Early Literacy and Oral Language Ties:
Extending the range of human-computer interface for early assessment

Alison L. Bailey,
Alejandra Martin,
Anahit Pogossian,
Marlen Quintero Pérez,
Department of Education,
UCLA
Gary Joseph Yeung,
Abeer Alwan,
&
Amber Afshan
Department of Engineering,
UCLA

Abstract

As a first step in evaluating the inclusion of extended oral discourse in the administration and scoring of language and literacy assessments using social robots, this pilot study reports on how well pre-kindergarten and kindergarten students performed with JIBO, a social robot, being developed to assist educators in the regular monitoring of children's language and literacy progress. A key finding was that a measure of contextualized vocabulary (i.e., produced during explanatory discourse) was related to letter recognition and a discrete measure of expressive vocabulary—important early literacy and literacy-related language skills, respectively. The study will help evaluate the overall educational value of automated language and literacy assessment and help in making adjustments to improve the assessment experiences of young students.

Objectives

This small-scale study explores connections between measures of early literacy and oral language as presented to students using the JIBO social robot that is being developed as part of a larger initiative to assist educators in the regular monitoring of students' progress in these domains. As part of the feasibility phase of this project, we were specifically interested in how pre-kindergarten and kindergarten students would perform in this human-computer interface context and whether the inclusion of extended oral discourse measures generated additional information about connections between students' language and literacy development for early childhood educators. This is the first step in evaluating the inclusion of oral explanations in the administration and scoring of the social robot's assessments.

The contribution of extended oral discourse skills and their connection to literacy measures and discrete oral language measures need to be better understood due to the technical challenges the inclusion of oral discourse will present for fully autonomous applications of JIBO, with speech recognition and deep learning models ultimately needed for automated evaluation of students' extended oral discourse. This initial pilot study can provide guidance on determining the value of taking on this challenge in the future. Our specific research questions as this stage were:

1) *What student performances are generated by the JIBO assessments and how are these influenced by key factors such as grade, program, language background, and gender?*
2) *How are the different oral language and literacy assessments correlated?*
3) *What role does extended oral discourse play in predicting early literacy and literacy-related language skills?*

Perspectives

Oral language and literacy ties

Approaches to early literacy development that view reading as a language-based skill have argued for the continuity between children's oral language skills and later literacy outcomes (Dickinson, McCabe, Anastasopoulos, Peisner-Feinberg, & Poe, 2003; Snow, 1991; Storch & Whitehurst, 2002). Oral language skills such as phonological awareness and vocabulary skills developed at the pre-kindergarten and kindergarten level can support the development of early literacy skills such as letter recognition and spelling skills (e.g., Paige, Rupley, Smith, Olinger, & Leslie 2018). These connections hold for emergent Spanish-English bilingual children and transfer across languages (e.g., Dickinson, McCabe, Clark-Chiarelli, Wolf, 2004). Increasingly, research suggests oral language skills at the discourse level (e.g., narratives, expository discourse) also support literacy development (e.g., Reese, Suggate, Long, & Schaughency, 2010) and experiences with discourse in interactive and engaging ways influence school readiness and later literacy (e.g., Leyva, & Smith, 2016). Collectively these findings suggest assessment of oral language skills for the prediction of literacy skills may need to include not only measures of discrete skills such as isolated words on a formal expressive vocabulary test but also language produced during extended discourse.

Use of social robotics for early assessment

Previous research for this pilot study has investigated the feasibility of implementing child-friendly robots for administering clinical and educational assessments with young children (Bailey et al., Oct., 2018; Yeung et al., April, 2019, Sept., 2019). Until recently, research suggested that speech recognition systems are not currently accurate enough for implementation with children (Kennedy et al., 2017; Yeung & Alwan, 2018). However, social robots such as JIBO (Spaulding et al., 2018) and Tega (Park et al., 2017) have seen some success when performing basic educational tasks with children such as acting like an engaged listener. Our research is extending this human-computer interface to be more useful to educators and their young students.

Methods and Data Sources

Participants

The 36 students recruited for this pilot study attended a university demonstration elementary school in the southwestern United States and are part of the larger, on-going project. The JIBO social robots (approx. 1 foot, 6 inches tall with a round display screen for a face and a sleek digital device body design, see Figure 1) were initially introduced to teachers and students as part of their science and technology inquiry-based curriculum. Approximately 40% of the school is enrolled in Spanish-English dual-language immersion classrooms. Table 1 provides the demographic data on the sample. Just eight of the (57%) of the pre-kindergarten students completed all the assessments. Thirteen (76%) of kindergartners completed all the assessments. Due to technical issues, fatigue or lack of willingness to interact with JIBO, nine students completed only a subset of the measures and five did not contribute data.

Procedures

The JIBO social robot was programmed to administer the 3rd Goldman Fristoe Test of Articulation (GFTA-3: Sounds in Sentences (story repetition task), Sounds in Words (picture naming task, Goldman & Fristoe, 2015), alphabet and number naming (letter and number recognition tasks) and prompts to elicit extended oral discourse in the form of two open-ended oral explanations of a personal routine (teeth cleaning) and an early science routine (mixing colors). Instructions, prompts, and periodic friendly interactions by JIBO were pre-recorded by a female researcher, with recordings pitch-shifted to sound like a young child's voice.

Each student individually interacted with JIBO and were video and sound recorded. One researcher sat next to the child as an "instructor" and interacted with JIBO along with the child. The other researcher sat behind JIBO as an "operator" controlling the display of items on JIBO's "face" with a computer (the larger project is in the process of also using the collected child speech samples to develop an autonomous version of JIBO using automatic speech recognition). JIBO first introduced itself and asked warm-up questions (e.g., "What is your name?", "What is your favorite color?") to put the child at ease. At the end of the session, JIBO would thank the child for playing, say goodbye, and laugh in response to being petted by the researcher and/or child. Sessions were approximately 30 minutes.

Letter recognition

Recognition of letters of the alphabet was measured based on the GFTA-3 letters subtest by displaying randomly generated sequences of letters on JIBO's screen with the audio prompt "What letter is this?" The score is the proportion of the 26 letters that a child produced correctly.

Expressive vocabulary

Children's expressive vocabulary was measured based on the GFTA-3 Sounds in Words. Children were prompted with a picture on the screen and an audio prompt of "What is this?" The score is the proportion of 58 discrete words correctly produced.

Explanatory oral discourse

Extended oral discourse, related to a personal context and an academic context, was elicited by showing the participants an image of a child cleaning his/her teeth and then audio prompts to explain how and why they cleaned their teeth (personal routine). This was followed by an image of a teacher and students mixing colors in the classroom and audio prompts to explain how and why they mixed to obtain different colors (academic task). In both cases, children were asked to explain to a naïve, hypothetical friend to elicit maximum explicitness. Tasks were evaluated for their word sophistication in context (increasing amount/variety of topic-related vocabulary), sentence sophistication (increasing syntactic complexity) and coherence/cohesion (increasing use of logical organization/discourse markers such as transition words) following an established protocol placing performances on language learning progressions (Bailey & Heritage, 2014). The three language features were placed at 0 (language feature is not evident), 1 (early emergent), 2 (emergent), 3 (developing) and 4 (controlled). Proportion of agreements between two raters ranged from .84 to .87 (personal routine) and from .81 to .84 (academic task). Cohen's kappa that takes account of chance agreements ranged from .75 to .80 (personal routine) and .65 to .73 (academic task), and are substantial (Landis & Koch, 1977). Disagreements were resolved by consensus.

Results

Student performances overall and by grade, program, language background, and gender

Table 2 shows the students' performances on the language and literacy measures overall. Grade was a significant factor for the formal test of expressive vocabulary ($t(21) = -2.972$, $p < .01$. Hedge's effect size (g) for unequal sample sizes was 1.27, considered large), and approached significance for letter recognition. In each case, kindergarten students, on average, outperformed the pre-kindergarten students. Interestingly there were no significant differences by grade on the language features at the word-, sentence- or discourse-levels of the two extended oral discourse tasks. There were significant differences by program classroom and language background in terms of whether English was spoken in the home (alone or in combination with another language, most frequently Spanish). The English medium classrooms, on average scored higher on vocabulary sophistication in personal routine explanations, the expressive vocabulary test and letter naming. Students exposed to English at home had higher ratings for all three language features in explaining a personal routine, as well as for the expressive vocabulary test and letter recognition (see Table 3 for test of means). Hedge's effect size (*g*) for unequal sample sizes ranged from .53 to 1.60, considered medium to large. However, neither program classroom nor English exposure at home distinguished between students on the three language features of the academic-themed oral discourse task that required students to explain color mixing, possibly because this was a challenging verbal task for most students. We found no significant differences in performances by gender.

Correlations between measures

Table 4 presents the correlations between the language and literacy measures. There was a high number of significant positive correlations among the measures. The vocabulary features of the two explanation tasks were correlated with each other and with the expressive vocabulary test. The vocabulary features of the personal routine explanation and the expressive vocabulary test were additionally correlated with letter recognition. The three language features of each of the extended oral discourse tasks were also largely correlated both within and across the two tasks, suggesting students who do well on one language feature in their explanations do as well on the other features.

Predictors of early literacy and literacy-related discrete oral language skills

Expressive vocabulary significantly predicted letter recognition ($\beta = .73$, $p<.01$), even controlling for grade ($\beta = -.02$, $p=.93$). The two predictors explained 52% of the variance ($R2=.52$, $F(2,20)=9.70$, $p<.001$). Controlling for exposure to English in the home ($\beta = .19$, $p=.41$), expressive vocabulary ($\beta = .61$, $p=.01$) and exposure explained 55% of the variance in letter recognition ($R2=.55$, $F(2,17)=9.27$, $p<.01$). Vocabulary sophistication in the personal routine explanation significantly predicted letter recognition, but not after controlling for expressive vocabulary or English exposure separately. Vocabulary sophistication and coherence/cohesion in both explanation tasks predicted expressive vocabulary, but not after controlling for English exposure.

These preliminary findings will be supplemented by analyses of an additional 85 students assessed with JIBO and still undergoing rating. Multivariate models will be built that can explore whether discourse measures predict alphabetic knowledge such as letter recognition beyond what is explained by the discrete skills expressive vocabulary test or whether they influence early literacy more indirectly through their impact on the formal measure of expressive vocabulary.

Scholarly Significance

This pilot study with pre-kindergarten and kindergarten students has promising implications for evaluating early childhood language and literacy development in a human-robot interface context. While previous studies have found that robots in education settings provide interactive language experiences (Sugimoto, 2011; Chambers et al., 2008; Bers, 2010; Chang et al., 2010; Young et al., 2010), such as teaching new words to children successfully (Kanero et. al., 2018), as well as assisting children to produce oral discourse (Westlund & Breazeal, 2015; Hyun, Kim, Jang, & Park, 2008), we need to better understand the predictors of early literacy, including what in turn predicts the early predictors (e.g., especially the discourse-embedded, vocabulary-building experiences that appear to be significant in these preliminary analyses). The contextualized vocabulary measure during oral discourse may help to distinguish the vocabulary skills that are needed in children's language and early literacy experiences.

This information will help evaluate the overall educational value of automated systems, including social robots, to children's literacy assessment and make adjustments to improve the experiences of students. Modifications include (1) improvement of speech recognition of young children, which is still elusive in robot-human interactions (Kennedy, et al. 2017), and (2) assessment of letters, words, discourse, and social interaction knowledge. Future work of the larger project aims to evaluate the efficacy of integrating such automated systems in classrooms.

References

Bailey, A. L., & Heritage, M. (2014). The role of language learning progressions in improved instruction and assessment of English language learners. *TESOL Quarterly*, *48*(3), 480-506.

Bailey, A.L. Yeung, G., Quintero Perez, M., Martin, A., Afshan, A., & Alwan, A., (October, 2018). *MEET JIBO! Demonstration of a Personalized Learning Companion Robot Designed for Speech and Language Assessment in Early Childhood Education Settings.* CRESST/UCLA Annual Conference, Los Angeles, CA.

Bers, M. U. (2010). The TangibleK robotics program: Applied computational thinking for young children. *Early Childhood Research & Practice, 12*(2), n2.

Chambers, J. M., Carbonaro, M., & Murray, H. (2008). Developing conceptual understanding of mechanical advantage through the use of Lego robotic technology. *Australasian Journal of Educational Technology, 24*(4), 384-401.

Chang, C. W., Lee, J. H., Chao, P. Y., Wang, C. Y., & Chen, G. D. (2010). Exploring the possibility of using humanoid robots as instructional tools for teaching a second language in primary school. *Educational Technology & Society, 13*(2), 13–24.

Dickinson, D.K., McCabe, A., Anastasopoulos, L., Peisner-Feinberg, E.S., & Poe, M.D. (2003). The comprehensive language approach to early literacy: The interrelationships among vocabulary, phonological sensitivity, and print knowledge among preschool-aged children. *Journal of Educational Psychology, 95* (3), 465-481.

Dickinson, D.K., McCabe, A., Clark-Chiarelli, N., Wolf, A. (2004). Cross-language transfer of phonological awareness in low-income Spanish and English bilingual preschool children. *Applied Psycholinguistics*, *25,* 323–347.

Goldman, R., & Fristoe, M. (2015). *GFTA-3: Goldman Fristoe 3 Test of Articulation.* PsychCorp.

Hyun, E. J., Kim, S. Y., Jang, S., & Park, S. (2008). Comparative study of effects of language instruction program using intelligence robot and multimedia on linguistic ability of young children. In Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication (pp. 187–192). Munich, Germany: HRI.

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 Development Perspectives, 12*(3), 146-151.

Kennedy, J., Lemaignan, S., Montassier, C., Lavalade, P., Irfan, B., Papadopoulos, F., Senft, E., & Belpaeme, T. (2017). Child speech recognition in human-robot interaction: Evaluations and recommendations. In *Proc. of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 82-90.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159-174.

Leyva, D., & Smith, M. (2016). Beyond book reading: Narrative participation styles in family reminiscing predict children's print knowledge in low-income Chilean families. *Early Childhood Research Quarterly*, *37*(4), 175-185.

Paige, David D., William H. Rupley, Grant S. Smith, Crystal Olinger, and Mary Leslie. (2018). Acquisition of letter naming knowledge, phonological awareness, and spelling knowledge of kindergarten children at risk for learning to read. *Child Development Research*, pp.1-10.

Park, H.W., Gelsomini, M., Lee, J.J., & Breazeal, C. (2017). Telling stories to robots: The effect of backchanneling on a child's storytelling. In *Proc. of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*.

Reese, E., Suggate, S., Long, J., & Schaughency, E. (2010). Children's oral narrative and reading skills in the first 3 years of reading instruction. *Reading and Writing, 23*(6), 627-644.

Snow, C. E. (1991). The theoretical basis for relationships between language and literacy in development. *Journal of Research in Childhood Education, 6*(1)*,* 5- 10.

Spaulding, S., Chen, H., Ali, S., Kulinski, M., & Breazeal, C. (2018). A social robot system for modeling children's word pronunciation. In *Proc. of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*.

Storch, S. A., & Whitehurst, G. J. (2002) Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*(6)*,* 934-947.

Sugimoto, M. (2011). A Mobile mixed-reality environment for children's storytelling using a handheld projector and a robot. *IEEE Trans Learning Technologies, 4*(3), 249-260.

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). Portland, OR: ACM.

Yeung, G., & Alwan, A. (2018). On the difficulties of automatic speech recognition for kindergarten-aged children. In *Proc. of INTERSPEECH*, 1661-1665.

Yeung, G., Bailey, A.L., Afshan, A., Quintero Perez, M., Martin, A., Alwan, A., et al. (April, 2019). *Toward the Development of Personalized Learning Companion Robots for Early Speech and Language Assessment.* Poster presented at the annual meeting of the American Educational Research Association, Annual Conference, Toronto, Canada.

Yeung, G., Bailey, A.L., Afshan, A., Tinkler, M., Quintero Perez, M., Martin, A., A. Pogossian, Spaulding, S., Park., H. W., Muco, M., Alwan, A. & Breazeal, C. (Sept., 2019). *A robotic interface for the administration of language, literacy, and speech pathology assessments for children.* SLaTE 2019: 8th ISCA Workshop on Speech and Language Technology in Education, Graz, Austria.

Young, S. S. C., Wang, Y. H., & Jang, J. S. R. (2010). Exploring perceptions of integrating tangible learning companions in learning English conversation. *British Journal of Educational Technology, 41*(5), E78-E83.

Table 1. *Demographic data for participants*

Demographic Variable	Frequency	Percentage
Grade		
Pre-K	17	47.2
Kinder	19	52.8
Total	36	100
Program Classroom*		
EMI	21	58.3
DLI	15	41.7
Total	36	100
Gender		
Boy	15	41.7
Girl	21	58.3
Total	36	100
Home Language Exposure		
English monolingual	7	19.4
Spanish monolingual	6	16.7
Bilingual Eng.-Span.	10	27.8
Other Bi/multilingual	7	19.4

Other Monolingual	2	5.6
*Total***	32	88.9

Notes: *EMI: English Medium Instruction; DLI: Spanish-English Dual-Language Immersion
** 4 students with missing data for home language exposure

Table 2. *Descriptive statistics of performances overall*

Measure	n	Mean (SD)	Min	Max
Personal Routine Explanation Vocabulary	31	2.65 (1.08)	0	4
Personal Routine Explanation Sentence Structure	31	2.58 (.56)	2	4
Personal Routine Explanation Coherence/Cohesion	31	1.71 (.86)	0	3
Academic Explanation Vocabulary	31	1.94 (1.03)	0	4
Academic Explanation Sentence Structure	31	2.55 (.68)	1	4
Academic Explanation Coherence/Cohesion	31	1.61 (1.02)	0	4
Expressive Vocabulary (GFTA-3)	23	.93 (.05)	.83	1.0
Letter Recognition (GFTA-3)	21	.90 (.17)	.32	1.0

Table 3. *Mean performances by home language background*

Measure	English spoken at home (alone or in combination with another language)		English never spoken at home		
	n	Mean (SD)	n.	Mean (SD)	t(df), p-value
Personal Routine Explanation Vocabulary	18	2.83 (.86)	9	1.89 (1.27)	t(25)=-2.30, p=.03
Personal Routine Explanation Sentence Structure	18	2.72 (.58)	9	2.22 (.44)	t(25)=-2.29, p=.031
Personal Routine Explanation Coherence/Cohesion	18	2.00 (.84)	9	1.22 (1.30)	t(25)=-1.88, p=.071
Academic Explanation Vocabulary	18	2.00 (.69)	9	1.55 (1.24)	t(25)=-1.21, p=.237
Academic Explanation Sentence Structure	18	2.50 (.51)	9	2.55 (1.01)	t(25)= 1.91, p=.850
Academic Explanation Coherence/Cohesion	18	1.72 (.67)	9	1.11 (1.36)	t(25)=-1.58, p=.127
Expressive Vocabulary (GFTA-3)	13	.95 (.03)	6	.89 (.05)	t(17)=-3.40, p=.003

Expressive Vocabulary (GFTA-3)	12 .95 (.06)	6 .75 (.25)	t(16)=-2.77, p=.014

Table 4. *Correlations between oral language and literacy assessments*

Measure	1	2	3	4	5	6	7	8
1.Personal Routine Vocabulary								
2. Personal Routine Sentences	.349							
3. Personal Routine Coherence/Cohesion	.576**	.600**						
4. Academic Task Vocabulary	.457**	.411*	.606**					
5. Academic Task Sentences	.138	.274	.391*	.484**				
6. Academic Task Coherence/Cohesion	.384*	.518**	.602**	.640**	.559**			
7. Expressive Vocabulary %	.538**	.125	.451*	.433*	.079	.403		
8. Letter Recognition %	.458*	-.043	.320	.089	-.023	.160	.720**	

Note: *Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Figure 1. *Example assessment setting of a child and JIBO interacting, along with an "instructor" to the side of the child and an "operator" behind JIBO.*