Speech Communication 170 (2025) 103206

Contents lists available at ScienceDirect

Speech Communication

journal homepage: www.elsevier.com/locate/specom

The Ohio Child Speech Corpus

Laura Wagner [a,*], Sharifa Alghowinhem [b], Abeer Alwan [c], Kristina Bowdrie [d], Cynthia Breazeal [b], Cynthia G. Clopper [e], Eric Fosler-Lussier [f], Izabela A. Jamsek [d], Devan Lander [d], Rajiv Ramnath [f], Jory Ross [e]

[a] *Department of Psychology, The Ohio State University, United States*
[b] *MIT Media Lab, Massachusetts Institute of Technology, United States*
[c] *Department of Electrical and Computer Engineering, UCLA, United States*
[d] *Department of Speech and Hearing Science, The Ohio State University, United States*
[e] *Department of Linguistics, The Ohio State University, United States*
[f] *Department of Computer Science and Engineering, The Ohio State University, United States*

ARTICLE INFO

Keywords:
Child language corpus
Social robot
School-readiness
Peri-pandemic

ABSTRACT

This paper reports on the creation and composition of a new corpus of children's speech, the Ohio Child Speech Corpus, which is publicly available on the Talkbank-CHILDES website. The audio corpus contains speech samples from 303 children ranging in age from 4 – 9 years old, all of whom participated in a seven-task elicitation protocol conducted in a science museum lab. In addition, an interactive social robot controlled by the researchers joined the sessions for approximately 60% of the children, and the corpus itself was collected in the peri-pandemic period. Two analyses are reported that highlighted these last two features. One set of analyses found that the children spoke significantly more in the presence of the robot relative to its absence, but no effects of speech complexity (as measured by MLU) were found for the robot's presence. Another set of analyses compared children tested immediately post-pandemic to children tested a year later on two school-readiness tasks, an Alphabet task and a Reading Passages task. This analysis showed no negative impact on these tasks for our highly-educated sample of children just coming off of the pandemic relative to those tested later. These analyses demonstrate just two possible types of questions that this corpus could be used to investigate.

1. Introduction

This paper reports on a new corpus of children's speech, the Ohio Child Speech Corpus (OCSC) currently available on the Talkbank-CHILDES corpora website (MacWhinney, 2000). All corpora on the CHILDES-Talkbank site can be accessed for free. The site contains data from several thousand children, including children acquiring English as well as dozens of other languages, and some children acquiring their language atypically. The data has been contributed to this resource by researchers over the course of more than 60 years and reflects the research questions and technology of those researchers. Thus the contents of individual corpus entries include children of a range of ages, doing a range of tasks (or no tasks at all), and recorded in a range of ways (video, audio, transcripts with no associated recordings). This website is widely used by language development researchers.

The OCSC contains speech samples from 303 children ranging in age from 4 – 9 years who all went through the same elicitation procedure. There are several distinctive features of this corpus that make it a unique addition to the field: the number and ages of the participants, the elicitation protocol used, the embedded study involving a social robot, the dual audio recording (i.e. both high and low fidelity), and the specific timing in which the corpus was collected (i.e. during the two summers after the COVID-19 pandemic lockdown). In addition, we note that the corpus was created at a lab embedded within a science museum and its participants were recruited from among the visitors to the museum and tested in a glass-enclosed room that was visible to other visitors. Thus, the entire process of creating the corpus was part of a museum exhibit with the aim of exposing the general public to what science "really" looks like. We discuss the value of each of these distinctive features and report on two preliminary analyses that make use of a selection of the corpus materials.

* Correspondence author at: Department of Psychology, 1835 Neil Avenue, Columbus, OH, 43210, United States
E-mail addresses: Wagner.602@osu.edu (L. Wagner), Sharifah@media.mit.edu (S. Alghowinhem), alwan@ee.ucla.edu (A. Alwan), Bowdrie.1@osu.edu (K. Bowdrie), Cynthiab@media.mit.edu (C. Breazeal), Clopper.1@osu.edu (C.G. Clopper), Fosler-lussier.1@osu.edu (E. Fosler-Lussier), Jamsek.1@buckeyemail.osu.edu (I.A. Jamsek), Lander.50@buckeyemail.osu.edu (D. Lander), Ramnath.6@osu.edu (R. Ramnath), Ross.1589@buckeyemail.osu.edu (J. Ross).

https://doi.org/10.1016/j.specom.2025.103206
Received 21 March 2024; Received in revised form 8 February 2025; Accepted 13 February 2025
Available online 4 March 2025

1.1. Distinctive corpus features

1.1.1. Large child sample in the early school year ages

The OCSC contains a large sample ($N = 303$) of young school-age children, ranging in age from 4 – 9 years old. The OCSC is publicly available on the Talkbank-CHILDES database (MacWhinney, 2000), where it is distinctive in terms of its size, its age demographic, and its composition. Among the approximately 80 regular North American and British English corpora currently available, only one third contain overlap with the current age span and only 7% include at least 15 children from a comparable age span. There are substantially more children in this age range in the Narrative corpora section (e.g. Peterson and McCabe, 1983; Hicks, 1990) and Frog Story corpora section (cf. Berman and Slobin, 1994) available on the Talkbank-CHILDES site, but children in those corpora typically engaged in just a single story-telling task and not the range of tasks included here. The Linguistics Data Consortium (LDC) is another large resource for corpora. In the LDC, there are almost two dozen corpora containing large numbers of children, many overlapping with the current age range (e.g. Eskenazi et al., 1997). However, in these corpora, children often provided only very small individual samples, sometimes consisting of solely reading a word list (e.g. Leonard and Doddington, 1993). Thus, those data do not allow for the range of possible analyses that the current corpus does.

Language skills grow and change between the ages of 4 and 9 years (C. Chomsky, 1969; Nippold, 2016) and this corpus will facilitate investigations of a range of different language skills in this age range. However, while the large sample size of the OCSC will provide some reasonable assurances of generalizability of findings, it should be noted that the demographics of the children in the corpus are highly specific (see Table 1).

1.1.2. Elicitation protocol

The same protocol was used to elicit speech from all of the children in the OCSC, and it consisted of a variety of tasks (see Table 2). The tasks were inspired by standardized tests used to assess children's language (Richard and Hanner, 2005; Zimmerman et al., 2011; Gillam and Pearson, 2017). They included some tasks designed to explicitly highlight school readiness skills, such as alphabet and number knowledge, as well as the inclusion of short reading passages for older children. Other tasks were designed to encourage more open discourse styles such as describing amusing pictures and narratively-oriented pictures, as well as explaining how to do common tasks. In addition, one task was a classic task from the language development literature: the Wug task (Berko, 1958), used to elicit plural forms. Not all children completed all items within each task and not all children completed every task. Nevertheless, the consistency of the speech samples across children, and especially across the age range covered by the corpus, makes it feasible to use the corpus to do cross-sectional age comparisons of specific linguistic skills.

1.1.3. Social robot assistant & embedded study

A particularly unique element of this corpus was the presence of a social robot, Jibo, for over half of the participating children. Social robots are designed to facilitate and participate in appropriate conversations and interactions with others. They are increasingly a part of children's lives, including being involved in educational testing and assessment contexts (Wik and Hjalmarsson, 2009; Belpaeme et al., 2018; Spitale et al., 2020; Shahab et al., 2024). As we will discuss in more detail below, the impact of these devices on children's language production has not yet been settled in the literature. Within our corpus, approximately 40% of the children went through the protocol with the social robot completely absent from the room while the remaining 60% of children were introduced to the robot and had periodic interactions with him throughout the session as it provided praise and encouragement. Moreover, we gathered some basic information about children's background with social robots from the parents which can be used as a predictor of the effect of the social robot on their linguistic performance. In Section 3, we report in detail on how the presence and interaction type of the social robot influenced children's language production in the Narrative Pictures task.

1.1.4. Dual recording of speech

The corpus was recorded in a museum setting which was a somewhat noisy environment. We collected the fluctuating ambient decibel levels in the room before each child was run, and these ambient levels ranged from 50 dB SPL to 75 dB SPL. To generate a clear audio track, children wore a high-fidelity lapel microphone that was placed on their clothing quite close to their mouths. In addition, we also recorded each session with a low-fidelity table microphone that was placed several feet from the child. The low-fidelity audio recordings feature much more of the environmental noise produced by the museum environment. To facilitate synching between the two audio streams, children pressed noise-making buttons after each task which provided clear audio boundaries at regular intervals in the session. The high-fidelity audio track is of interest to researchers whose goals involve understanding the development of children's language skills; this track is the one linked directly to the transcripts inside the Talkbank-CHILDES database site. However, the low-fidelity audio recording may be of interest to researchers in the field of automatic speech recognition (ASR). In the real world, noisy environments are common but they often pose challenges to automatic

Table 1
Participant information.

	4-yr-olds	5-yr-olds	6-yr-olds	7-yr-olds	8-yr-olds	9-yr-olds	All Children
N	26	54	60	63	57	43	303
Mean age (years)	4.57	5.56	6.5	7.54	8.49	9.40	7.18
(range)	4.1 - 4.98	5.03 – 5.97	6.0 – 6.99	7.01 – 7.99	8.02 – 8.98	9.01 – 9.94	4.1 – 9.94
Girls/Boys/Other[1]	18/8/0	27/27/0	33/27/0	34/28/1	29/28/0	25/18/0	166/136/1
Race	White: 19	White: 38	White: 52	White: 50	White: 54	White: 35	White: 248
	Black: 3	Black: 4	Black: 4	Black: 4	Black: 3	Black: 4	Black: 22
	Asian: 2	Asian: 3	Asian: 3	Asian: 3	Asian: 0	Asian: 1	Asian: 12
	Multiple: 0	Multiple: 8	Multiple: 1	Multiple: 2	Multiple: 0	Multiple: 1	Multiple: 12
	Unknown: 2	Unknown: 1	Unknown: 0	Unknown: 4	Unknown: 0	Unknown: 2	Unknown: 9
Ethnicity	Hispanic: 2	Hispanic: 2	Hispanic: 2	Hispanic: 4	Hispanic: 2	Hispanic: 4	Hispanic: 24
N multilingual	3	8	2	7	5	2	27
N with recent speech/language problems	1	1	4	4	2	3	15
Average Session Length in Minutes (range)	25.1	28.8	30.6	32.9	32.4	32.6	30.9
	(9.5 – 37.7)	(9.2 – 49.4)	(9.8 – 53.5)	(19.9 – 55.3)	(13.8 – 47.2)	(21.8 – 50.0)	(9.2 – 55.3)
Robot Absent	14	22	26	28	19	12	121
Robot Present (Encouragement)	6	9	10	9	19	11	64
Robot Present (Instruction)	5	9	16	14	9	9	62
Robot Present (Presents images)	1	14	8	12	10	11	56

[1] One parent indicated that their child should not be classified as either a boy or a girl

Table 2
Elicitation Materials.

Task	Items	Specifications	Sample Item
Alphabet	26 items	For each letter: • Label it • Label a picture that starts with it • Come up with a new word starting with it	
Numbers	28 numbers (1 – 15, 25, 20 – 100 by tens, 100, 200, 500, 1000) 4 math problems	• Label numbers • Read simple math problems (and attempt computations)	**9 + 7 =**
Wug	10 familiar words 10 nonsense words	• Singular form of the word is provided along with a picture • Child is encouraged to produce the plural form	Example nonsense picture: 
Narrative Pictures (embedded study)	4 Items	• Describe what's happening in this picture	
Routines	16 items	• Explain how to do everyday activities	Example: breakfast 
Reading Passages	1 or 2 passages (from Cartledge et al., 2015)	• Read a short passage	Sample passage titles: *Arthur Ashe, Bike Race When I Grow Up*
Whimsical Pictures	10 items	• Describe what's in this picture	

Note: The pictures depicting common words in the Wug task were drawn from the Massive Memory database (Brady et al., 2008) and the pictures depicting nonsense words were drawn from the NOUN database (Horst and Hout, 2016). All other pictures were commissioned for this project and were drawn by Rebecca Hinkelman.

systems (Li et al., 2014). Our noisier recording offers an opportunity for researchers to test the robustness of their systems and compare it to performance with clear audio. The low-fidelity audio track is available by request.

1.1.5. Peri-Pandemic timing of data collection

We had not intended the timing of the data collection to be a notable feature of the OCSC. However, we began collecting speech samples in June 2021 and finished in early September 2022. In Columbus, Ohio, where the corpus was collected, schools closed for in-person instruction in March 2020 and remained virtual in most locations for the rest of that school year. Regular in-person classroom instruction was broadly resumed in the fall of 2021 for the 2021–2022 school year. The first wave of children we worked with were just coming off of a year of virtual schooling. Moreover, the lab itself in 2021 was still subject to a variety of COVID-related precautions. For example, the experimenters wore face-masks and most of the children wore face-shields during the sessions. The museum where the lab is located (see next section) had been closed for 15 months and had just re-opened that June. We worked with 106 children over the summer (June – August) of 2021. An additional 48 children were run over the following school-year (September 2021 – May 2022) which coincided with schools resuming in-person instruction in Ohio. The next major wave of data collection was in summer 2022 (ending on Labor Day) when the remaining 149 children were run. By this second summer, all COVID restrictions had been lifted, both in Ohio and in the lab. This corpus provides a snapshot record of linguistic skills – including school-readiness skills – for two notable times: children who were just coming off of a year of virtual schooling (and general pandemic-related restrictions) and children who were one year past that moment. The OCSC thus allows for investigations of how the pandemic may have influenced children of different ages, both in the immediate aftermath of broad shut-downs as well as potential longer-term effects. That said, we note that each child was run only once, so any comparisons are necessarily between-subjects and cross-sectional in nature. Moreover, we have no specific information about what each child's individual experience was during the pandemic and therefore can only classify children based on the typical experience of children in Ohio at that time. In Section 4, we report on an investigation of two tasks – the Alphabet task and Reading Passages task – to see if performance was influenced by the specific pandemic-related timing when the children performed them.

1.1.6. Museum setting

This corpus was collected inside a unique lab space. The Language Sciences Research Lab is embedded inside the Columbus, Ohio, Center of Science and Industry (COSI) and the physical space consists of large "fish-bowl" style rooms where visitors to the museum can watch research as it is happening (Wagner et al., 2015). Families are recruited from the floor of the museum and invited to participate in contributing to a "real" science experiment. Research assistants were trained to provide educationally rich explanations of the work that are accessible to the museum-going public. The creation of a corpus of children's speech is not a canonical science activity for most people and thus this project served as a large-scale public demonstration about how language scientists conduct their research.

1.2. Potential uses for this corpus

This corpus was created by a large team (see the co-authors list) with varied research interests. The uses of this corpus are as varied as the people who collected it. First and foremost, the corpus is a large sample of children's natural (elicited) speech. It is useful for language acquisition researchers who want to investigate the development of core linguistic skills in children's lexicons, syntax, and phonology. Moreover, some of the specific tasks allow for more pointed investigations of children's language skills: for example, the inclusion of the Wug task facilitates examinations of children's plural production and the Routines task facilitates examinations of children's ability to linguistically sequence events. The second major use we envision for this corpus is computational. Children's speech is not as well represented in ASR models as adult speech is and this corpus contains audio recordings of a large sample of children, along with good transcriptions. These data could be used to train ASR models. Moreover, as noted in Section 1.1.4, a second, low-quality, audio stream is available upon request. This second audio was recorded specifically to allow for more robust ASR tests. Finally, the distinctive features of this corpus – the embedded robot study, the peri-pandemic timing, and the location in a science museum – all create opportunities for other kinds of research questions. Below, we highlight two ways that the corpus can provide information in those domains, specifically the potential impact of a social robot on children's speech and the potential impact of the pandemic on school readiness. Like all corpora of naturally produced speech, there are inherent limitations to the kinds of causal conclusions one can draw from it. However, the size and composition of this corpus make it an excellent starting point for developing research questions.

2. Speech sample collection methods

2.1. Participants

All participants were recruited at a local science museum and run in a glass-enclosed space within the museum. A total of 303 children were included in the corpus, and the full demographic description of these children can be found in Table 1. We note that the designation of having "recent speech/language problems" refers to information from parental report and indicates that the children were either currently experiencing problems, currently in therapy, or had been in therapy within the last year. An additional element of demographic information, not represented on the table, is that the sample was drawn from highly educated families: 75% of children had at least one parent with a college degree or higher and for only 6% of children had neither parent attended college. An additional 5 children were run but were not included in the corpus because parents either declined permission for the data to be made public or did not provide full demographic information about the child. Basic demographic information (age, gender) is noted on each transcript; a spreadsheet containing full demographic information for every child is available on the Talkbank-CHILDES website.

2.2. Elicitation stimuli

There were seven elicitation tasks used in the session, most which were adaptations of tasks commonly used in standardized assessment tests of children's language (cf. Richard and Hanner, 2005; Zimmerman et al., 2011; Gillam and Pearson, 2017). The Alphabet task asked children to identify letters and words beginning with them; the Numbers task asked children to identify numbers and (when possible) do simple math problems; the Narrative Pictures task showed complex scenes that encouraged stories and asked children to describe them (this task was also used in the embedded robot task described below), the Routines task asked children to describe how to do everyday tasks, and the Whimsical Pictures task asked children to describe pictures containing amusing combinations of animals and objects. In addition, we used the Wug task which prompts children to produce plural forms of familiar and novel words (cf. Berko, 1958) and a Reading Passages task in which children read short passages at a 2nd grade reading level from the set used in Cartledge et al. (2015). All tasks were supported by visual items (pictures or words/numbers) which were presented on laminated cards. Table 2 describes the tasks in the order in which they were administered, including number of items and sample pictures.

2.3. Running procedures

Children were recruited from the floor of the museum space and invited (along with a parent or guardian) to come into the onsite testing space. The child sat at a small table beside a "testing" experimenter. The testing experimenter conducted all the main interactions with the child, including putting on the lapel microphone (Audio-technica MT830 Omnidirectional Condenser Lavalier Microphone). The lapel microphone was placed on the child's collar, near their mouth. At a separate table facing the child, the "technology" experimenter controlled all the logistical elements of the study, including measuring the sound levels in the room, controlling the social robot (when present), monitoring the recordings, as well as administering surveys and permission forms to the child's parent. A second, lower fidelity, microphone (Fifine Mini Gooseneck USB Microphone) was placed on the technology table and

Fig. 1. The Testing Set-up.

Fig. 2. Jibo, the social robot.

also recorded the session. Children were not video-taped for this corpus. Fig. 1 shows a picture of a child being tested; the technology experimenter was positioned at a table with the same view that the picture shows while parents were sitting off to the left. Note that this picture was taken during the summer of 2021, when full COVID-19 precautions were in use.

Parents (and occasionally, other family members such as younger siblings) sat near the technology experimenter. Parents were provided with questionnaires that asked for demographic information as well as information about the child's familiarity with socially interactive robots and computers and their reading skills and frequency. The survey questions and responses can be found at the OCSC's Talkbank-CHILDES site. Parents were provided with information about the Talkbank-CHILDES online corpus and signed permission forms allowing their child's audio recordings to be made public on this site.

All children went through the seven elicitation tasks in the order listed in Table 2. The testing experimenter guided children through the full session. She first attached the lapel microphone to the child's collar and made sure the child could reach the pictures and the noisemaking buzzers. She presented the pictures for each task which were collated in a large binder with color-coded tabs. She began each task by explaining its general purpose ("We're going to talk about some numbers now") and then followed up as needed with a set of general prompts (e.g., "What's this? Can you tell me more?). For the Alphabet task, children were shown all 26 letters in sequence, each accompanied by a word beginning with that letter and its associated picture. Children were asked to identify the letter, to label the picture, and when feasible, to think of another word that began with the same letter. For the Numbers task, children were asked to label 28 individual numbers and then solve some simple math problems. Children who were unable to do the necessary arithmetic were asked instead to label the numbers in the equation. For the Wug task, children were presented with 20 cards, half depicting sets of common objects and half depicting novel objects. Children were provided with the label for one object on each card and prompted to describe a set of the objects ("This is a wug. What would you call these? These are…"). For the Narrative Pictures task, children were presented sequentially with four different richly detailed pictures. For example, one picture showed a girl holding a giant radish in a garden and another showed a scientist in a lab with a microscope and an octopus. For each picture, they were asked to "tell me what's happening in this picture." However, for many children this was the task in which Jibo provided those instructions and sometimes presented the pictures on his screen (see Section 2.4 below). For the Routines task, children were presented with a card depicting a familiar event, such as a child washing their hands or making breakfast. The child was asked to describe the card and then to "tell me how you do that." For the Reading Passages task, the child was presented with a card showing one passage in a large font. Passages were returned to the binder at the back of their section so that all the passages would eventually be cycled through in the study. For the Whimsical Pictures task, children were presented with a small card showing an unusual pairing of items, such as a pig in a teacup or a giraffe eating a sandwich. Children were asked to "tell me what's on this picture."

If children were hesitant to talk, the testing experimenter could also provide hints, such as directing the child's attention to items in the pictures or just telling them what a letter or an object was. Children were encouraged to complete each task, but if they became frustrated or bored, the testing experimenter would move to the next task. For the Reading Passages task, if children (or their parents) indicated that they couldn't read, the task was not started. In general, we note that the priority for the testing experimenter was to encourage the child to speak even if that meant digressing from the task at hand. However, once digressions had run their course, the experimenter resumed the protocol where she had left off.

After each task, children pressed a noisemaking buzzer (their choice from among 4 buzzers). The buzzers served as mini-rewards within the session, and also provide a means for synching the two audio streams (see Section 1.1.4). Children were given a sticker for their participation at the end of the session. The average duration of each session was approximately 31 minutes long (see Table 1 for a breakdown by age group).

2.4. Social robot and embedded study

For approximately 60% of the children (see Table 1), a Jibo social robot was present throughout the session (see Fig. 2). Jibo is 12″ tall and was placed on the testing table along with the elicitation pictures. All children in the robot sessions began their session by being introduced to Jibo, who asked the child his/her name and his/her favorite color. Jibo provided encouragement (e.g. "you're doing a great job") intermittently throughout the entire session. Jibo uses a computer generated voice and he swivels his top when interacting in a reasonably anthropomorphic way. Jibo does have a variety of social capabilities that were not tapped in this experiment (he can, for example, make a virtual pizza). In order to keep the interactions with Jibo as similar as possible across the children, his interactions were restricted to a limited set of sentences and the timing of their use was controlled by the technology experimenter through a laptop computer.

During the Narrative Pictures task, Jibo could take on two additional responsibilities beyond praise and encouragement. In the Instruction condition, Jibo provided the specific instructions for the task: "Today we are going to do an activity together. I'm going to show you pictures and you're going to tell me about them." The testing experimenter would then present the child with the narrative pictures and Jibo would prompt the child for a description: "Look at the picture. What is this?" In the Picture Presentation condition, the elicitation pictures themselves were presented on Jibo's "face" screen, which measures 4.5 × 2.5 inches. Thus, there were four possible conditions for this particular task: Robot Absent (for the 40% of children for whom Jibo was not present in the session at all); Robot Encouragement (when Jibo provided praise as it did throughout the whole session); Robot Instruction (when Jibo provided praise and instructions for this particular task); Robot Picture Presentation (when Jibo provided praise, instructions, and presented the pictures on its face). Table 1 shows how many children in each age group were in each of these robot conditions.[1]

2.5. Transcription procedures

All sessions were transcribed using the audio from the high-fidelity lapel microphone. Every utterance of the child, experimenters, and any others who spoke (occasionally mothers or siblings interjected utterances) was transcribed. We used standard CLAN conventions as laid out in the Talkbank-CHILDES site to format the files, including following conventions for noting unintelligible utterances, common children's words, etc. Each transcriber went through a standardized training protocol in which their work was checked closely and repeatedly by an experienced transcriber. In addition, trained transcribers participated in regular group meetings to receive refresher tips and trouble-shoot difficult cases. Every transcript was reviewed and corrected (as needed) by an experienced transcriber. A third transcriber did a final review and checked that all conventions were consistently applied. Any information that would make the child clearly identifiable (e.g., last names, home addresses, etc.) was redacted from the transcripts and bleeped out of the audio recordings.

3. Example 1 for using the corpus: looking at potential effects of a social robot on children's speech

As noted in Section 1.1, one of the distinctive features of this corpus is the inclusion of the social robot, Jibo, in approximately half of the sessions. Social robots are becoming more common and they have a great deal of potential to facilitate interactions in educational and assessment contexts (Wik and Hjalmarsson, 2009; Belpaeme et al., 2018; Spitale et al., 2020; Esfandbod et al., 2023b; Shahab, et al., 2024). Previous research on the impact of social robots consistently shows that children are engaged by them (Kory-Westlund and Breazeal, 2015; Breazeal et al., 2016; Kanero et al., 2018; Esfandbod et al., 2023a). Measurements of children's productive language when interacting with them is mixed, but largely positive. In one study (Kory-Westlund and Breazeal, 2015), the presence of a robot teacher led typically developing preschoolers to produce more words and greater lexical diversity (relative to a human teacher) and in Esfandbod et al. (2023b), a highly interactive lip-synching social robot led to gains in a speech language therapy context. However, in Spitale et al. (2020), the effect of a robot teacher did not change behavior compared to a human teacher and in Xu et al. (2021), the robot teacher (relative to a human one) led to lower rates of language production and quality although the children did achieve higher rates of comprehension. More generally, the conclusion of the review paper by Kanero et al. (2018) is that there are not enough studies of the impact of social robots on children's language abilities to draw firm conclusions.

The study embedded within this corpus was designed to add to our understanding of how social robots impact children's general language usage. Our Jibo robot offered a rather modest type of interaction with a social robot: his speech was controlled by an experimenter and he was deployed sparingly through most of the corpus tasks. However, for one of the tasks, the Narrative Pictures task, we purposefully provided different "doses" of Jibo to different children. We noted that some of the robots with the most dramatic impacts on children involved robots that were especially active with the children and we hypothesized that increasing Jibo's responsibilities in that task would lead to increased impact. Thus, Jibo's role in that task had four levels of increasing intensity: Absent, friendly Encouragement, encouragement plus providing Instructions, and encouragement plus both providing instructions and Presenting pictures. Given the nature of our corpus, our measures of impact are both related to children's production: how much do they say (a quantity measure) and how complex are their utterances (a quality measure).

3.1. Methods

Section 2.4 above describes the Narrative Pictures task implementation, including the four roles that Jibo could play within it. We analyzed the transcripts of the 301 children within the OCSC who completed the Narrative Pictures task (two children, mean age 5.8 years, could not be included because they did not participate in that task). The dependent measures in this analysis were the number of utterances the child produced, a measure of language quantity, and the child's mean length of utterances (MLU), a very general measure of language complexity. Our analyses focused exclusively on children's performance in the Narrative Pictures task, which was the only one where we systematically varied the role of the Jibo robot. We chose this task for this embedded experiment because it was a straightforward task to administer requiring little back-and-forth interaction, thus making it easy for Jibo to provide appropriate directions. It was also a task that emphasized the child's free production as there were many ways to describe, explain, and elaborate on the rich pictures. We note that this task represented on average 26% of each child's utterances in their entire session, making it a representative task of their linguistic abilities.

3.2. Results

Fig. 3 shows the impact of the robot on the linguistic quantity measure of number of utterances for the children as a function of their age. A linear regression analysis was conducted with Robot role (Absent, Encouragement, Instruction, Presentation), Child's age (measured continuously), and their interaction as the independent variables and number of utterances as the dependent variable. This model significantly predicted the number of utterances ($R^2 = 0.074$, $p = .002$). Both Robot role and Child's age were significant predictors, but the interaction was not significant.[2] As can be seen by the positive slopes in Fig. 3, the age effect reflects the fact that children say more as they get older ($\beta = 8.169$, $p = .019$). As can be seen in Fig. 3, children produced fewer utterances in the Robot Absent condition (the solid black line) than they did in the other conditions (the remaining lines). However, the only significant difference across Robot roles was between Robot Absent and Robot Encouragement ($\beta = 82.335$, $p = .043$).

A second regression analysis was conducted with Robot role and Child's age as independent variables and the dependent variable of MLU, our measure of linguistic complexity. This model significantly predicted MLU ($R^2 = 0.112$, $p <$ 001) but for this variable, only the Child's age was a significant predictor ($\beta = 0.321$, $p < .001$), while the Robot role was not ($\beta = 0.071$, $p = .198$). As can be seen in Fig. 4, the age effect is reflected in the positive slope of all the lines (MLU increased with age), while the lack of effect for Robot role can be seen in the way

[1] The intention was to have equal numbers of children in each age group in each robot condition. The over-sampling of children without the robot present at all reflects the fact that we went through several periods where we were having technical difficulties with Jibo. We opted to run children in the Robot Absent condition rather than forego testing opportunities entirely. The uneven distribution of children in the three conditions with the robot reflects an interaction of our overall strategy of minimizing procedural changes for the experimenters within a shift of testing (thus, each experimenter pair typically ran all of their participants in a given shift in the same condition) and the fact that recruitment was done opportunistically on the museum floor (thus, some days and some shifts yielded greater numbers of participants in different age groups than others). While we endeavored to even out the cells over the lifetime of the project, we did not achieve a fully counter-balanced design.

[2] We also checked if the child's gender or the amount of familiarity with social computers influenced the number of utterances produced but neither item was a significant predictor in the model (either alone or in combination with Robot condition and Age). A similar null effect was found for Gender and Familiarity for the MLU analysis as well.

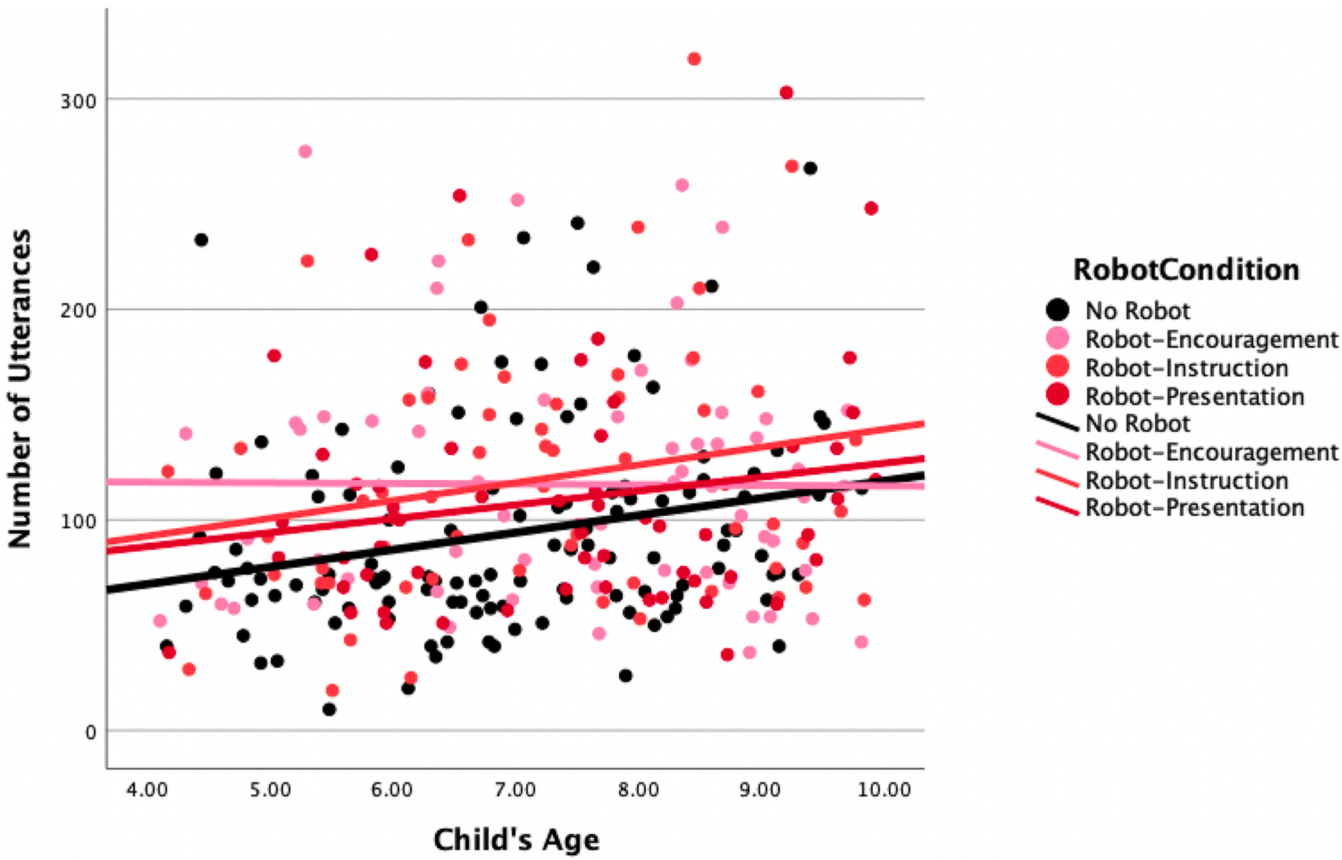

Fig. 3. Number of Utterances for Each Robot Condition.
The number of utterances each child said during the Narrative Pictures task for each Robot Condition. The black dots show the condition when the robot was absent and the colored dots show the conditions when the robot was present in different roles. The lines in matching colors show how the number of utterances changed as a function of the children's age. Children produced significantly more utterances as they got older (shown in the positive slopes of all the lines). Children also said fewer utterances when the robot was absent (shown by the black line being consistently lower than all the colored lines).

all of the lines overlap.

3.3. Discussion

These results showed that in a focused task that emphasized creating an extensive, self-contained speech sample, the presence of the social robot led children to talk more, but did not change the complexity of their speech, as measured by MLU. The extent of the robot's role, however, was not important: children talked on average more in all of the conditions when the robot was present relative to when it was absent. The lack of effect for the robot dosage, however, may reflect the fact that in practice, we did not differentiate the conditions very effectively. A review of the frequency of Jibo's interactions (when he was present) showed that he talked very little relative to the testing experimenter and that the amount he talked did not change across the conditions, ranging from just 3.7% to 4.7% of the utterances addressed to the child. The interactions with the child were primarily carried by the experimenter, regardless of condition. That said, despite the lack of differentiation among the conditions when Jibo was present, we note that the Robot Absent condition did lead to less talking by the child relative to all of the conditions with the Robot Present. Thus, even though children spent very little time overall interacting with Jibo in any condition, even the small dose of interaction had a measurable impact on the amount of child talking. These data support the idea that social robots encourage engagement from children (e.g. Kory-Westlund and Breazeal, 2015).

4. Example 2 for using the corpus: looking at potential effects of COVID-19 on children's school readiness

As noted above, approximately one third of the children in the corpus were run during the summer immediately following the COVID-19 lockdown. We were therefore able to investigate whether these children were behind in basic school skills relative to children of the same age who were tested the following year, after a comparatively normal year of schooling. We examined two reading-related tasks: the Alphabet task and the Reading Passages task. The Alphabet task asked children to identify each letter of the alphabet, identify a picture beginning with that letter, and to come up with their own word that also began with the letter. Alphabet knowledge is a critical pre-reading skill, and one that children are expected to master by the time they are in first grade (Ohio Department of Education, 2017). The Reading Passages task asked children to read short texts drawn from the set used in Cartledge et al. (2015) that were targeted towards children at a 2nd grade reading level.

We hypothesized that young children who had just missed a traditional year of early schooling – preschool or kindergarten – during the pandemic would be worse at the Alphabet task relative to children who had just completed those school years. We further hypothesized that older children who had just missed a traditional year of later schooling – 1st through 3rd grade – would be less willing and less able to complete the reading passages relative to children who had just completed those school years. We omitted all children who were tested during the school year from the analysis to avoid the potential impact of ongoing school instruction. We note that we did not get detailed information about children's schooling and thus we classified children based on their age,

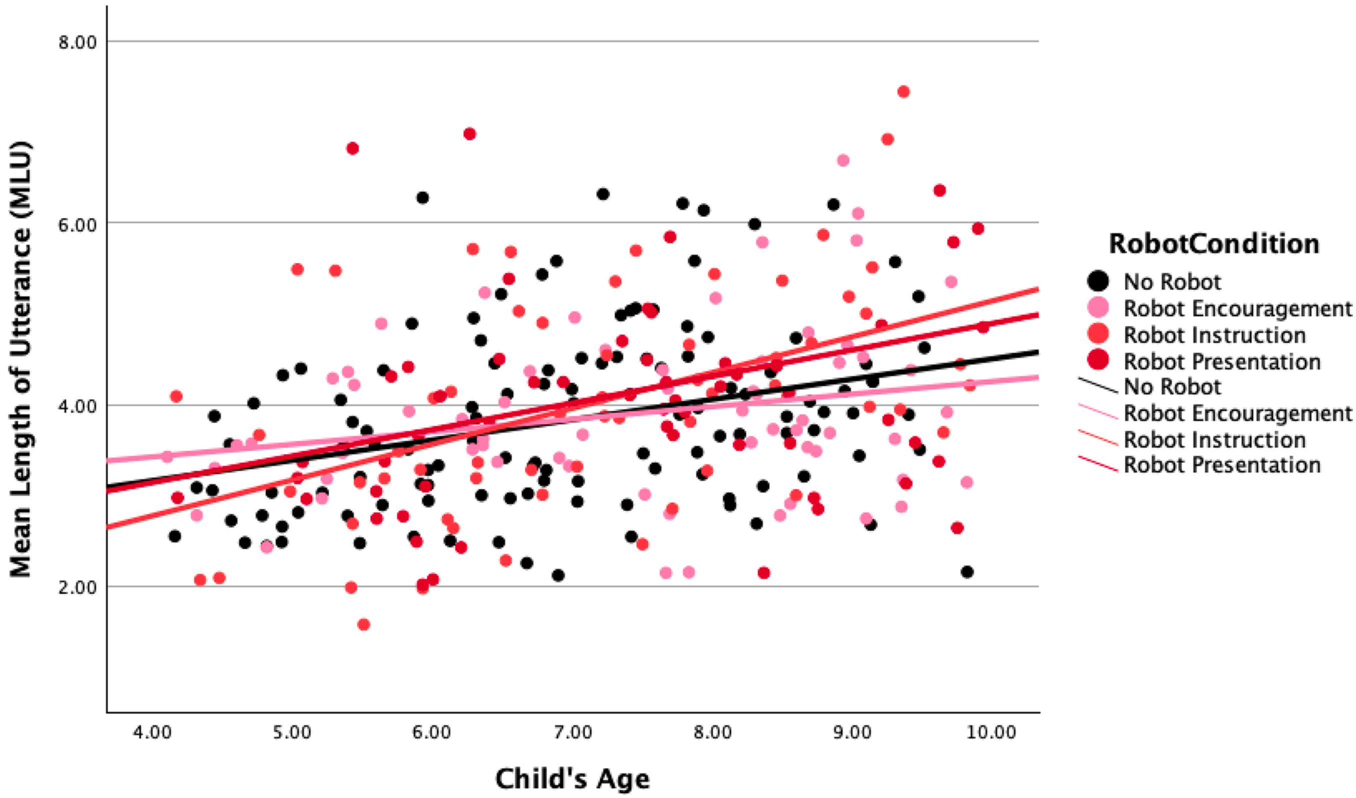

Fig. 4. Mean Length of Utterance (MLU) for Each Robot Condition.
The MLU each child said during the Narrative Pictures task for each Robot Condition. The black dots show the condition when the robot was absent and the colored dots show the conditions when the robot was present in different roles. The lines in matching colors show how the number of utterances changed as a function of the children's age. Children produced significantly longer utterances as they got older (shown in the positive slopes of all the lines). There was no significant effect of the robot condition on MLU.

the cut-off dates for local schools in Ohio, and the timing of the lockdown in the area.

4.1. Alphabet analysis

4.1.1. Participants

We focused on the 112 children in the database who were tested in the summertime and who should have just completed a year of either Kindergarten or Pre-K schooling. The children tested in the summer just post-pandemic (2021) most likely missed that schooling while the children tested in the following summer (2022) most likely just completed it. The Just-Post-Pandemic cohort consisted of 43 children, 24 of whom likely missed Kindergarten (*M* age = 6.34 years) and 18 of whom likely missed some kind of Pre-K schooling (*M* age = 5.07 years). The Year-Later cohort consisted of 70 children, 6 of whom likely missed Kindergarten and 64 (*M* age = 6.97 years) of whom likely missed some kind of Pre-K schooling (*M* age = 5.62 years).

4.1.2. Coding

We adopted strict criteria for success: children needed to provide their answers spontaneously without any hints or prompting beyond a basic command ("What letter/word is this? What other words start with that letter?"). However, as not all children were prompted for all three parts of the task on all letters, the proportion correct was calculated based on the number of opportunities children were offered to succeed. In some cases, a child was never asked to complete any task beyond letter identification and thus the N's for word identification and word generation are lower.

4.1.3. Results

We ran three separate ANOVA analyses with Cohort (Just-Post-Pandemic vs. Year-Later) and School Year (Kindergarten vs. Pre-K) as independent variables, and the separate dependent variables of percentage of letters identified, percentage of words identified, and percentage of letters for which children could generate new words. For the percentage of letters identified, we found no effect of either Cohort (F (1, 111) = 0.212, p = .65) or School Year (F (1, 111) = 0.198, p = .66), and no interaction between the two variables (F (3, 108) = 1.23, p = .27). These non-effects likely represent ceiling effects, as the success rate in this task was quite high, with children identifying 87.5% (or 23) letters on average overall. Looking at the slightly more challenging task of word identification, the ANOVA revealed an effect of Cohort (F (1, 109) = 4.05, p = .047, partial eta squared = 0.037) and School Year (F (1, 109) = 8.933, p = .003, partial eta squared = 0.078), but no interaction between the two variables (F (3, 106) = 0.232, p = .63). As can be seen in Fig. 5, children performed better on this task if they had just completed/missed Kindergarten relative to Pre-K, and also performed better if they were in the Year-After cohort than in the Just-Post-Pandemic cohort. Both of these results are in the hypothesized direction. On the even more challenging task of generating their own words, over half of the children did not complete the task. However, that decrement was approximately equal across the cohorts (55% attrition rate in Just-Post-Pandemic cohort and 59% attrition in the Year-Later cohort). As can be seen in Fig. 5, the lower N led to substantial variability in the scores, making the tests less sensitive overall. The ANOVA showed no significant effect of either Cohort (F (1, 47) = 2.69 p = .11) or School Year (F (1, 47) = 0.038, p = .85), and no interaction between the two variables (F (3, 44) = 0.12, p = .73).

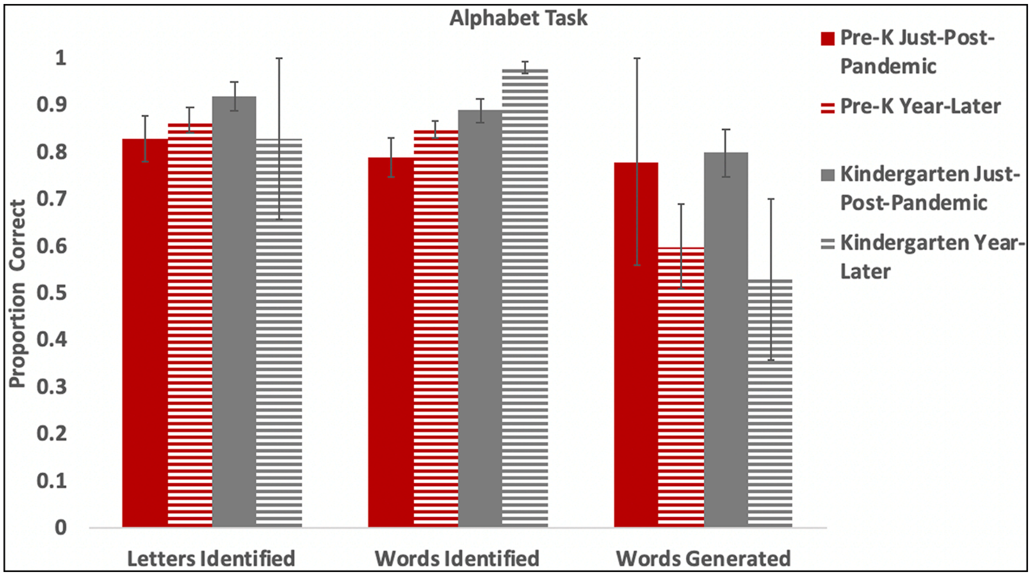

Fig. 5. Children's Performance on the Alphabet Task.
There were no significant differences of cohort or grade level on the number of letters identified or number of new words generated. However, for the number of words identified, there was a significant effect for both the Cohort and Grade Level of the child.

4.2. Reading passages

4.2.1. Participants

We focused on the 129 children in the OCSC who were tested in the summertime and who should have just completed a year of first, second, or third grade in school. The children tested in the summer just post-pandemic (2021) most likely missed that schooling, while the children tested in the following summer (2022) most likely just completed it. The Just-Post-Pandemic cohort consisted of 59 children, 23 of whom likely missed first grade (*M* age = 7.46 years), 21 of whom likely missed second grade (*M* age = 8.49 years), and 15 of whom likely missed third grade (*M* age = 9.45 years). The Year-Later cohort consisted of 70 children, 29 of whom likely just completed first grade (*M* age = 7.55 years), 25 of whom likely just completed second grade (*M* age = 8.48 years), and 16 of whom likely just completed third grade (*M* age = 9.31 years).

4.2.2. Coding

We coded the number of passages each child started, which can be seen as a measure of children's interest and enthusiasm for the task. We also coded the number of passages that each child completed, which can be seen as a measure of children's persistence.

We did not undertake the more complex coding of assessing the quality of children's reading of each passage so our measures do not capture the nuance of their reading ability.

4.2.3. Results

Our first analysis looked at the number of passages that children started to read. The range in values was 0 – 4, and by definition, only whole numbers were possible. Thus, we used two separate non-parametric Kruskal Wallis tests to determine if there were differences on reading enthusiasm as a function of the School Year that the child likely most recently missed/completed (First, Second, Third grade) or testing Cohort (Just-Post-Pandemic vs. Year-After). The results showed no difference for School Year (H (2) = 2.5, p = .29) on how many

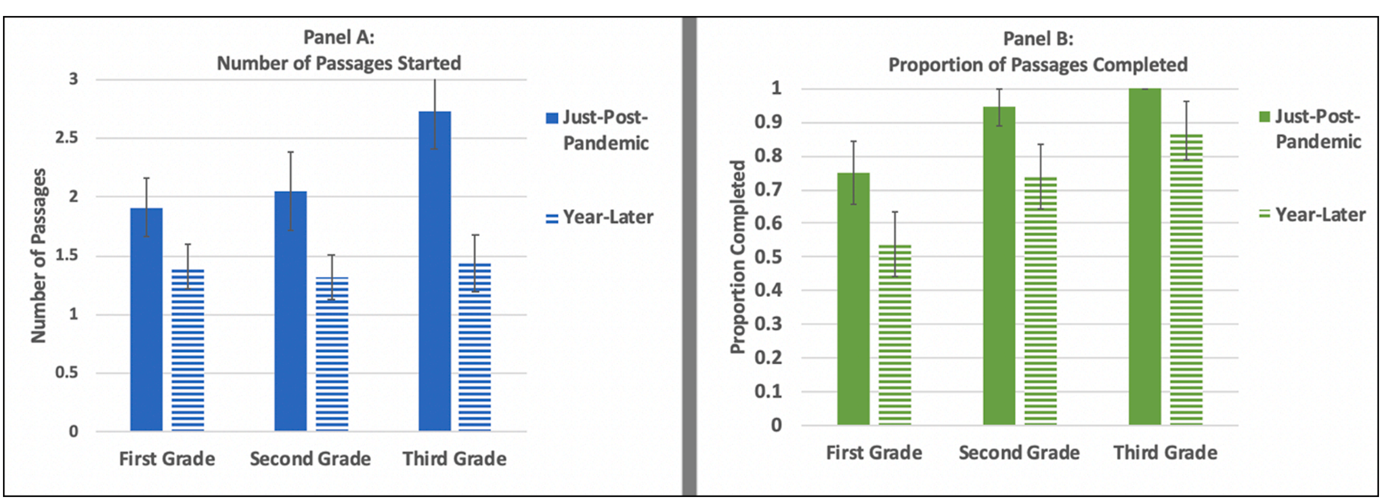

Fig. 6. Children's Performance on the Reading Passages Task.
Children who missed/completed higher grades completed significantly more passages. In addition, children who were tested in the year Just-Post-Pandemic started and completed significantly more passages than children in the Year-After.

passages children started to read. However, there was a significant effect of the Cohort, with children run in the Just-Post-Pandemic year starting more passages on average than those tested in the Year-After (2.2 passages vs. 1.4 passages: H (1) = 11.08, p < .001). These results run counter to our hypothesis and can be seen in panel A of Fig. 6.

To investigate whether the children finished the passages they started, we calculated the proportion of completed passages for each child. We chose this more continuous measure rather than just the number completed to adjust for the differences in the number of passages children started. Children who did not start any passages (N = 9) were not included in this analysis. We conducted an ANOVA with School Year and Cohort as independent variables, and the proportion of completed passages as the dependent variable. The results showed significant effects for both School Year (F (2, 114) = 6.03, p = .003, partial eta squared = 0.096) and Cohort (F (1, 114) = 6.32, p = .013, partial eta squared = 0.053), but no interaction between the two variables (F (2, 114) = 0.14, p = .87). As can be seen in panel B of Fig. 6, children completed more passages in the Just-Post-Pandemic year than they did in the Year-After, and they completed more passages as they got older (by Least Squared difference tests, children who had likely just finished first grade completed significantly (p < .05) fewer passages than children who had likely just missed/completed second or third grade although there was no statistical difference between the number of passages finished by the two older groups).

4.3. Discussion

Overall, these results suggest that for the children who participated in this corpus, the pandemic may not have disrupted their schooling strongly. For younger children, alphabet knowledge was mostly unaffected by the pandemic. While children who had just missed a year of Pre-K or Kindergarten did identify fewer words beginning with the target letters than children a year later, even that result should be seen in the context of overall quite high performance. For older children, the pandemic actually increased their performance in the reading task. In the year immediately following the lock-down, children were more willing to begin reading passages and also more successful at completing the ones they started relative to a year later.

We had hypothesized the reverse results. One possible reason for our findings is that our dependent measures were not sensitive enough. We used simple measures easily extracted from the corpus and they may simply not reflect children's real difficulties. Another severe limitation to this particular study is that we had no specific knowledge about children's educational background, and especially no knowledge of what kind of schooling they received during the pandemic. We classified children based on what was typical for children of that age in our location but we have no guarantees that children followed the typical patterns.

Nevertheless, given those typical patterns, we should perhaps not be quite so surprised that the children in our sample were not so adversely affected. Recent results (Fahle et al., 2024) have found that while the pandemic was detrimental to school skills for children from mid- and low-income school districts, children in high-income school districts did not see declines on their test scores; moreover, the state of Ohio (where our testing was done) was notable for widening the achievement gap in reading between high- and low-income families during the pandemic (see figure 11 in Fahle et al., 2024). The children in our corpus came from families with high educational levels – 75% of them had at least one parent with a college degree and many had multiple parents with college degrees or post-graduate degrees. Many of our children likely came from the demographic in Fahle et al. that showed pandemic resilience in their reading skills.

However, our results did not merely show that these children were unaffected by their pandemic year in this domain. For the Reading Passages task, they showed that children were doing significantly better immediately post-pandemic. We suspect that a second factor, "pandemic-coping", may be at play. Anecdotally, we noted that people who were willing to come to the museum in the Just-Post-Pandemic summer were extremely eager to interact with new people. Both parents and children may have brought a special enthusiasm for doing a task that allowed children to talk with people from outside of their own families. Relatedly, parents who were willing to bring their children into a very public space which was subject to a variety of pandemic-oriented regulations (e.g., masks, distancing, sanitizing routines) may have been the type of parents who were particularly motivated to ensure that their children received educational enrichment – something they may also have provided during the lockdown period. All pandemic regulations were over by the Year-After summer, and perhaps also the pandemic-coping bump of highly educated families may also have been over.

5. General discussion

The OCSC is a large cross-sectional corpus of children's speech which can be used to investigate a variety of questions in children's language development. We reported on two such questions that took advantage of distinctive features of the corpus and its collection. One distinctive feature of this corpus was that roughly 60% of the children were tested with the assistance of a social robot. In one study, we investigated the effects of having that robot present on children's speech during one specific task. We found that the presence of the robot increased children's willingness to talk, as measured by the number of utterances produced, but did not change the complexity of their speech, as measured by their MLU. This result suggests that social robots might be a benefit in situations when one wishes to encourage children to speak – for example as part of administering tests to assess language skills. A second distinctive feature of the corpus was that it was collected during the peri-pandemic time period. In a second study, we investigated whether children who were in the immediate aftermath of the pandemic lockdown would show worse pre-reading and reading skills relative to children tested a year later. We found that not only did the children show no detrimental effects in the first post-lockdown summer relative to the second, they in fact were significantly more likely to start and complete reading passages in that first post-lockdown summer. This result points to a demographic feature of our corpus, namely, that the children came from highly educated families. Their performance on the school readiness tasks is consistent with national trends in achievement gaps that were exacerbated during the pandemic.

The two studies reported here give a flavor of the kinds of things that could be examined in our corpus, and we identify several further potential uses in Section 1.2. However, even for the studies reported here, our investigations so far have only looked at a fraction of the richness of this data set. We have not, for example, examined whether the presence of the robot materially changes the kinds of things that children talk about: one might imagine that Jibo primes distinctive vocabulary words. We have also not examined the nature of the errors that children make when reading different passages. Nor have we looked at the different ways children describe their own reading ability. Further studies using this corpus are thus possible and very much encouraged. All of the audio data and associated transcripts have been posted on the Talkbank-CHILDES database for others who wish to make use of this corpus to investigate questions of their own.

Funding

This work was supported by the Ohio Department of Higher Education and by the National Science Foundation (#IIS-2008043, #SMA-2146474).

CRediT authorship contribution statement

Laura Wagner: Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Sharifa Alghowinhem:** Writing – review & editing, Resources. **Abeer Alwan:** Writing – review & editing, Methodology, Conceptualization. **Kristina Bowdrie:** Writing – review & editing, Supervision, Project administration, Investigation. **Cynthia Breazeal:** Writing – review & editing, Resources, Conceptualization. **Cynthia G. Clopper:** Writing – review & editing, Visualization, Resources, Funding acquisition, Formal analysis, Conceptualization. **Eric Fosler-Lussier:** Writing – review & editing, Resources, Methodology, Funding acquisition, Conceptualization. **Izabela A. Jamsek:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Data curation. **Devan Lander:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Investigation, Data curation. **Rajiv Ramnath:** Writing – review & editing, Methodology, Funding acquisition, Conceptualization. **Jory Ross:** Writing – review & editing, Supervision, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank all the students who assisted with the creation of this corpus: Sumurye Awani, Melike Baspinar, Nicki Besse, A'Niyah Brown, Lucas Butler, Hali Clark, Tema Cohen, Emma Cronin, Katlyn DeHart, Jane Farrell, Leslie Gutierriez Cervantes, Kara McClain, Madolyn McDonald, Keerat Sandhu, Kelly Schroeder, Abby Simon, Maya Soundappan, Riley Stalter, Melanie Subina, Claire Taylor, Chloe Thomas, Nicole Tonyan, and Kailin Zhang. We thank our artist, Rebecca Hinkelman, for her beautiful pictures. We thank Jan Weisenberger for facilitating the origin of this project. Finally, we thank the Center of Science and Industry and all the families who participated.

Data availability

The data used in this paper have been posted publicly at the Talkbank-CHILDES website: https://childes.talkbank.org/access/ENG-NA/OCSC.html. All transcripts and audio recordings are available there for public use.

References

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., Tanaka, F., 2018. Social robots for education: a review. Sci. Robot. 3, eaat5954. https://doi.org/10.1126/scirobotics.aat5954.
Berko, J., 1958. The child's learning of English morphology. Word 14 (2–3), 150–177.
Berman, R.A., Slobin, D.I., 1994. Relating Events in narrative: A crosslinguistic Developmental Study. Lawrence Erlbaum Associates, Hillsdale, NJ.
Brady, T.F., Konkle, T., Alvarez, G.A., Oliva, A., 2008. Visual long-term memory has a massive storage capacity for object details. Proc. Natl. Acad. Sci. U.S.A. 105 (38), 14325–14329.
Breazeal, C., Harris, P.L., DeSteno, D., Kory Westlund, J.M., Dickens, L., Jeong, S., 2016. Young children treat robots as informants. Top. Cogn. Sci. 8 (2), 481–491.
Cartledge, G., Bennett, J.G., Gallant, D.J., Ramnath, R., Keesey, S., 2015. Effects of culturally relevant materials on the reading performance of second-grade African Americans with reading/special education risk. Multiple Voices for Ethnically Diverse Exceptional Learners 15 (1), 22–43.
Chomsky, C., 1969. The Acquisition of Syntax in Children from 5 to 10. MIT Press.
Esfandbod, A., Rokhi, Z., Meghdari, A.F., Taheri, A., Soleymani, Z., Alemi, M., Karimi, M., 2023a. Fast mapping in word-learning: a case study on the humanoid social robots' impacts on Children's performance. Int. J. Child Comput. Interact. 38, 100614.
Esfandbod, A., Rokhi, Z., Meghdari, A.F., Taheri, A., Alemi, M., Karimi, M., 2023b. Utilizing an emotional robot capable of lip-syncing in robot-assisted speech therapy sessions for children with language disorders. Int. J. Soc. Robot. 15 (2), 165–183.
Eskenazi, M., Mostow, J., and Graff, D. (1997). The CMU Kids Corpus. Philadelphia: the Linguistics Data Consortium. doi:10.35111/b4v0-ff65.
Fahle, E., Kane, T.J., Readon, S.F., Staiger, D.O., 2024. The first year of pandemic recovery: a district-level analysis. Education Recovery Scorecard. https://educationrecoveryscorecard.org/wp-content/uploads/2024/01/ERS-Report-Final-1.31.pdf.
Gillam, R.B., Pearson, N.A., 2017. Test of Narrative Language – Second Edition, Pro-Ed.
Hicks, D., 1990. Kinds of texts: narrative genre skills among children from two communities. In: McCabe, A. (Ed.), Developing Narrative Structure. Erlbaum, Hillsdale, NJ.
Horst, J.S., Hout, M.C., 2016. The Novel Object and Unusual Name (NOUN) Database: a collection of novel images for use in experimental research. Behavioral Research Methods 48 (4), 1393–1409.
Kanero, J., Geçkin, V., Oranç, C., Mamus, E., Küntay, A.C., Göksun, T., 2018. Social robots for early language learning: current evidence and future directions. Child Dev. Perspect. 12 (3), 146–151.
Leonard, R.G. & Doddington, G.R. (1993). TIDIGITS. Philadelphia: Linguistics Data Consortium. doi:10.35111/72xz-6x59.
Kory-Westlund, J.K., Breazeal, C, 2015. The interplay of robot language level with children's language learning during storytelling. In: Proceedings of the 10th Annual ACM/IEEE International Conference on Human–Robot Interaction, pp. 65–66. https://doi.org/10.1145/2701973.2701989. March 2–5, 2015, Portland, OR.
Li, J., Deng, L., Gong, Y., Haeb-Umbach, R., 2014. An overview of noise-robust automatic speech recognition. IEEE/ACM. Trans. Audio Speech. Lang. Process. 22 (4), 745–777. https://doi.org/10.1109/TASLP.2014.2304637.
MacWhinney, B., 2000. The CHILDES project: Tools For Analyzing Talk. Erlbaum, Mahwah NJ.
Nippold, M.A., 2016. Later Language development: School-age children, adolescents, and Young Adults. PRO-ED, Inc, Austin, TX, p. 78757, 8700 Shoal Creek Boulevard-6897.
Ohio Department of Education, 2017. Ohio's Learning standards: English language Arts. https://education.ohio.gov/getattachment/Topics/Learning-in-Ohio/English-Language-Art/English-Language-Arts-Standards/ELA-Learning-Standards-2017.pdf.aspx?lang=en-US.
Peterson, Carole, McCabe, Allyssa, 1983. Developmental psycholinguistics: Three ways of Looking At a Child's Narrative. Plenum Press, New York.
Richard, G.J., Hanner, M.A., 2005. Language Processing Test 3: Elementary. LinguiSystems.
Shahab, M., Taheri, A., Mokhtari, M., AsemanRafat, A., Kermanshah, M., Shariati, A., Meghdari, A.F., 2024. Manufacture and development of Taban: a cute back-projected head social robot for educational purposes. Intell. Serv. Robot. 1–19.
Spitale, M., Silleresi, S., Cosentino, G., Panzeri, F., Garzotto, F., 2020. Whom would you like to talk with?" exploring conversational agents for children's linguistic assessment. In: Proceedings of the interaction design and children conference, pp. 262–272.
Wagner, L., Speer, S.R., Moore, L.C., McCullough, E.A., Ito, K., Clopper, C.G., Campbell-Kibler, K., 2015. Linguistics in a science museum: integrating research, teaching, and outreach at the Language Sciences Research Lab. Lang. Linguist. Compass. 9, 420–431.
Wik, P., Hjalmarsson, A., 2009. Embodied conversational agents in computer assisted language learning. Speech. Commun. 51 (10), 1024–1037.
Xu, Y., Wang, D., Collins, P., Lee, H., Warschauer, M., 2021. Same benefits, different communication patterns: comparing Children's reading with a conversational agent vs. a human partner. Comput. Educ. 161, 104059.
Zimmerman, I.L., Steiner, V.G., Pond, R.E., 2011. Preschool Language Scales–Fifth Edition (PLS-5). Pearson, Bloomington, MN.