Using Formative Conversation-based Assessments to Support Students' English Language Development

Alexis A. Lopez, Christine Luce, Diego Zapata-Rivera and Carol Forsyth

***Abstract*—In this article, we discuss the use of prototype formative conversation-based assessments designed to measure English learners' language skills. Conversation-based assessments are technology enhanced assessment systems that simulate interactions between a test taker and one or more virtual agents. We discuss preliminary findings from two studies exploring the use of conversation-based assessments to gather evidence of students' English language skills. The findings suggest that conversation-based assessments have the potential to provide useful information to teachers and students about the students' language skills and can be used to enhance and support English language development.**

***Index Terms*—Conversation-based assessments, English language assessments, English learners, English language development**

I. Introduction

Recently, there has been an increased interest in using a variety of innovative assessment approaches to support and enhance language development. Assessments are an integral part of learning as they help teachers obtain information about where the language learners are along their personal learning trajectory [1]. Assessments that have the potential to enhance learning involve the collection and interpretation of evidence about performance so that judgments can be made about further language development [2]. Thus, it is critical to develop language assessments that provide teachers and students with valuable information about the learners' language learning process.

We have leveraged advances in natural language processing in the field of intelligent tutoring systems [3] and applied Evidence-Centered Design principles [4] in the creation of innovative Conversation-Based Assessment (CBA) prototypes to gather evidence of English language proficiency through carefully planned conversations [5], [6], [7]. Conversations with computer agents can be used to gather evidence of student's English language proficiency that may be difficult to obtain using traditional assessment approaches.

The rest of the paper is organized as follows. First, Section II presents background of conversation-based systems, CBAs, and the structure of CBAs. Next, Section III describes the design of two formative CBA prototypes that were developed to assess young English learners. Then, Section IV describes two studies that were carried out to investigate and evaluate the two prototypes and Section V summarizes preliminary findings. Finally, Section VI presents conclusions and future research directions.

Alexis A. Lopez is with Educational Testing Service (e-mail: alopez@ets.org).

Christine Luce is with Educational Testing Service (e-mail: cluce@ets.org).

Diego Zapata-Rivera is with Educational Testing Service (e-mail: dzapata@ets.org).

Carol Forsyth is with Educational Testing Service (e-mail: cforsyth@ets.org).

II. Framework

A. Conversation-based Systems

Conversation-based systems can simulate conversations between a student and one or more virtual agents [8]. Conversation-based tasks are structured to provide opportunities for individuals to demonstrate their knowledge, skills, and abilities (KSAs), as well as to scaffold learning and provide useful feedback to students. Through the natural flow of the conversation, the virtual agents can repeat or rephrase their questions, ask follow-up questions, provide feedback on the quality of the responses, and provide scaffolds to elicit additional or missing information [5, 9]. Technology-enhanced conversation systems with virtual agents have been successfully used in the past as part of tutoring systems to facilitate learning for instructional purposes [3], [10], [11], [12], [13].

B. Conversation-based Assessments

Technology-enhanced conversation systems have also been used for assessment purposes. Specifically, conversation-based systems have been used to measure argumentation skills [14], science inquiry [15], and English language proficiency skills [5], [6]. These automated conversations involve dialogues between the test-taker and one or more virtual agents [7]. The main features of CBAs are the following:

- Students interact with virtual agents to create engaging and interactive environments. The interaction with the virtual agents allows for elicitation of more detailed information about the students' KSAs.
- Additional prompting is provided to redirect students to provide a different response or to elicit more information.
- Task-level feedback is provided in real-time immediately after the student answers a question.
- Scaffolds are provided when students are unable to answer a question; the type of support depends on how the student responds.

C. CBA Structure

In a CBA, each conversation starts with a main question and attempts to understand the students' responses during the

interaction with the virtual agents using computer natural language processing and automated speech recognition. The students' responses for each question are compared against a pre-defined expected answer with two natural language processing algorithms that have proven to be successful in the past, regular expressions and latent semantic algorithm [16]. Regular expressions are a way of capturing a key word or expression [17] and latent semantic analysis is a geometric matching algorithm that is designed to capture the contextual meaning of the phrase [18]. The conversations adapt to what students say/type in order to measure what students know about a particular construct. Once the system interprets the student's response, it will send the student through a particular conversational path (see Figure 1).

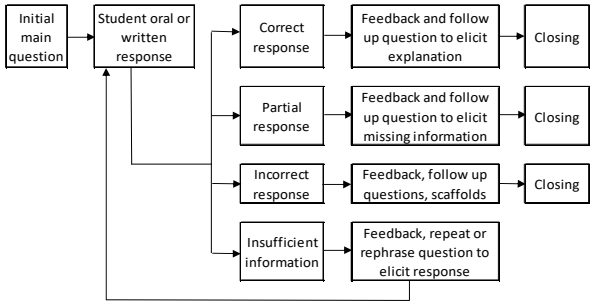

Figure 1. Conversation Paths in Formative CBAs.

If the student's response matches the predetermined correct response, it is interpreted as correct and it is sent through the correct conversation path. The virtual agents provide feedback on the quality of the response and ask follow-up questions in order to allow the student to explain his/her responses. If the student's response partially matches the predetermined correct response, it is sent through the partial conversation path and the virtual agents provide feedback and ask follow-up questions to elicit more information. If the student's response matches the predetermined incorrect response, it is sent through the incorrect conversation path and the virtual agents provide feedback, provide scaffolds, and ask follow-up questions to elicit a different response. Sometimes the system is unable to make a decision because it cannot interpret what the student says/types. This happens when the student says/types something off topic (not relevant to the topic of the question), requests clarification (e.g., I don't know, what did you say?), or does not provide a response. In this case, the response is sent through the insufficient conversation path and the system will prompt the student and give them additional opportunities to respond (e.g., by rephrasing the question or asking the student to elaborate).

III. THE ASSESSMENTS

We developed two formative CBA tasks to evaluate how they can be used to support young English learners in developing their English language skills. Prototype 1 is intended to measure both the English language proficiency skills and math content knowledge of middle school English learners and identify their language and mathematics difficulties for instruction and intervention. Some students might have low English proficiency but have high math skills and vice versa. Prototype 2 is intended to provide information about how to support the student's language development so they can learn mathematics. The prototype simulates a small group activity in which the student interacts with three virtual agents, a teacher and two peers. Figure 2 shows the interface for Prototype 1. This prototype includes seven conversations, each assessing a different facet (or aspect) of a broader construct of language (e.g., following oral directions, understanding information in texts and tables, and getting information from math problems and tables) and mathematics (e.g., ratios and unit rate). Students participate in the conversations by typing their responses in the computer interface.

Figure 2. Prototype 1 Interface.

Prototype 2 was designed to measure the English proficiency of students approximately 8-11 years old learning English in countries where English is not used as a first language. This CBA targets different listening, speaking and reading constructs: understanding oral and written directions, answering simple questions, and identifying and summarizing key ideas. Prototype 2 is divided into seven conversations which simulate interactions between the student and four virtual agents to learn about the weather in the United States. In this CBA, the student interacts with a teacher, a school librarian and two peers in three different settings: a classroom, a school library and a science museum. Students participate in the conversations by recording their spoken response using the computer interface. Figure 3 shows the interface for this prototype.

Figure 3. Prototype 2 Interface.

IV. THE STUDIES

We conducted two studies to evaluate whether the intended KSAs were elicited by these CBAs. In the first study, we

administered Prototype 1 to 82 middle school English learners from two schools in the United States. The main goals of this study were to examine how middle school English learners interacted with the CBA and to compare the students' performance on the CBA to their performance on more traditional measures. In particular, we addressed these research questions: 1) How reliable was the CBA system in interpreting student responses? 2) How do students perform on traditional measures compared to their performance on Prototype 1? 3) Can Prototype 1 be used to evaluate students' strengths and weaknesses in both language and mathematics? And 4) What perceptions do students have of Prototype 1? The CBA and the more traditional measures for the study were administered on a desktop computer using Internet Explorer. Students began the study by completing the CBA. Then, they completed a survey related to user perceptions. Following the survey were three abbreviated independent measures to collect data on students' skills related to mathematics (proportional reasoning), listening comprehension, and reading comprehension. The data collected for this study included the students' responses and scores for each of the questions in the CBA, scores on the math measure, scores on the listening comprehension measure, scores on the reading comprehension measure, and responses on the survey.

We conducted another study using Prototype 2 to examine the students' perceptions in regard to their experience completing this prototype and to explore the use of different feedback types to elicit missing or additional information. The main research questions were: 1) What perceptions do students have of Prototype 2? And 2) Do students respond differently to different feedback types? We included three types of feedback in the conversations to create three different versions of the same task. The corrective feedback was provided only when the response was partial or incorrect. In the first version, the virtual agents provided feedback by asking a specific question to elicit more information (focused/cued feedback). The feedback in the second version consisted of the virtual agents indirectly rephrasing the students incorrect or incomplete responses to provide the correct response (recast feedback). In the third version, the virtual agents provided feedback by telling the students their response was incorrect and prompting them to change it (try again feedback). Table 1 illustrates the three feedback types. In this example, students are expected to say that they are going to learn about the weather in the United States.

Table I
FEEDBACK TYPES IN PROTOTYPE 2

Type	Sample feedback
Focused	Lisa: Yes, we are learning about the weather. Ron: In what country are we learning about the weather?
Recast	Lisa: Yeah, we are learning about the weather in the United States. Ron: Wait, Lisa. Let the new student tell me. What did the teacher say we are going to learn today?
Try again	Lisa: Yes, but it's not about any weather. You need to tell Ron more. Ron: What did the teacher say we are going to learn today?

The virtual agents are trying to help the student provide the missing information (i.e., the United States). The main questions and the expected responses were the same for all versions. Eighty English learners from an elementary school in South Korea participated in this study. Students were randomly assigned to one of the three feedback versions. The data collected for this study included student responses to Prototype 2, background information questionnaire, and post-study survey.

V. PRELIMINARY FINDINGS

The results from Study 1 showed that the CBA system was able to score responses in Prototype 1 reliably. The agreement between the scores assigned by the CBA system in Prototype 1 and the consensus scores assigned by two human raters was high (94% agreement). We also found that the language and math scores on Prototype 1 had a significant positive correlation with the external measures (r=.52 language, r=.56 math). An ANOVA revealed the Prototype 1 language score accounted for a significant portion of variance on the external language measures (21.6% of the variance in listening and 28.8% in reading) and the Prototype 1 math score accounted for a significant portion of variance on the external math measure (33.3% of the variance). Prototype 1 also provided evidence of the students' strengths and weaknesses, both in language and mathematics. The student responses provided information about the students' ability to understand a word problem, to find information in a table, to understand the meaning of ratio, and to understand how to write a ratio. The majority of the students were able to answer the virtual agents' questions (87%), and used the feedback and support provided by the virtual agents to engage in the task and to correct their responses or to articulate missing information. We also found that students had positive perceptions about Prototype 1 and about interacting with the virtual agents (82.1%).

Similarly to what we found in the first study, the majority of the participants in the second study also had a positive perception of Prototype 2 (75%). Some students reported that it was interesting to interact with the virtual agents because they could practice speaking in English and they could learn new words (37.5%). We also found that some students were able to provide more information or change their responses based on the feedback they got from the virtual agents (51.3%). Although more students were able to change their responses with the *try again* feedback (18 students) than with the other two types of feedback (14 students with the *recast* feedback and nine students with the *focused/cued* feedback), only four of them were able to correct their responses or provide the missing information. Six of the nine students who received the *focused/cued* feedback and eight of the 14 students who received the *recast* feedback were able to either correct their responses or provide the missing information.

VI. DISCUSSION

The main contribution of our studies is that we provide empirical evidence that formative CBAs have the potential to create engaging and interactive environments to measure the language skills of young English learners [5]. Engagement is especially useful for promoting language learning [19]. In both studies, we found that the majority of the students enjoyed interacting with the virtual agents and liked that the agents supported them in completing the activities. Students in the second study reported that the CBA gave them an opportunity to practice their English skills by interacting with the virtual agents. The students' level of engagement while completing the CBAs was reflected by their willingness to spend time and effort to complete all the tasks.

The use of additional prompting and scaffolds allow formative CBAs to collect rich evidence of the students' progress and achievement. In our two studies, we found that the responses to both the initial and follow-up questions contributed to assessing the students' language skills. The interactions with the virtual agents allowed for elicitation of more detailed evidence of students' English language skills [5]. Likewise, the additional prompting helped not only to progress the conversations but also provided useful information about the students' strengths and weaknesses in their language development. Although additional research should be conducted to substantiate the CBA system-derived diagnostic information, it could potentially inform teachers and learners about areas in need of improvement.

In formative CBAs, students are able to interact with the virtual agents to complete a specific goal. In this interaction, students receive timely and frequent feedback that varies according to how they respond to each question. Students could use this feedback to monitor their progress toward achieving their goal [20]. The effectiveness of the feedback is essential to successful language development [21]. Thus, it is important to examine different ways to provide effective feedback to English learners.

Planned future work includes more empirical evidence to validate the use of CBAs to measure the language skills of young learners, including using CBAs for diagnostic purposes. Other work includes exploring the use of formative CBAs to assess different language skills. Equally important is examining and improving the accuracy of CBAs to interpret English learners' spoken and written responses.

REFERENCES

[1] P. Rea-Dickins, "Understanding teachers as agents of assessment," Language Testing, vol. 21, pp. 249–58, 2004.

[2] J. E. Purpura, Assessing grammar. Cambridge, UK: Cambridge University Press, 2004.

[3] A. C. Graesser, N. Person, and D. Harter, "Tutoring Research Group: Teaching tactics and dialog in AutoTutor," International Journal of Artificial Intelligence in Education, vol. 12, pp. 257–279, 2001.

[4] R. J. Mislevy, and G. D. Haertel, "Implications of evidence-centered design for educational testing," Educational Measurement: Issues and Practice, vol. 25, pp. 6–20, 2006.

[5] Y. So, D. Zapata-Rivera, Y. Cho, C. Luce, and L. Battistini, "Using trialogues to measure English language skills," Journal of Educational Technology & Society, vol. 18, pp. 21-32, 2015.

[6] K. Evanini, Y. So, J. Tao, D. Zapata-Rivera, C. Luce, L. Battistini, and X. Wang, "Performance of a trialogue-based prototype system for English language assessment for young learners," in Proceedings of the Interspeech Workshop on Child Computer, Interaction (WOCCI 2014), Singapore, 2014.

[7] D. Zapata-Rivera, T. Jackson, and I. R. Katz, "Authoring conversation-based assessment scenarios," in Design recommendations for intelligent tutoring systems Volume 3: Authoring tools and expert modeling techniques, R. A. Sottilare, A. C. Graesser, X. Hu and K. Brawner, Eds., Orlando, FL: U.S. Army Research Laboratory, 2015, pp. 169–178.

[8] A. C. Graesser, H. Li, and C. Forsyth, "Learning by communicating in natural language with conversational agents," Current Directions. Psychological Science, vol. 23, pp. 374-380, 2014.

[9] A. C. Graesser, C. M., Forsyth, and P. Foltz, (2016). "Assessing conversation quality, reasoning, and problem solving performance with computer agents," in On the nature of problem solving: A look behind PISA 2012 problem solving assessment, B. Csapo, J. Funke, and A. Schleicher, Eds.,. Heidelberg, Germany: OECD Series. 2016, pp. 275-297.

[10] T. W. Chan and A. B. Baskin, "Learning companion systems," in Intelligent tutoring systems: At the crossroads of artificial intelligence and education, C. Frasson and G. Gauthier, Eds. Norwood, NY: Ablex Publishing, 1990, pp. 6–33.

[11] Z. W. Hong, Y. L. Chen, and C. H. Lan, "A courseware to script animated pedagogical agents in instructional material for elementary students in English education," Computer Assisted Language Learning, vol. 27, pp. 379-394, 2014.

[12] W. L. Johnson, J. W. Rickel, and J. C. Lester, "Animated pedagogical agents: Face-to-face interaction in interactive learning environments," International Journal of Artificial Intelligence in Education, vol. 11, pp. 47–78, 2000.

[13] H. C. Yang, and D. Zapata-Rivera, "Interlanguage pragmatics with a pedagogical agent: The request game," Computer Assisted Language Learning, vol. 23, pp. 395-412, 2010.

[14] Y. Song, J. R. Sparks, J. W. Brantley, T. Jackson, D. Zapata-Rivera, and M. E. Oliveri, "Developing argumentation skills through game-based assessment," in Procedings of the 10th Annual Game Learning Society Conference, Madison, WI. 2014

[15] D. Zapata-Rivera, T. Jackson, L. Liu, M. Bertling, M. Vezzu, M., and I. R. Katz, "Science inquiry skills using trialogues," in the 12th International conference on Intelligence Tutoring Systems, pp. 625-626. 2014

[16] Z. Cai, A. C. Graesser, C. Forsyth, C. Burkett, K. Millis, P. Wallace, D. Halpern, and H. Butler, "Trialog in ARIES: User input assessment in an intelligent tutoring system," In Proceedings of the 3rd IEEE International Conference on Intelligent Computing and Intelligent Systems, W. Chen and S. Li, Eds. Guangzhou: IEEE Press, 2011, pp. 429-433.

[17] D. Jurafsky, and J. Martin, Speech and language processing. Englewood, NJ: Prentice Hall, 2008.

[18] T. Landauer, D. S. McNamara, S. Dennis, and W. Kintsch, Handbook of latent semantic analysis. Mahwah, NJ: Erlbaum, 2007.

[19] K. A. Noels, L. C. Pelletier, and R. J. Vallerand, "Why are you learning a second language? Motivational orientations and self-determination theory," Language Learning, vol. 50, pp. 57-85, 2000.

[20] V. J. Shute, "Focus on formative feedback," Review of Educational Research, vol. 78, pp. 153-189, 2008.

[21] M. Fernández-Toro, and S. Hurd, "A model of factors affecting independent learners' engagement with feedback on language learning tasks," Distance Education, vol. 35, pp. 106–125, 2014.