a prime example of a high-quality teaching practice that supports better outcomes in human-human teaching. This can inform the design of social robots. Current research on robot use in education has focused on curriculum-level contingent mediations where the difficulty level of subsequent tasks are adjusted to the current capabilities of a learner. However, task-level contingent mediations that provide support for students' learning during a specific task remain unexplored. This research investigates whether patterns of task-level mediations differ between a robot, virtual, and human agent, as well as their effects on learning outcomes. To investigate these research questions, we designed instruction that utilizes contingent mediational flows, based on formative assessment data, to be delivered by a human, robot, and virtual agent to teach grade 3-5 children science words. We identified 23 unique instructional patterns. Then, we compared these patterns across teaching agents and their effects on learning outcomes. Overall, our study demonstrated that high-quality contingent mediations support children's science vocabulary in learning regardless of the teaching agent.**

I. Introduction

Social robots provide cognitive and affective benefits in educational contexts ranging from language to math [1]. These benefits are often attributed to the embodiment of social robots through studies comparing their effects on learning gains with virtual and disembodied agents [2]. However, studies comparing robots to human teachers are scarce [1]. To understand how robots can fit into the educational ecosystem, we must understand how robots compare to human teachers so we can effectively design robots for education and identify the role they can play within an instructional toolbox.

A fundamental skill in high-quality human teaching is the ability to provide appropriate contingent mediation to a learner [3], [4]. Contingent mediation is an approach stemming from Vygotskian theory that disposes of the theory that a learner is an empty vessel to passively receive information and suggests that a learner's higher mental processes are formed by the presence of mediating agents during a learner's interaction with their environment [5]. Contingent mediation are the actions of an agent that guide/constrain a learner's interactions with their learning environment and applied depending on a learner's actions as well as learning progress [5]. Contingent mediations include modifying task difficulty, modeling, contingency management (e.g., praise), feedback, and scaffolding [6]. Literature with human teacher-learner interactions show that mediations provide support for improving learning outcomes and learning results from contingent responses are even more superior [3].

There are two types of contingent mediations explored in educational robotics [7], [8]. The more frequently studied type of contingent mediation is curriculum-level mediation [7]. These mediations adjust the difficulty of the subsequent tasks, such as adjusting the difficulty of subsequent stories presented to preschoolers, based on their learning performance for previous words [7]. Another critical contingent mediation is task-level mediation, which occurs during the learning of a specific task (e.g., while learning a particular word). Task-level mediation refers to adapting instruction through targeted information and content that addresses gaps discovered in a learner's knowledge based on formative assessment. This provides individualized pathways for a student to learn more complex tasks. While limited research on robot use in education has explored task-level mediations, such as having robots use gesture or gaze direction to elicit children's attention to pictures that depict a particular word's meaning [8], we found no research that explores contingent task-level mediations.

This research investigates what patterns of task-level mediations occur when third- to fifth-grade children interact with human, robot, and virtual teaching agents to learn science words with two morphemic parts. All agents utilized task-level contingent meditations for children's learning of each word. A mixed methods analysis was then conducted to investigate the following research questions (RQ):

RQ1: What patterns of contingent mediations occur?

RQ2: Do patterns of contingent mediation differ across teacher types (human, virtual, robot) for different words? If so, how?

RQ3: Do patterns of contingent mediation differ across teacher types for different age students? If so, how?

RQ4: Do numbers of contingent mediations differ across teacher types? If so, how?

RQ5: When utilizing the same contingent mediations, do learning outcomes differ amongst teacher types?

This work was supported by the National Science Foundation grant #2238088

[1]Intelligent Robotics Laboratory, Oakland University, MI, USA (e-mail: louie@oakland.edu)

[2] Department of Teaching and Learning, Oakland University, MI, USA

II. Literature Review

Two areas of research are most relevant to our study. First, we present a review of studies that have compared children's learning outcomes across teaching agents. While research also explores differences in children's engagement and preferences across types of teaching agents, we do not review those studies here as they are beyond our immediate scope, which focuses on learning outcomes. Second, we present the small available body of research on task-level mediations.

A. Comparisons of Children's Learning Outcomes across Human, Robot, and Virtual Teaching Agents

1) Robots vs. Virtual Teachers: When studies have compared students' learning outcomes across instruction by robots versus tablets, findings have been mixed. Zhexenova and colleagues found that primary students' handwriting of Latin-based alphabet letters is similarly well supported by robots and virtual teachers via tablets [9]. However, Hyun and colleagues showed that four-year-old children had greater word recognition and vocabulary gains when taught by a robot, as compared with a virtual teacher via a digital notebook, likely due to the robot engaging in more bi-directional interactions with children [10]. Likewise, Konijn and colleagues compared second language learning outcomes for young children across storytelling activities with a robot versus a virtual agent via tablet and found that the children had greater learning gains when working with the robot [11]. Additionally, Hsiao and colleagues compared preschoolers' literacy outcomes (e.g., comprehension and retelling) across instruction provided by a robot versus a virtual agent and showed that children's outcomes were better when they were taught by the robot. They attributed this to the additional features of the robot, such as light and sound features and the robot's ability to move its arms, which seemed to draw children's attention to the instruction better than the virtual teacher that did not have these features [12].

In contrast, Tozadore compared Portuguese students' English word learning outcomes when taught by a tablet versus a robot teacher and found that students with the tablet teacher had better English word learning outcomes, paid more attention, and self-reported better learning than the students taught by the robot [13]. Likely this was related to the value of tangible features, such as subtitles on tablets, which were not available in the robot's condition.

We conclude that the differences in findings across these studies seem to be related to differences in affordances across the teaching agents. When the affordances are similar [9], then teaching agents have equivalent effectiveness. However, when one teacher agent has more helpful affordances (e.g., bi-directional interactions with robots or subtitles for virtual agents on tablets) then that agent is more effective [9].

2) Robots vs. Human Teachers: We found just two previous studies that compared students' learning outcomes across instruction by robots versus humans. Westlund and colleagues examined how preschool children learned new words from a robot versus a human [8]. They found that children used non-verbal cues, such as gaze direction and bodily orientation, to learn equally well from both robot and humans. Likewise, Westlund and colleagues compared preschoolers' word learning across robot, human, and virtual teachers and found that children learned words similarly across teacher types [7]. Again, it seems apparent that when the mediational affordances across agents are similar, both agents will perform teaching equally as well and yield similar student outcomes.

3) Task-level Mediations: Task-level mediation provides support for students' learning during a specific task. De Haas and colleagues demonstrated that task-level mediations (i.e., providing extra attempts to learn animal names with varied feedback), was more effective for supporting five year-olds' word learning from robots than instruction from robots without task-level mediations (i.e., just one response attempt with automatic feedback regarding whether their response was correct) [14]. Despite this, we found only one other study that employed task-level mediations in robot instruction. Robots in Westlund and colleagues' studies used task-level mediations, such as oral naming of pictures, eye gaze, and bodily orientation to draw preschool students' attention toward pictures that depicted the concept, as supports for children's learning of a particular word. Thus, further research is needed on robots' and virtual agents' use of task-level mediations to support students' learning outcomes [7], [8].

Further, the task-level mediations explored so far in this body of research have not included contingent mediations, which use formative assessment data to inform the immediate next mediation that will be used to support the child's learning. However, the importance of exploring contingent mediations has been raised by several researchers [1], [15], [16]. Thus, this is another important next direction for research to explore.

III. Methods

An Institutional Review Board approved mixed methods study was conducted with children in grades 3-5 to investigate the effects of task-level meditations on human-robot teaching and how they differ from other social agents. A repeated measures design was utilized where children interacted with three social teaching agents: a human, a robot, and a virtual agent. All teachers employed the same teaching materials, instructional content, and flow of formative-assessment-based mediations to teach science words with two morphemic parts.

A. Participants

Twenty grade 3-5 children were recruited from a university-based STEM camp. All participants spoke English fluently. Written informed consent was obtained from all participants' parental guardians prior to the study and parental guardians were informed that their children could withdraw at any time. Parents were seated out of view of the child and did not participate in the sessions. Before beginning the sessions, a researcher explained what would take place during the session, asked for participants' assent, and told participants they were free to stop participating at any time.

B. Setting

Sessions took place in a university classroom. Room partitions were used to divide the classroom into an 8'

(a)

(b)

Fig. 1. (a) A participant interacting with the robot teacher. (b) The virtual agent, Ari.

x 16' space, Figures 1a. The space was decorated by an experienced teacher with science-related objects and posters to replicate the ambiance of being in a classroom environment. Participants were seated at a desk in front of a 27" touchscreen monitor with speakers placed on either. A webcam was placed on top of the monitor and facing the participant. In the robot and human teacher conditions, the teacher was next to the monitor so that they could refer to the teaching materials on the screen. In the virtual agent condition, the teacher was depicted on the screen. A camp counselor or parent, familiar with the child, always remained in the room for the security of the participants. Behind the partitions was a computer used by a researcher to ensure the reliability of the interactional flow.

C. Pedagogical Design

1) Vocabulary Word Selection: First, we chose nine science words that aligned with Beck and her colleagues' Tier 3 words, which are specific to a particular discipline's content (e.g., biodiverse is a word within the discipline of biology) [17].

Second, each word we chose also had two morphemes. Both morphemes in the chosen words were either bases or roots, but not affixes. This was a way to control for word-meaning learning difficulty because bases and roots are more similar in quality as compared with affixes. We also controlled for word difficulty by selecting words with morphemes that could be represented concretely with visual examples. No words were chosen that shared any of their morphemes with other words (e.g., we used "biodiverse", so we would not include "biosphere" because they both contain the root "bio"). This avoided confounding the learning conditions and results.

Third, we used the Corpus of Contemporary American English (COCA[1]) to identify the frequency of word occurrence (over a billion words of text), number of texts that the word occurs (out of 485,202 total texts), and percentage of all texts that have the word. The nine words, their morphemes, and COCA frequency data are presented in Table I. All words have low occurrences to decrease the likelihood that children already know their meanings.

[1]https://www.english-corpora.org/coca

Fig. 2. Formative assessment for the word "Biodiverse"

Finally, using the COCA data, we grouped words into three levels: level 1, level 2, and level 3 (see Table I). These levels serve as a proxy for word difficulty differences, because when word occurrence is lower, children have fewer opportunities to learn the word. Then, for each participant, we randomly assigned one word from each level to construct groups of words, which each contained one word from each level.

2) Instructional Design and Mediational Flow: We used formative assessment data to guide the flow of contingent mediations. On the screen, students were first presented with a target word, four possible images that might reflect its meaning, and an "I don't know" button. Students were told the following by a teacher:

> *I'm going to ask you the meaning of a hard word. So, if you don't know its meaning, you can click on the "I don't know" button at the bottom. But, if you do know the word's meaning, you can click on the picture that shows the word's meaning. The word is* ***[target word]****.*

The four picture choices represented three different levels of the student's word meaning understanding [18]:

- Accurate knowledge (response reflected accurate understanding of the word meaning)
- Partial knowledge (response reflected understanding related to one morpheme in the word; picture choices were provided for each of the word's two morphemes)
- No knowledge (response reflected inaccurate knowledge of the word meaning or selecting the "I don't know" button)

Figure 2 presents an example of the formative assessment for the word "biodiverse".

Based on a student's response to the formative assessment, contingent task mediation was provided [3]. If the student chose the incorrect picture (i.e., no knowledge) or the "I don't know" button, they received the full mediation (Figure 3(a)). If they chose one of the two pictures that reflected one of the word's morpheme meanings (i.e., partial knowledge), they received partial mediation for the unfamiliar morpheme (Figure 3(b)). If they chose the picture that accurately reflected the word's meaning, they were shown a screen that said "Good work!" and then continued to the next target word.

Full mediation included the following components:

- Teacher requests the student to **say the word** – this creates a phonological imprint [17].

TABLE I

WORDS USED FOR THE STUDY AND THEIR RELEVANT COCA DATA

Level	Word	Morpheme 1	Morpheme 2	Word occurrence	Number of texts in which word occurs	% of all texts that have the word
3	biodiverse	bio	diverse	65	58	0.01%
	lithosphere	litho	sphere	124	46	0.01%
	unicellular	uni	cellular	153	85	0.02%
2	hydropower	hydro	power	529	295	0.06%
	homeostasis	homeo	stasis	592	292	0.06%
	pseudoscience	pseudo	science	583	380	0.08%
1	geothermal	geo	thermal	1451	633	0.13%
	photosynthesis	photo	synthesis	1395	637	0.13%
	topography	topo	graphy	1709	1300	0.27%

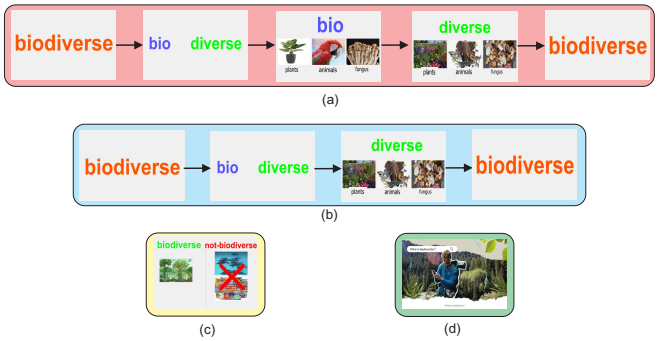

Fig. 3. Contingent mediation flows: (a) full mediation for the word "biodiverse", (b) partial mediation when "diverse" is the unfamiliar morpheme, (c) the first extended mediation, and (d) the second extended mediation

- Teacher requests the student **write the word** – this creates a graphic imprint [17].
- Teacher provides an oral meaning and illustration for **morpheme 1** [19].
- Teacher provides an oral meaning and illustration for **morpheme 2** [19].
- Teacher provides a student-friendly **definition** of the target word's meaning [17].

Partial mediation included the following components:

- Teacher provides an oral meaning and illustration for the **unfamiliar morpheme** [19].
- Teacher provides a student-friendly **definition** of the target word's meaning [17].

After full or partial mediation, the student was provided with the formative assessment again. If the student chose the incorrect picture or the "I don't know" button, they received the first extended mediation Figure 3(c). If they chose one of the two pictures that reflected one of the word's morpheme meanings, they received partial mediation for the unfamiliar morpheme. If they chose the picture that accurately reflected the word's meaning, they were shown a screen that said "Good work!" and then continued to the next target word.

The first extended mediation included the following [20]:

- Teacher presents illustrations of **examples** and **non-examples** of the target word's meaning on opposite sides of a visual chart.
- Teacher explains why some illustrations are examples, such as discussing their key characteristics, and why some illustrations are non-examples, such as identifying key characteristics that are missing.
- Teacher provides a student-friendly **definition** of the target word's meaning.

After this second round of mediation, the student was provided with the formative assessment again. This time, if the student chose the incorrect picture or the "I don't know" button, they received the second extended mediation Figure 3(d). If they chose one of the two pictures that reflected one of the word's morpheme meanings, they received partial mediation for the unfamiliar morpheme. If they chose the picture that accurately reflected the word's meaning, they were shown a screen that said "Good work!" and then continued to the next selected target word.

The second extended mediation included the following:

- Teacher provides a student-friendly **definition** of the target word's meaning [17].
- Teacher presents a **video** that explains the target word's meaning [21].

Since the mediational flow depended on the child's formative assessment response after each round of mediations, the number of rounds of mediations varied for each child. Additionally, across mediations, we presented the target word using salient text (large, bold, and brightly colored) to draw attention to the target word [22].

D. Social Mediating Agent Designs

Three social mediating agents served as teachers to deliver the pedagogical design: a Pepper robot named Lexi (Figure 1a, a virtual agent named Ari (Figure 1b), and a human teacher who had experience teaching children. Lexi, Ari, and the human all followed the same script that outlined the oral and body language mediations to deliver the instructional and mediational flows described in Section III-C.2.

Lexi is a 120 cm tall robot with a humanoid upper body, an omnidirectional base, and a custom robotic voice alongside co-speech gestures. In order to implement the oral and body language script to deliver the instructions and mediations,

we broke down the script into individual behaviors that were then designed in Choregraphe[2]. Choregraphe is a block-based programming application for creating autonomous behaviors and interaction flows for Lexi. The prosody, intonation, and motions were carefully designed by engineers and verified by a teacher with 23 years of experience in literacy education with young children. There were two main body language meditations: 1) pointing to a specific location on the screen that related to the oral content being delivered and 2) turning the robot's torso towards the screen when it was orally referring to content on it. The teacher verified the degree to which the behaviors authentically replicated human teacher behaviors and the understandability as well as the naturalness of these behaviors for this demographic group. The robot's behaviors were then linked to touchscreen events so that when a participant made a selection on the screen it would automatically guide the individual through the appropriate instructional and mediational flow, which were synchronized with the robot behaviors.

Ari is an animated dictionary, designed by a graphical artist, with anthropomorphic features such as eyes, mouth, and hair, and uses Google text-to-speech for its voice[3]. The prosody, intonation, and motions of Ari were designed in a similar manner to the robot. The body language mediations differed from the robot condition because 1) it moved to the specific location on the screen when orally referring to specific content and 2) it was situated in the middle of the screen when orally referring to the overall screen content. Similar to the robot, Ari's behaviors were linked to touchscreen events. In both the robot and virtual agent conditions, a researcher needed to indicate to the application whether the word was vocalized and written correctly by the participant in the full mediation. This was due to limitations and lack of reliability of speech and handwriting recognition techniques for children [23].

E. Procedures

Participants were seated at a table where the vocabulary instruction application was set up on a touchscreen. During each session, the participant interacted with the robot, computer, and human teacher to learn nine science words (three words per condition) using the pedagogical design described in Section III-C. The presentation of words in each condition was randomized and counterbalanced across participants. A complete session with a child took 30–45 minutes. It had four kinds of activities: (1) teacher introduction, (2) physical activity to provide mental breaks and reduce stress, (3) science word meaning instruction, and (4) post-test.

1) Teacher Introduction: Scripted teacher introductions were used by each teacher to introduce themselves. While each teacher script varied slightly, they all included these same components: teachers' name, their interest in vocabulary words, and their excitement to work with the child.

2) Physical Activity: After the introduction, the teacher offered the child a choice of two possible physical activities: (1) a Roblox Run that consisted of an obstacle course depicted on the screen or (2) a Fortnite Dance that consisted of dance moves performed by Fortnite characters on the screen. The human and robot teachers participated in the activities to build rapport. This was not possible with the virtual agent.

3) Science Word Meaning Instruction: Next, the child received approximately 10-15 minutes of instruction from the teacher for three science words. See section III-C for the instructional flow and types of mediations that were used.

The child's learning process was recorded by logging the interactions the child was having with the touchscreen. This log provided insights for assessing the child's initial response for each target word, all the subsequent contingent mediations the child received, and whether they identified the meaning of the target word correctly after their last mediation. A total of 20 log data records were collected.

4) Science Vocabulary Post-Test: At the end of the session, the human teacher conducted a post-test interview on the science words learned. For each target word, the teacher read a definition and four word choices (all target words from the study) and asked the child which word went with the definition. Then, the teacher recorded the child's response. This post-test evaluation was designed to differ from the task children performed during instruction in order to ensure that children learned the definition of the word instead of simply deducing the correct answer from the four possible choices during the learning task's formative assessment.

IV. Data Analysis

RQ1: To answer RQ1, we used the log files to identify each child's initial assessment response, all subsequent mediations, and whether the child got the word correct on the assessment after the last mediation. For example, the pattern "IDK→ M1 & M2→ M2→ M1→ Correct" means that the child initially responded "I don't know" when asked the word meaning, the first (M1) and second (M2) morphemes were taught (i.e., mediation 1), the second morpheme was taught again (i.e., mediation 2), the first morpheme was taught again (i.e., mediation 3), and finally the child got the word correct.

RQ2: To answer RQ2, we created a database of all the target words in the study and identified each mediation pattern that occurred for that word and the type of teacher that provided the mediation. Then, we sorted the database by types of teachers that provided the mediation. Next, we sorted the database within each type of teacher by the mediation patterns. Given that there were 23 patterns, we focused on patterns that occurred more than once within at least one of the teacher conditions. For each pattern that met this criterion, we calculated the percentage of occurrence within each teacher condition for each pattern.

RQ3: To answer RQ3, we created a database that included all the mediation patterns that occurred in the study for each teacher condition. We also identified the grade of the child who received each mediational pattern. Then, we sorted all the mediation patterns in each teacher condition by the grade of the child who received the mediations. Next, we sorted all the mediation patterns in each teacher condition for each grade level by the mediation patterns. Again, due to the

[2]https://www.aldebaran.com

[3]https://cloud.google.com/text-to-speech

large number of mediational patterns, our analysis focused on patterns that occurred more than once within at least one of the teacher conditions for a particular grade level. For each pattern that met this criterion, we calculated the percentage of occurrences within each teacher condition for each pattern.

RQ4: To answer RQ4, we extended the database used for RQ3 by adding columns after each teacher condition to add the total number of mediations provided in each mediational pattern. For example, recall from the description of the data analysis for RQ1 that the pattern "IDK→ M1 & M2→ M2→ M1→ Correct" included three mediations. Next, the average number of mediations for each condition were calculated.

RQ5: To answer RQ5, we statistically analyzed the children's target vocabulary knowledge pre and post-test results. For the pre-test data, the children's initial answer to the formative assessment question for each target word meaning was collected using the log files. Children's level of word-meaning knowledge on the pre-test was scored as follows: 0 for wrong or "I don't know" answers, 0.5 for the partially correct answers (i.e., they demonstrated understanding of one of the two morphemes in the word), and 1 for the correct answers, and then aggregated for each teachers' condition. Similarly, children's level of knowledge on the post-test was scored 0 for incorrect answers (there was not an "I don't know" option) and 1 for correct (there were no partial knowledge response options on the post-test) and aggregated to have a final score for that teacher's condition. To investigate the effects of teacher type on children's vocabulary learning, we used the post-test results as a dependent variable and the teacher type as the independent variable. Also, we considered the pre-test results as the covariate and fit an Ordinal Logistic Regression model to examine the effects of teacher types on the children's learning performance.

V. Findings

RQ1: We identified 23 distinct mediation patterns from the log data These patterns included a range of 0-5 mediations.

RQ2: As you can see in Table II, there are no clear mediation pattern differences across the three teacher conditions. In fact, 18 mediational patterns occurred just once per teacher condition, and thus were excluded from the table. This demonstrates the broad variation in mediational patterns provided to children across words for all conditions.

Further, while patterns often occurred more than once across two conditions, in which conditions each pattern occurred varied by word. For example, for the word litho-sphere, the mediational pattern "IDK→ M1 & M2→ Correct" occurred 50% of all mediations in the human condition and 40% of all mediations in the robot condition, but not at all in the virtual condition. In contrast, for the word pseudoscience, the mediational pattern "M2→ M1→ M2→ Correct" occurred 50% of all mediations in the virtual condition and 25% of all mediations in the robot condition, but not at all in the human condition.

Based on this initial analysis, we conclude that the variation in mediational patterns occurred based on each child's need for specific contingent meditations for each word, rather than due to teacher type.

TABLE II
Mediation patterns for each target word by teacher type

Word	Teacher Condition		
	Human	Virtual	Robot
Bio-diverse	-	38% IDK→ M1 & M2→ Correct	40% M1→ M2→ Correct
Geo-thermal	-	-	-
Homeo-stasis	-	-	-
Hydro-power	30% M2→ M1→ Correct	-	-
Litho-sphere	50% IDK→ M1 & M2→ Correct	-	40% IDK→ M1 & M2→ Correct
Photo-synthesis	33% IDK→ M1 & M2→ Correct	-	29% IDK→ M1 & M2→ Correct
Pseudo-science	-	50% M2→ M1→ M2→ Correct	38% IDK→ M1 & M2→ M2→ Correct
Topo-graphy	33% IDK→ M1 & M2 → Correct	60% IDK→ M1 & M2→ Correct	-
Uni-cellular	29% IDK→ M1 & M2 → Correct	50% M2→ M1→ Correct	75% IDK→ M1 & M2 → Correct

RQ3: As you can see in Table III, there are no clear mediation pattern differences across the three teacher conditions for each grade level. In fact, 16 mediational patterns occurred just once per teacher condition in each grade and thus were excluded from the table. This demonstrates the broad variation in mediational patterns provided to children across words for all conditions.

Further, while patterns often occurred more than once across two teacher conditions for a particular grade, which conditions and the percentage of occurrence varied. For example, in grade 3, the mediational pattern "M1→ M2→ Correct" occurred 17% of all mediations in the virtual condition and 17% of all mediations in the robot condition, but not at all in the human condition.

Additionally, patterns occurred differently across grade-levels. For example, for grade 4, the same mediational pattern occurred 17% of all mediations in the human condition, but not at all in the virtual or robot conditions.

Based on our analysis, we conclude that the variation in mediational patterns occurred based on each child's need for specific contingent mediations for each word, rather than due to teacher type.

RQ4: The average number of mediations in each teacher condition were very similar. There was an average of 1.53, 1.76, and 1.87 mediations in the human, virtual, and robot conditions respectively. Less than 0.4 of one mediation difference between the human and robot teacher conditions is not of much practical significance.

Based on this initial analysis, it seems that while the

TABLE III

MEDIATION PATTERNS FOR EACH GRADE LEVEL BY TEACHER TYPE

Grade	Pattern	Human	Virtual	Robot
3	M1→ M2→ Correct	0%	17%	17%
3	M1→ M2→ M1 & M2→ Correct	10%	0%	0%
3	IDK→ M1 & M2 → Correct	15%	0%	13%
3	M2→ M1→ Correct	5%	13%	13%
3	M2→ M1→ M2→ Correct	10%	13%	0%
4	M1→ M2→ Correct	17%	0%	0%
4	IDK→ M1 & M2→ Correct	17%	17%	17%
4	IDK→ M1 & M2 → M2→ M1→ t-table→ Correct	11%	0%	6%
4	M2→ M1→ Correct	6%	11%	11%
4	IDK→ M1 & M2→ Correct	0%	11%	11%
5	IDK→ M1 & M2→ Correct	47%	33%	27%
5	M2→ M1→ Correct	20%	7%	7%
5	IDK→ M1 & M2→ M2→ M1→ t-table→ Correct	0%	13%	0%
5	IDK→ M1 & M2→ M2→ Correct	0%	0%	13%

TABLE IV

PRE/POST-TEST COMPARISON RESULTS USING ORDINAL LOGISTIC REGRESSION

Factor	Estimate	Std. Error	Sig.	Lower Bound	Upper Bound
Pre-test (covariate)	-0.122	0.290	0.673	-0.690	0.446
Virtual (level 0)	-0.405	0.654	0.535	-0.876	1.686
Robot (level 1)	-0.344	0.663	0.603	-1.644	0.955
Human (level 2)			*baseline		

number of mediations varies (between 0-5) per child for a particular word, the overall variation in the number of mediations provided by teacher type is quite small.

RQ5: Results of the Ordinal Logistic Regression did not show any significant differences amongst the teachers considering the pre-test scores in the model. The model's statistics are presented in Table IV. None of the factors (e.g., pre-test score or teacher type) showed a significant association with the dependent variable (i.e., post-test score).

VI. DISCUSSION

Overall, our study demonstrates that high-quality contingent mediations support children's science vocabulary learning during instruction, regardless of the teaching agent, as we found no significant differences across teacher types in children's post-test knowledge of target vocabulary meanings. These findings extend Westlund and colleagues' research [7], [8], which also found no differences across types of teachers when instruction was held constant, by exploring and showing that this was also the case for contingent mediations that were provided systematically across teacher types as well.

Our findings underscore the importance of contingent meditations for maximizing children's vocabulary learning. While students in our study learned 100% of the target words during instruction with contingent mediations, and accurately identified meanings for 84% afterwards, children in Westlund and colleagues [8] studies learned just 43% and 71% of words, respectively, with non-contingent mediations.

Additionally, we extended prior research by exploring the instructional patterns that occurred when meditations were delivered contingently based on formative assessment of student responses. While prior research explored the use of curriculum-level contingent mediations, such as adjusting word difficulty based on a child's performance [7], ours is the first to explore task-level contingent mediations delivered by robot and virtual teaching agents. We identified 23 unique contingent mediational flow patterns in our dataset, demonstrating the varied needs of children to support their vocabulary learning. Taken together, these findings suggest that future research should focus more on contingent task-level mediations to optimize children's learning.

While we found no consistent patterns of mediations across teacher types, we conjecture this is because mediation patterns occurred based on each child's needs. This is supported by the high rate of correct answers during instruction and post-test. Further, like prior studies that compared robot and virtual teachers that had similar affordances and found no differences in children's learning when affordances were held constant [9], we found that when contingent mediational responses were held constant in response to learners' formative assessment data, learning was not different across teacher types. That is, our research provides further evidence that any teaching agent is as good as another if they have the same affordances and quality of mediations. Future research should investigate how affordances of different teachers can uniquely address varying student learning styles and needs.

These findings are notable for children's education because they indicate that if robots are developed to deliver high-quality teaching meditations, they can offer pathways for children to learn complex concepts and compete on par with the learning gains produced by a human teacher. Robots also have the unique effect that children are often more eager to work with them than human or virtual agent teachers during instructional activities [9]. Hence, the combination of high-quality teaching mediations and the motivational effects of social robots has the potential to lead to greater learning gains within the classroom. Robots also offer the opportunity for scaling personalized pathways for learning a new concept through contingent mediations, whereas this would be infeasible for a human teacher within a classroom of 30 children who each gneed unique mediational flows.

It can be argued that virtual agents have similar advantages to robots in terms of scalability and, consequently, it is important to consider what educational activities they can most benefit. Educational activities that require embodiment and a teacher to be physically situated in the same space as a learner could leverage the benefits of robots. One such example is the use of tactile practices as a part of contingent

mediations to allow for a multisensory approach (oral, visual, tactile), which better supports some children's learning as compared with using non-tactile approaches [24]. For example, multisensory phonics instruction teaches children the relationship between letters of a written language and sounds of the spoken language, and a part of that instruction has students writing letters in the sand, or manipulating letter tiles to blend sounds in words [25]. A robot could demonstrate the tactile activities as part of small group instruction, but a virtual agent cannot offer such capabilities.

Overall, this study emphasizes the need to carefully consider how contingent mediations are crafted in a pedagogical design aimed to be deployed with a robot, as it is a crucial factor that leads to a child's learning outcomes. From a technical standpoint, robots need to not only be adaptive to whether a child's response is correct or not, but also be capable of formative assessment of a child's response and subsequent selection of the best contingent mediation that should follow in the instructional flow. Such capabilities would provide a two-fold benefit of enabling a robot to identify appropriate task-level instructional mediations that are contingent to a child's current learning development and ensure a child remains engaged and motivated by receiving instruction that best fits their learning needs.

VII. Limitations

Our study has three limitations: (1) The human teacher was constrained to a script for the delivery of the educational content, which enabled a more controlled study design, but it would be valuable to investigate whether an unconstrained human teacher would lead to similar results. (2) Children sometimes deviated from the planned interaction, such as starting a side conversation, at which point the human teacher had to respond because the robot and virtual agent were not programmed to do so. Future research might explore designing such behavioral management abilities in robots and virtual agents. (3) There is the risk that the child's formative assessment responses did not accurately reflect their target vocabulary knowledge, but rather may have been good guesses. This problem likely had a low occurrence rate due to the high percentage of correct responses on the post-test (84%).

Acknowledgements

The authors would like to thank Keira Askew for the design and artwork of the virtual agent Ari.

References

[1] T. Belpaeme, J. Kennedy, A. Ramachandran, B. Scassellati, and F. Tanaka, "Social robots for education: A review," *Science robotics*, vol. 3, no. 21, pp. 1–9, 2018.

[2] D. Leyzberg, S. Spaulding, M. Toneva, and B. Scassellati, "The physical presence of a robot tutor increases cognitive learning gains," in *Proceedings of the annual meeting of the cognitive science society*, vol. 34, no. 34, 2012, pp. 1882–1887.

[3] E. Rassaei, "Tailoring mediation to learners' zpd: Effects of dynamic and non-dynamic corrective feedback on l2 development," *The Language Learning Journal*, vol. 47, no. 5, pp. 591–607, 2019.

[4] I. Bakhoda and K. Shabani, "Bringing l2 learners' learning preferences in the mediating process through computerized dynamic assessment," *Computer Assisted Language Learning*, vol. 32, no. 3, pp. 210–236, 2019.

[5] A. Kozulin, *Vygotsky's educational theory in cultural context*. Cambridge, United Kingdom: Cambridge University Press, 2003.

[6] R. G. Tharp and R. Gallimore, *Rousing minds to life: Teaching, learning, and schooling in social context*. Cambridge, United Kingdom: Cambridge University Press, 1991.

[7] J. K. Westlund, *et al.*, "A comparison of children learning new words from robots, tablets, & people," in *Proceedings of the 1st international conference on social robots in therapy and education*, 2015.

[8] J. M. K. Westlund, *et al.*, "Children use non-verbal cues to learn new words from robots as well as people," *International Journal of Child-Computer Interaction*, vol. 13, pp. 1–9, 2017.

[9] Z. Zhexenova, *et al.*, "A comparison of social robot to tablet and teacher in a new script learning context," *Frontiers in Robotics and AI*, vol. 7, pp. 1–14, 2020.

[10] E.-j. Hyun, S.-y. Kim, S. Jang, and S. Park, "Comparative study of effects of language instruction program using intelligence robot and multimedia on linguistic ability of young children," in *RO-MAN 2008-The 17th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 2008, pp. 187–192.

[11] E. A. Konijn, B. Jansen, V. Mondaca Bustos, V. L. Hobbelink, and D. Preciado Vanegas, "Social robots for (second) language learning in (migrant) primary school children," *International Journal of Social Robotics*, pp. 1–17, 2022.

[12] H.-S. Hsiao, C.-S. Chang, C.-Y. Lin, and H.-L. Hsu, ""irobiq": the influence of bidirectional interaction on kindergarteners' reading motivation, literacy, and behavior," *Interactive Learning Environments*, vol. 23, no. 3, pp. 269–292, 2015.

[13] D. C. Tozadore, A. H. Pinto, C. M. Ranieri, M. R. Batista, and R. A. Romero, "Tablets and humanoid robots as engaging platforms for teaching languages," in *2017 Latin American Robotics Symposium (LARS) and 2017 Brazilian Symposium on Robotics (SBR)*. IEEE, 2017, pp. 1–6.

[14] M. De Haas, P. Vogt, and E. Krahmer, "The effects of feedback on children's engagement and learning outcomes in robot-assisted second language learning," *Frontiers in Robotics and AI*, vol. 7, pp. 1–18, 2020.

[15] V. Lin, H.-C. Yeh, and N.-S. Chen, "A systematic review on oral interactions in robot-assisted language learning," *Electronics*, vol. 11, no. 2, pp. 1–37, 2022.

[16] R. Van den Berghe, J. Verhagen, O. Oudgenoeg-Paz, S. Van der Ven, and P. Leseman, "Social robots for language learning: A review," *Review of Educational Research*, vol. 89, no. 2, pp. 259–295, 2019.

[17] I. L. Beck, M. G. McKeown, and L. Kucan, *Bringing words to life: Robust vocabulary instruction*. New York City, USA: Guilford Press, 2013.

[18] T. Christ, "Moving past "right" or "wrong" toward a continuum of young children's semantic knowledge," *Journal of Literacy Research*, vol. 43, no. 2, pp. 130–158, 2011.

[19] K. Apel, E. B. Wilson-Fowler, D. Brimo, and N. A. Perrin, "Metalinguistic contributions to reading and spelling in second and third grade students," *Reading and writing*, vol. 25, pp. 1283–1305, 2012.

[20] D. A. Frayer, W. C. Fredrick, and H. J. Klausmeier, *A schema for testing the level of concept mastery*. Wisconsin, USA: Wisconsin University Research & Development Center for Cognitive Learning, 1969.

[21] S. B. Neuman, "Giving all children a good start: The effects of an embedded multimedia intervention for narrowing the vocabulary gap before kindergarten," in *Technology as a support for literacy achievements for children at risk*. Springer, 2012, pp. 21–32.

[22] L. M. Justice, L. Skibbe, A. Canning, and C. Lankford, "Pre-schoolers, print and storybooks: An observational study using eye movement analysis," *Journal of Research in Reading*, vol. 28, no. 3, pp. 229–243, 2005.

[23] V. Bhardwaj, *et al.*, "Automatic speech recognition (asr) systems for children: A systematic literature review," *Applied Sciences*, vol. 12, no. 9, p. 4419, 2022.

[24] K. Warnick and P. Caldarella, "Using multisensory phonics to foster reading skills of adolescent delinquents," *Reading & Writing Quarterly*, vol. 32, no. 4, pp. 317–335, 2016.

[25] Brianspring, *Phonicsfirst curriculum guide*. Brianspring, 2022, dimensions: 9.5 x 11 inches, Binding: 2 volumes, spiralbound, Page Count: Volume I - 303; Volume II - 264.