(O)

2020　2023

Development of an open-source speaking evaluation platform with automated scoring

Daniels, Paul

1,700,000

Through the development of an open-source system with automated scoring of speech, language learners in large classes had more opportunities for extensive speaking practice. In addition, the system proved beneficial for speaking practice for standardized language learning tests.

Successful development of an open-source automated speaking assessment platform was achieved. The platform automated the scoring of speaking tasks, allowed the instructor to add custom speaking tasks, and offered learners individualized speaking feedback. Data collected from student use of the system indicated that student's overall speaking abilities improved. Learners also indicated that the speaking tasks with individualized feedback was helpful. In addition, a positive correlation was observed between the human-generated and computer-generated scores for speaking tasks.

Teaching English a a foreign language

automated speech assessment language learning speaking

Background of the Study

In Japan, there has been a persistent push to significantly improve the English-speaking ability and communicative competency of Japanese students. Faced with a lack of authentic speaking opportunities and limited time allotted for English instruction in elementary school, Japanese students often struggle to make sufficient improvements in their speaking abilities. To overcome these challenges, the Ministry of Education, Culture, Sports, Science and Technology (MEXT) is planning substantive reforms to English education in 2020, from the elementary to the higher secondary school levels, specifically aiming to improve English-speaking skills. Additionally, the National Center for University Entrance Examinations plans to recognize several standardized tests, such as Cambridge, TOEFL, TOEIC, GTEC, TEAP, and IELTS, as part of the new university admission exam system in 2020. Standardized test developers are striving to find more efficient methods to evaluate speaking abilities in a computer testing environment.

To better prepare students for MEXT's 2020 reforms and for the entrance exam speaking component, creative solutions need to be explored to maximize speaking opportunities for language learners. Students need to become more familiar with the speaking tasks they will encounter on the latest exams to feel more relaxed during the actual test. They also need to be exposed to more extensive speaking tasks to improve their speaking skills before the exams. Due to limited opportunities to practice speaking, the use of computer-assisted language learning (CALL) together with automatic speech recognition (ASR) technology can provide learners with additional extensive speaking opportunities.

Purpose of the study

The purpose of this research is to advance the development of an open-source automated speaking assessment platform. The assessment platform will offer educators a low-cost and reliable solution to efficiently assess the speaking abilities of large groups of language learners. The platform will also give learners greater opportunities to practice their speaking skills for the Japanese entrance exams in 2020. This platform can be used to create general speaking practice activities that offer self-evaluative personalized learner feedback and to administer standardized language placement or proficiency tests. The assessment platform is intended to be low cost and flexible, enabling educators to create customized speaking activities based on learner levels and curriculum goals.

Methods of the research

In the first year of the research project, the open-source automated speech scoring platform was developed and updated. Extensive testing was conducted on the platform to investigate both the accuracy of computer-based scoring and the effectiveness of the computer-based speaking tasks.

During the second year of the research project, the platform was successfully deployed in the language learning classroom during the first half of the year, resulting in issues when users accessed the speaking platform via a mobile device. In the second half of the academic year, learners mainly accessed the platform via on-campus PCs, resulting in fewer issues. An efficient coding process was considered to facilitate the development and improvement stages of the platform. Frequent and productive online communication between researchers and the system programmer helped keep the project on track. As a result, the system proved to be both stable and robust, with several hundred users accessing the online speaking tasks.

The computer-scored speech assessment system was fully integrated with the institution's in-house English proficiency test. Over 400 first-year students completed the computer-scored speaking proficiency test. The computer-generated speaking scores were compared against standardized English proficiency test scores such as TOEIC and CASEC.

The initial platform was designed to assess speaking tasks that had specific 'correct' target

phrases or sentences. Research focused on implementing open-ended speaking tasks to support extensive speaking tasks for lower-level language learners. The latest version of the system was able to support translation, listening, speaking, and evaluation. Language learners were able to speak a phrase in Japanese, listen to the English translation, speak the English phrase, and receive feedback. They were also able to listen to the pronunciation of new words or phrases used in a specific context and practice their production skills by repeating the language in the L2. Practice language tasks or language tests were generated based on a database of items created by a learner or an entire class using this latest speech assessment system. In addition to assessing open-ended speaking tasks, the system was improved to provide individualized feedback. Learners' language output was compared with the target language to determine how similar the two are. The new feedback implementation was used with shadowing and dictation tasks to provide learners feedback on their spoken or written production.

In the third year of the study, the rapid advancements in AI proved to be pivotal for language recognition and assessment. Using various AI tools, personalized grammar and content feedback for language learners were more easily and more accurately generated. By leveraging Whisper, OpenAI's speech-to-text engine, accurate text transcripts from students' speaking activities, including group presentations, were successfully generated. The transcribed text was then fed to both 'DeepL Write' and ChatGPT via APIs. These AI tools checked for grammar and punctuation mistakes and rephrased sentences. Using the AI-generated feedback, we established basic rubrics to automatically score language tasks.

In the final year of the research project, to further enhance the platform, previous speaking and writing activities were updated to allow for automatic scoring using freely available generative AI tools, with a specific focus on evaluating these productive tools. The applications for generative AI in language learning are rapidly expanding. To better understand how generative AI could assist with speaking skills, ChatGPT was integrated with some of the speaking activities, allowing language learners to practice their spoken conversation skills with a chatbot. The spoken conversation content was converted to text for reflective purposes or speaking assessment.

Research Results

In the first year of the research, results of the open-source automated speech scoring platform indicated an acceptable reliability level of the computer-based scores generated by the platform. Additionally, improved speed and accuracy of feedback were observed when accessing the platform from a mobile device. Extensive in-class research on the automated speaking assessment system indicated that students' overall listening comprehension scores improved between pre- and post-testing. A descriptive analysis conducted through observation and an online survey revealed that students found the listening-focused class activities engaging and felt their listening abilities improved, especially with intensive listening and speaking tasks.

Further research explored using smart devices rather than personal computers or mobile devices to administer the computer-scored speaking tasks. Initial tests showed that learners were better able to collaborate in groups when completing speaking activities administered via smart devices compared to personal devices.

The adoption of the computer-scored speaking assessment was successful, with fewer than 5% of 400 test-takers experiencing technical issues with the browser-based audio capture feature. Furthermore, a new 'extensive speaking' option was completed and functioned as intended after initial testing. The translation function used an API offered by DeepL.com. While generally accurate, there were instances of inaccurate translations, especially with kanji like '   ,' which can be read in different ways, leading to variations in the English translation. Additionally, proper nouns or katakana English presented challenges for the translation API.

A qualitative study conducted in the final year of the research project compared grammar scores generated by human raters with those generated by generative AI tools using 152 English as a Foreign Language (EFL) student writing samples. The main objective was to

evaluate the effectiveness of generative AI in scoring grammar and providing feedback on writing. The study demonstrated the potential advantages of generative AI in automated essay scoring (AES) in EFL settings.

4	3	2	3
Paul Daniels and Aya Yamasaki			1
Designing an In-house Language Placement Test for Technology Students			2022
TEFL Praxis Journal			43-55
DOI 10.5281/zenodo.7455692			

Paul Daniels			26
Auto-scoring of Student Speech: Proprietary vs. Open-source Solutions			2022
TESL-EJ			1-14
DOI 10.55593/ej.26103int			

Paul Daniels &			18
English and Career Planning: A Case for Students in the School of Environmental Science and Engineering			2021
			139 --146
DOI 10.32149/00002395			

			19
:			2021
			55-68
DOI			

Paul Daniels
Exploring Beliefs, Attitudes, and Experiences of Generative AI in Education
3rd International TEFL Praxis Conference
2024

Paul Daniels
Group presentation evaluation
East Shikoku JALT
2022

Paul Daniels
Moodle presentation feedback database
MoodleMoot Japan
2022

Paul Daniels
A new extensive speaking tool for Moodle
JALTCALL 2022
2022

Paul Daniels
AI and the self-directed language learner
Inaugural research meeting of the TEFL Praxis Association
2022

Paul Daniels
Using automatic speech recognition to develop speaking skills
JACET ESP SIG Kanto
2021

Paul Daniels
Designing computer-scored speaking tasks
JALT 2021 International Conference
2021

Paul Daniels
Designing auto-scored speaking tasks in Moodle
JALT CALL 2021 Conference
2021

Paul Daniels	
Computer-scored speaking activities in Moodle	
Moodle Moot Japan 2021	
2021	

Paul Daniels	
Lighten the load - computer-scored assignments in Moodle	
46th Annual International Conference on Language Teaching and Learning & Educational Materials Exhibition	
2020	

Paul Daniels	
An open-source speaking practice and testing application	
JALTCALL 2020	
2020	

Paul Daniels	
Moodle 101	
Japan Association of Language Teaching, East Shikoku Chapter Event	
2020	

Paul Daniels

A new Moodle quiz question type for speaking practice

Moodle Moot Japan 2020

2020

Paul Daniels

Promoting active learning in the language classroom

The Japan Association for Language Teaching, Matsuyama Chapter

2020

O

(Kumai Nobuhiro)		
(00248999)	(32606)	

O