



Article

AI-Assisted Enhancement of Student Presentation Skills: Challenges and Opportunities

Julia Chen ^{1,*}, Pauli Lai ², Aulina Chan ³, Vicky Man ³ and Chi-Ho Chan ²

¹ Educational Development Centre, The Hong Kong Polytechnic University, Hung Hom, Hong Kong SAR, China

² Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong SAR, China

³ Language Centre, Hong Kong Baptist University, Kowloon Tong, Hong Kong SAR, China

* Correspondence: julia.chen@polyu.edu.hk

Abstract: Oral presentation is a popular type of assessment in undergraduate degree programs. However, presentation delivery and grading pose considerable challenges to students and faculty alike. For the former, many students who learn English as an additional language may fear giving oral presentations in English due to a lack of confidence. For the latter, faculty who teach multiple classes and have many students may find it difficult to spend adequate time helping students refine their communication skills. This study examines an AI-assisted presentation platform that was built to offer students more opportunities for presentation training without the need for faculty intervention. Surveys with students and teachers were conducted to inform the design of the platform. After a preliminary platform was developed, two methods were employed to evaluate its reliability: a beta test with 24 students and a comparison of AI and human scoring of the presentation performance of 36 students. It was found that students are highly receptive to the platform, but there are noticeable differences between AI and human scoring abilities. The results reveal some limitations of AI and human raters, and emphasize the potential benefit of exploring collaborative AI–human intelligence.

Keywords: AI-assisted evaluation; oral presentation training; oral presentation scoring; human vs. AI scoring; higher education; English as an additional language



Citation: Chen, J.; Lai, P.; Chan, A.; Man, V.; Chan, C.-H. AI-Assisted Enhancement of Student Presentation Skills: Challenges and Opportunities. *Sustainability* **2023**, *15*, 196. <https://doi.org/10.3390/su15010196>

Academic Editors: Di Zou and Lucas Kohnke

Received: 10 October 2022

Revised: 5 December 2022

Accepted: 8 December 2022

Published: 22 December 2022



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Effective communication is a key attribute for university graduates. For many higher education institutions (HEIs), “students’ capabilities to . . . read, write, listen, and speak effectively are all now in the spotlight and are an institutional concern” [1] (p. 55). Language training often focuses on developing students’ productive language skills, including speaking, which are especially crucial for HEIs that employ English as the medium of instruction (EMI) and require oral presentations in English as a common form of assessment. Most courses within a student’s major include at least one oral presentation assessment, such as an individual/group PowerPoint or poster presentation. However, it is not feasible for these courses to include all the potential speaking genres that undergraduate students from different disciplines will encounter during their academic pursuits. It is, therefore, necessary to think outside the box—in this case beyond courses and lessons—to explore other means of helping students understand and meet the requirements of different academic speaking tasks, as well as providing them with practice opportunities and useful feedback for improvement.

This paper reports on a study that attempted to incorporate readily available AI tools into a one-stop platform, on which HEI students could access self-directed learning materials and automated feedback for enhancing their presentation skills. It lays out the journey by an interdisciplinary team of academics to identify the right AI tools in order to build a self-directed elearning platform that can intrigue learners, offer some actionable

feedback, and be deployed economically and swiftly enough to support unsupervised learning. The applicability, reliability, limitations and further development opportunities of the said AI tools were evaluated based on a beta test with 24 students and a comparison of AI and human scoring of the presentation performance of 36 students.

2. Literature Review

2.1. *Speech Training Needs of Students Who Learn ENGLISH as an Additional Language (EAL)*

Students' fear of public speaking in public has been widely reported in the literature. This includes higher education students, who are often so anxious about giving oral presentations that fear may become "a contributing factor in student mental health and wellbeing issues" [2] (pp. 1290-91). Another study with 1135 undergraduate students in a country where English is not the first language found that "63.9% of college students reported fear of public speaking" [3] (p. 127.e7). It is commonly observed that "students were hampered by negative associations with past public speaking experiences" [4] (p. 174). On top of a fear of public speaking, students also found it difficult to express themselves effectively in oral presentations [5]. These anxieties are even more common among non-native students who learn English as an additional language (hereafter referred to as EAL students). This apprehension takes several different forms. Some students' anxiety is related to their English proficiency. They feel they have a poor command of English and harbor a phobia of the language; they worry about grammar, vocabulary, and pronunciation, and lack the confidence to speak in English in public [6]. Others fear being the center of attention [7], developing stage fright, and experiencing psychological distress. Some find question-and-answer (Q&A) sessions after the planned delivery element of the presentation assessment horrifying and stressful, as they cannot thoroughly prepare for its necessarily unpredictable aspect.

Engineering students are no exception. Some are concerned that their lack of sufficient linguistic skills affects their explanation of concepts and procedures, and hence they cannot fully exhibit their technical ability or theoretical understanding [8]. Some are aware they do not possess the paralinguistic skills, such as eye contact and body language, that are needed to enhance their verbal delivery. Some are worried that, as presenters, they are not only the focus of attention but also of criticism [9]. Sometimes nervousness and stage fright lead to temporary mental blocks during the presentation and students forget or lose track of ideas [10].

HEIs have adopted a variety of measures to improve EAL students' language abilities. Some have begun to offer more English training courses [11], while others attempt to offer additional language support beyond coursework. The "English Across the Curriculum" movement in Hong Kong is an example of the latter and offers additional, discipline-related speaking and writing resources to students [12]. Capitalizing on students' assessment-oriented mentality, some teachers have introduced speaking tests with interaction and peer assessment to improve students' speaking skills [13]. An investigation conducted in the engineering discipline showed that students found it helpful to "[record] a rehearsal of their presentation, and [examine] their language and organization" [1] (p. 63). In this way, students can watch their recordings and focus sequentially on different aspects of their presentations. This can be a useful method, as focusing students on particular tasks helps reduce communication anxiety and leads to improved communication efficiency. This notion is supported by a study in an engineering course, which observed a "reduction of anxiety and an improved level of confidence" in students who used this method [14] (p. 183).

2.2. *Student Preference for Online English Learning*

Studies find that online learning platforms can motivate students to learn English [15]. Results from research justify the development of software applications to teach EAL courses for engineering and technical study programs, as these tools can lead to improvement in a "learner's cognitive capacity and motivation to study" [16] (p. 11). With the aid of proper

apps or software, students can apply “data-driven critical thinking to academic language learning” [17] (p. 65). Additional research revealed that online learning for students via platforms such as Microsoft Teams is a novel strategy that makes it easier for students to understand the learning resources [18] and improve their language abilities [19]. Indeed, according to Sülter et al., technology, such as AI, virtual reality, and mobile apps, can alleviate students’ anxiety of public speaking [20].

2.3. Technological Advances in the Improvement of the Teaching and Learning of English-Speaking Skills

In recent years, technological advances in the processing of speech have changed interactions between humans and digital devices. Progress has been made in several key areas, including the development of multicore processors, access to larger quantities of data, and the increasing popularity of mobile, wearable, and intelligent living room devices, as well as in-vehicle infotainment systems [21]. Swift advances in voice recognition have enabled the exploration of AI for speech training and have been found to improve students’ presentation skills [22].

The deployment of VR and AI technology in HE English teaching has been found to significantly help the teaching of college English [23]. With real-time data and analyses, computer-aided assessment enables teachers and learners to gain insight into strengths and weaknesses in student performance and consequently allows timely rectification of problems in teaching and learning [24].

2.4. Opportunities for AI in English Speech Acquisition

The automatic speech recognition (ASR) technique is reported to encourage real-time feedback and interactive oral practice in self-paced learning [25]. Given the emerging technology, scholars have called for a clearer understanding of the “unique attributes of speech recognition, in terms of both input data and output labels” [26] (p. 48). According to Baby et al., “[t]he most common approach to automatic segmentation of speech is to perform forced alignment using monophone hidden Markov models (HMMs) that have been obtained using embedded re-estimation after flat start initialization” and “these results are then used in neural network frameworks to build better acoustic models for speech synthesis/recognition” [27]. Kong et al. assert that to resolve problems related to sound event detection and separation, a segmentation framework must be used to compile “weakly labelled data” [28] (p. 777).

One study suggests the setup of a structure for domain adaptation of probabilistic linear discriminant analysis (PLDA) in speaker recognition. This structure consists of several existing supervised and unsupervised domain adaptation methods and, with the introduction of new techniques, enables a more flexible use of available data in different domains, thereby increasing the accuracy of speech recognition [29].

2.5. Considerations in Developing an Online English Learning Platform

Designing an online English learning platform cannot be done by IT professionals alone; it is also essential to incorporate English teachers’ ideas regarding teaching design [30] (p. 2). One study found that “the parties involved in the platform construction processes may, first of all, have a lot to offer in terms of the platform functionalities and should therefore be involved in the platform construction process” [31] (p. 101). Professionals with different expertise can contribute innovative ideas to the creation of a new platform, which requires a flexible and adjustable design to enable its application in different scenarios [32].

Many factors have to be taken into consideration when designing an online English learning platform with the capability to evaluate the pronunciation of English. For example, one study employs “state-of-the-art speaker recognition systems that comprise an x-vector (or i-vector) speaker embedding front-end followed by a probabilistic linear discriminant

analysis (PLDA) backend” [33] (p. 6619). However, a huge pool of data must be available to enable these components to function.

In another study with novel algorithms for measuring phoneme and lexical segmentation errors of people with dysarthria, “although the automated phoneme measures were highly correlated with the manually coded measures, there was a systemic bias from human coding results” [34] (p. 3365). The algorithm could not work in coding transcribed phrases that exceeded or did not include the target syllables [34].

The creation of a full set of ASR algorithms with satisfactory performance would be a challenging task [35]. There are currently five players in this field: Google, Apple, Microsoft, Amazon, and IBM [35]. In addition, the complicated nature of many algorithmic systems often hinders the comprehension of the reasons behind the selection of evaluation tools [36].

2.6. Applicability and Reliability of AI in Language Evaluation

Administration of tests and exams, including language assessment, has been severely disrupted by the pandemic as social distancing requirements have rendered it challenging to conduct in-person assessment. New options for proficiency tests that can be done at home are available, but concerns about validity are pervasive [37]. Some online language tests, particularly in cases where teachers use them instead of assignments that are aligned with specific academic learning outcomes, are found to be limited in terms of the “evidence of their construct alignment and relevance to academic settings” [37] (p. 614).

Evaluation is challenging; naturally, AI will not be appropriate for assessment in some parts of the curriculum [38]. Furthermore, ensuring the reliability of evaluating speaking and writing skills (that is, open-ended, productive skills) has always been a significant challenge [39,40].

Meanwhile, “the design and validation of measurement instruments” may be deployed to improve teaching, and hence the results can be adequately measured [41] (p. 58). Some scholars are optimistic that advancements in AI technology will gradually resolve current technical problems in assessing English speaking, and the use of AI in English teaching and learning in the future is promising [42] (p. 5). Hence, the use of AI to help English teaching and learning has potential, subject to the right selection of tools.

3. Methodology

3.1. Needs and Obstacles in the Present Research Context

The pandemic has disrupted both language learning activities and high-stakes testing. Students, regardless of academic discipline, require good English presentation skills. In the two HEIs in which this study was conducted, many of the students admitted to undergraduate degree programs had English levels equivalent to International English Language Testing System scores of 5.5 to 6. The HEIs employ EMI, which means all oral presentation assignments are delivered completely in English. Due to the large number of students in many courses, grading students’ oral presentations takes considerable time, which means teachers may only be able to focus on content, as opposed to communication skills, when grading presentation assignments.

Although teachers may not have the time to focus on the communication aspects of oral presentations, such as language use, delivery, and pronunciation, they want to offer students pre-assessment training opportunities. Technological advancement has facilitated the diagnosis of speech problems and the integration of AI technologies in foreign language education can support flexible, interactive, and learner-centered approaches. Seeing this as an opportunity, a team consisting of an educational developer, discipline-specific teachers, and language teachers decided to collaborate on the use of ASR to offer prompt feedback and interactive oral practice to support self-paced learning. It is, however, not easy to develop a complete ASR algorithm with reliable performance, which calls into question the objectivity of grading by AI. The ongoing development of AI technology must overcome many obstacles to resolve the existing technical problems in oral English assessment.

3.2. Rationale for the Study

There have been studies on the learning of English or English presentations by EAL students. However, research on the challenges and opportunities in the construction of an online platform to address the needs of EAL students specifically in learning English presentation is scarce. Seeing students struggle with English presentations and drawing on the literature on the potential of using AI in language evaluation, we formed a team of academics from an engineering department, an English language center, and an educational development center in two universities to explore the development of a platform for EAL university students to learn and practice English presentations. The platform would provide learning units focused on the skills required for delivering a strong presentation and allow AI evaluation of oral presentations. The team members from the engineering department lead the technological development of the platform; those from the English language center identified the oral presentation training needs of EAL students and tailored the platform to create learning units that targeted problems specific to those needs, and the team member from the educational development center advised on the educational needs and appropriate teaching methods for HE students.

3.3. Research Questions

To understand whether our AI-assisted learning platform is useful to students and to inform future development, we posed the following research questions:

- (1) Which quantifiers could be included in an AI-assisted presentation training platform?
- (2) What are the challenges and opportunities regarding the development and use of an AI-assisted presentation training platform?

3.4. Research Participants

To answer the research questions, the study involved both engineering and non-engineering students, targeting engineering students from a polytechnic university and humanities students from a mainly liberal arts university. Undergraduate students with different first languages in different years of study were invited to participate in an online survey. Table 1 provides a summary of their demographic composition. Consent was obtained from five students to participate in a follow-up focus group interview.

Table 1. Demographic composition of student participants of the needs analysis survey.

Demographics		PolyU (76 Respondents)	HKBU (28 Respondents)
Gender	Male	52	5
	Female	23	23
	Prefer not to say	1	0
Major	Engineering	72	0
	Science	2	2
	Arts and humanities	0	20
	Business	0	4
	Others	2	2
Current year	Year 1	8	8
	Year 2	22	0
	Year 3	27	6
	Year 4	14	13
	Others	5	1

Table 1. *Cont.*

Demographics	PolyU (76 Respondents)	HKBU (28 Respondents)	
First languages	Cantonese	50	24
	Mandarin	15	4
	English	1	0
	Others	10	0

We also sent out a needs analysis survey that targeted the teachers of the invited students and received 9 responses. All of them were non-language teachers and non-native English speakers.

3.5. Research Methods and Instruments

This study employed a mixed research approach as the dataset from any single procedure would not fully answer the research questions [43,44]. Quantitative data came from surveys that were conducted before and after the team developed the AI-assisted presentation training platform. Numerical data were also obtained from AI and human scoring of student presentations, allowing direct comparisons. Qualitative data were derived from focus group interviews to deepen the understanding of students' needs and preferences.

To investigate students' genuine needs in acquiring presentation skills, we conducted a needs analysis via an online survey, using Microsoft Forms, which consisted of a total of 35 questions in English. The survey for students comprised 14 questions to gauge their feelings about giving oral presentations, their oral presentation needs, and areas in which an AI-assisted presentation training platform could help. The teachers of these students also completed an online survey which comprised 21 questions about their perceptions of students' feelings about oral presentations, their grading experience, and areas in which an AI-assisted presentation trainer could help in assessing presentations.

Besides, focus group interviews with students were designed to follow up on the survey results and to deepen understanding of students' views and responses. All survey respondents were invited to join the interviews, which were transcribed verbatim.

3.6. Research Procedure

The team sent mass invitation emails to students and teachers of engineering and humanities disciplines about participation in the online baseline survey. Participation in the surveys and interviews was voluntary. Respondents were given a timeframe to submit their responses online.

Based on the results of the needs analysis, the team designed an online AI-assisted presentation training platform that features learning units catering to the needs of EAL speakers. This platform also features AI assessment of key components of oral presentations, such as pronunciation accuracy, fluency, vocal fillers, and facial emotions.

After a quick pilot build of the AI-assisted platform, the team conducted a beta test, using a Microsoft Forms survey with students to test the reliability of the system. This pilot build of the platform offered three tools for testing: pronunciation, facial expression, and vocal filler. Participants began by completing a 5-min survey to test their baseline understanding of oral presentations. They were then required to study the learning units and submit two video-recorded presentations for AI assessments. Finally, students were asked to complete a 10-min survey to provide their feedback on using our platform. The duration of the beta test process took as little as 45 min, or longer when respondents chose to study the provided content in depth.

To compare AI and human scoring, the team hired an English teaching professional—a teacher who is not currently teaching at the two universities involved in the study but has over 10 years of experience in English teaching, curriculum development, course delivery, and assessment of EAL courses at university level—to rate students' oral presentations

that had been submitted on our platform. The team then compared the manually graded results with the AI-graded results to discern differences in human and AI scores, shed light on the strengths and weaknesses of AI evaluation, and identify areas where AI would need to evolve to streamline the assessment process and enhance its precision.

4. Results and Discussion

4.1. Baseline Study on Needs Analysis

The baseline needs analysis aimed to determine how students and teachers perceive oral presentations as an assessment form, as well as to identify the training needs of students and the grading needs of teachers. We received survey responses from 104 students and nine teachers and conducted focus group interviews with five students who indicated on the needs analysis survey that they were willing to be interviewed.

From the nine responses from non-language teachers, we noted that most teachers would provide feedback on content (75%) and delivery (87.5%) (see Figure 1).

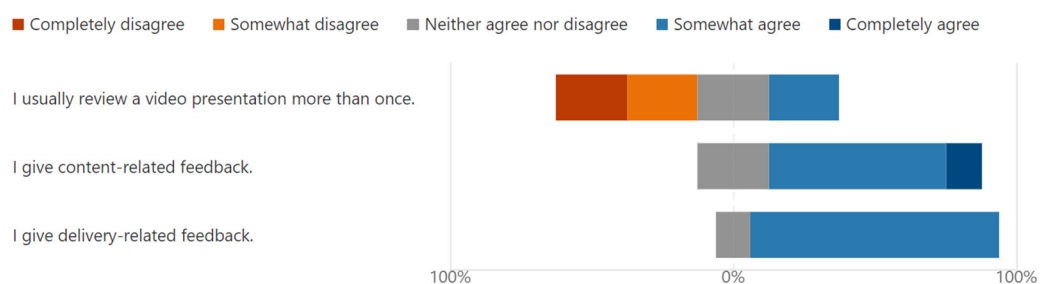


Figure 1. Teachers' feedback pattern.

Given that 88% of the non-language teachers grade 20 to over 101 presentations per semester (see Figure 2) and 50% review a video presentation more than once, there appears to be a strong grading burden for teachers, and therefore that an AI-assisted system with automated customizable grading assistance would be helpful.

Presentations to grade per semester

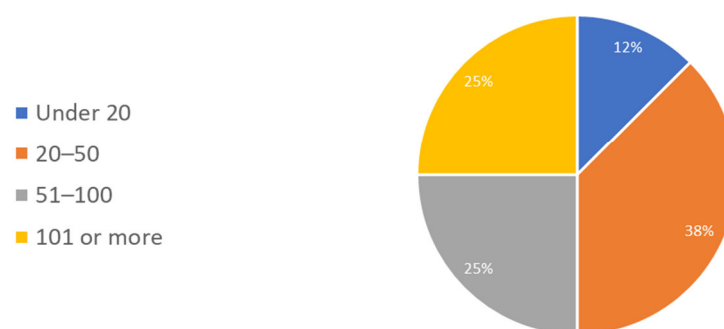


Figure 2. Presentations to grade per semester.

When asked about their preferred AI features, respondents indicated that they would like to include automation in the grading of fluency (78%), accuracy (78%), and eye contact (67%) (see Figure 3). This shows that teachers can potentially use an AI grading platform to help with non-content grading. Teacher respondents also noted a strong need for additional training in students' delivery skills (78%), structure and organization (67%), and referencing (33%) (see Figure 4).

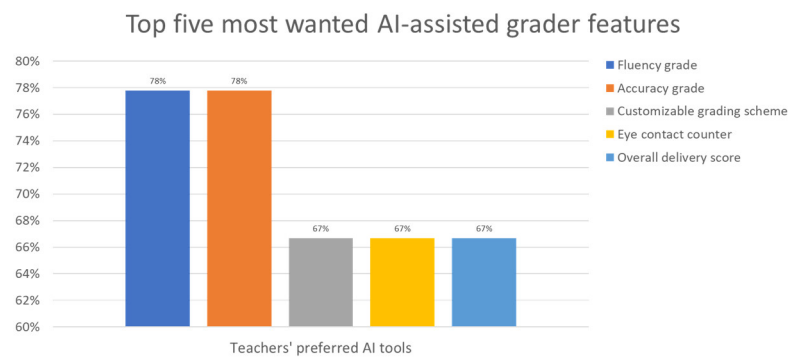


Figure 3. Top five most wanted AI-assisted grading features.

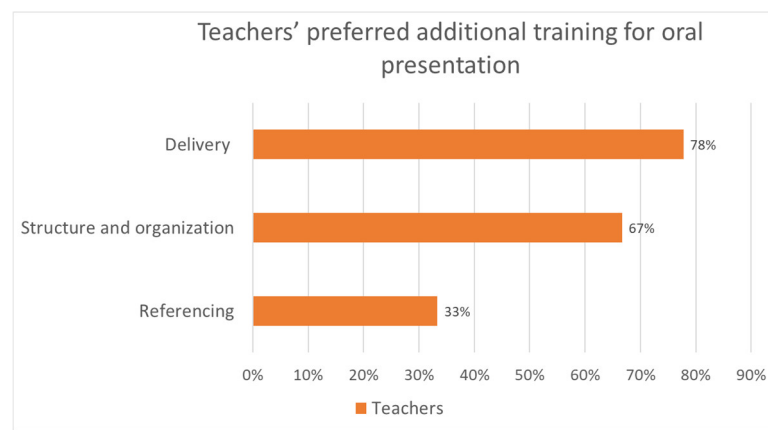


Figure 4. Teachers' preferred additional training for oral presentation.

Student respondents came from two broad categories, engineering and humanities. We noted a marked difference between the two broad disciplines, with engineering students disliking oral presentations more than their humanities peers—50% of humanities students responded they liked to do oral presentations, compared with only 26% of engineering students. We noted from some neutral comments that students were well aware of the pain points of oral presentations as an assessment. Student responses included, “oral presentations can showcase public speaking skills and charisma, so there is a higher chance of getting better scores”, and “it can be stressful for the presenter who may focus more on presenting techniques than professional skills.” They noted that those with stronger presentation techniques have a higher chance of scoring better and that nervousness, weakness in English speaking skills, and poor eye contact are undesirable for this kind of assessment. In fact, 24% of engineering students and 8% of humanities students had had poor scoring experiences in the past for presentation tasks. Respondents also reported they were uncomfortable with public speaking and had accent anxiety. Similar to teachers, they considered additional training in delivery skills (57%) more important than structure and organization (48%) and referencing (30%). Their top three most wanted AI-assisted training features were “filler alerts”, “silence checker”, and “eye contact counter”.

The findings from this baseline needs analysis helped us gauge students' and teachers' preferences for an AI-assisted presentation training and grading platform, as well as which features to prioritize in its development. Prioritization is particularly important for our small development team, as the AI features currently available from the market are not tailored for oral presentation training. Customization of these features required the team to analyze collected samples, compare perceived accuracy against machine performance, as well as incorporate elements for editable feedback by humans where the automation falls short. This is a very resource-intensive process that may not be viable without prioritization. We learned from the baseline study that presentation delivery is an area where training

is a top priority and therefore tried to map existing, well-trained AI features—including accuracy, fluency, vocal fillers, and facial expressions—to support self-paced learning with instant feedback and AI-assisted grading.

4.2. Development of an AI-Assisted Training Platform

Based on the results of the needs analysis, we designed an online AI-assisted presentation training platform, which was developed as a web application so that users could access the platform using any browser. It was designed to be an all-in-one platform for students to learn practical tips about oral presentations and rehearse with AI tools. The platform contains two modules: “Learning Units” and “Course and Assignment”. The “Learning Units” module provides presentation tips to help students learn how to deliver a good presentation. The “Course and Assignment” module allows students to submit their presentation assignments to the platform for AI evaluation and teacher grading. The hierarchical structure of the platform is shown in Figure 5.

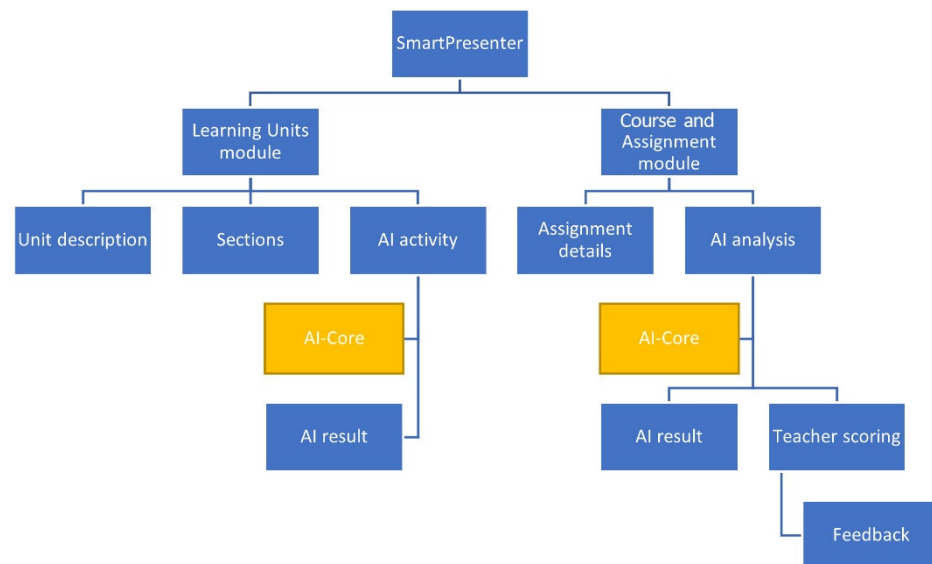


Figure 5. The hierarchical structure of the AI-assisted presentation training platform.

For the “Learning Units” module, our language team designed customized presentation tips for our university students, including topics such as content and structure, delivery, and pronunciation. Each learning unit includes a description and different sections. At the end of each unit, there is an AI activity for students to submit a presentation to practice what they have learned, which will then be evaluated by AI tools for the assessment of elements such as facial expression, vocal fillers, and pronunciation. An example of a learning unit is shown in Figure 6.

In the “Course and Assignment” module, students can submit a presentation under an assigned course. They can access learning units to receive presentation tips and submit presentation assignments to their assigned course(s) with multiple attempts allowed. After submission, they can check their attempt history with AI results and view the assessment results with the teacher’s feedback. Figure 7 shows students’ submission of presentation assignments, and Figure 8 shows the assessment results with the teacher’s feedback.

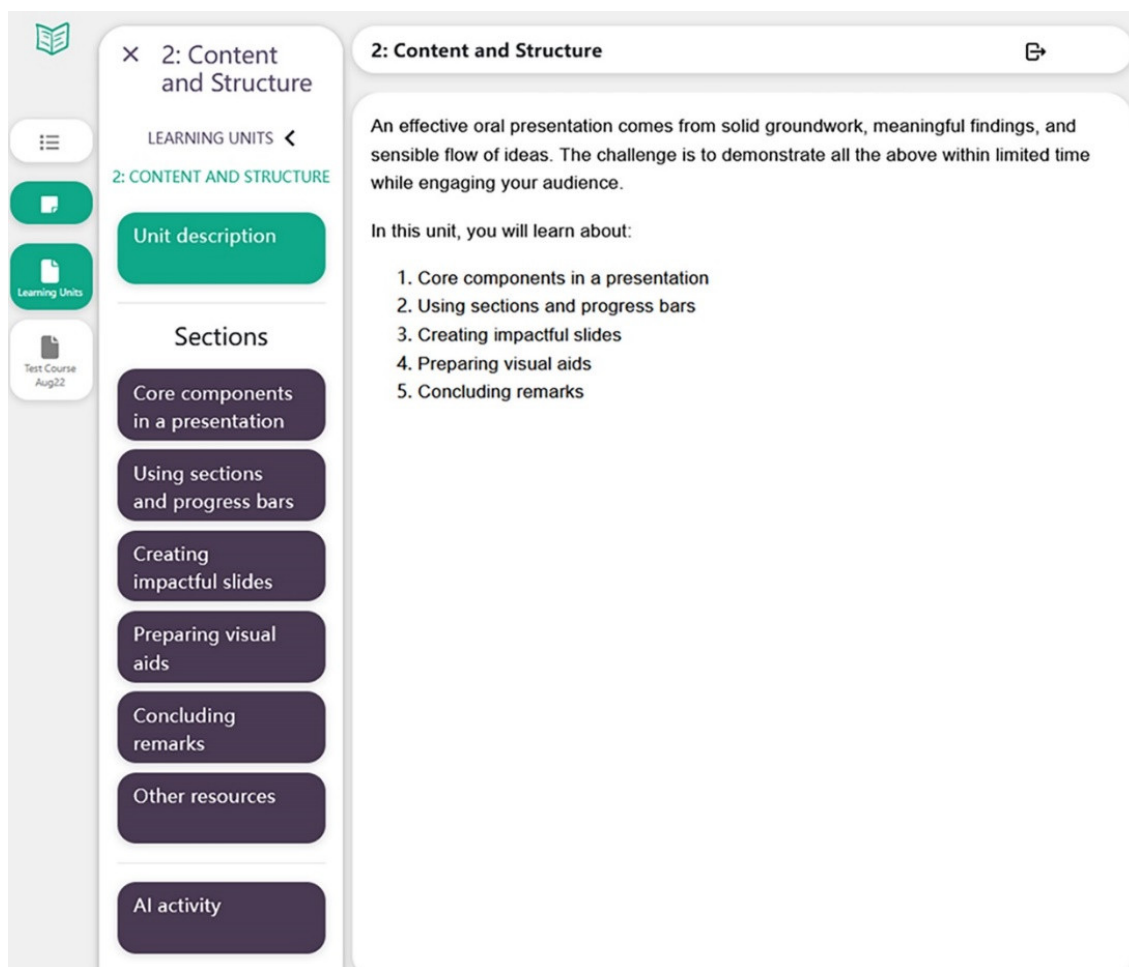


Figure 6. An example of a learning unit.

To evaluate the presentation, the platform uses mature AI web services that are trained by machine learning and readily available on the market. The AI tools adopted include facial expression, vocal filler detection, and pronunciation assessment. For facial expressions, the adopted AI tool can detect happiness, sadness, neutral feelings, anger, contempt, disgust, surprise, and fear. Figure 9 shows the results from a facial expression analysis.

Vocal filler detection detects the use of filler words in speech. Our language team supplies the list of customized filler words that are commonly used by local university students. This list is fed into our database. We then use an AI tool to detect the appearance of these words in a speech. The result displays the frequency of each filler word, as shown in Figure 10.

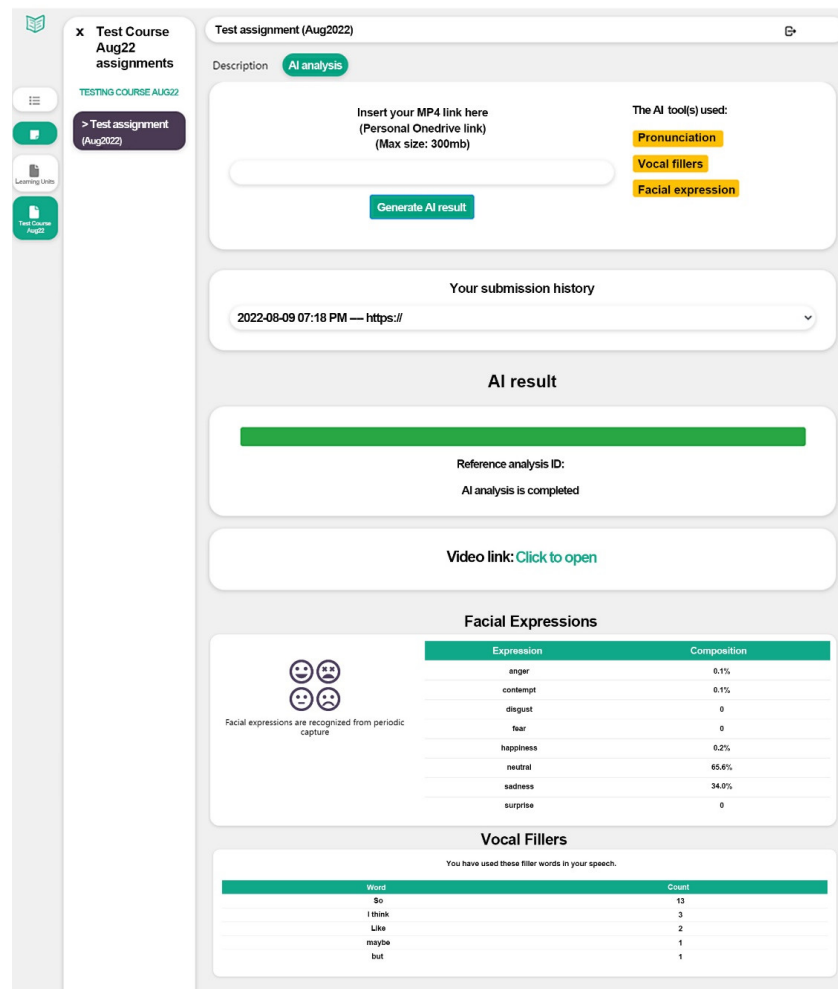


Figure 7. Sample page of presentation assignment submission by students.

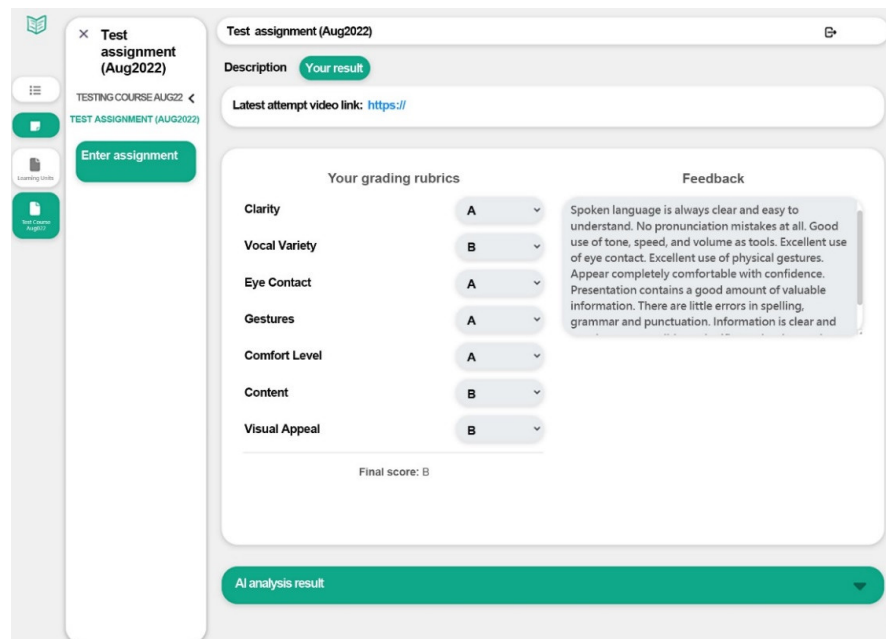


Figure 8. Sample page of assessment results with the teacher's feedback.

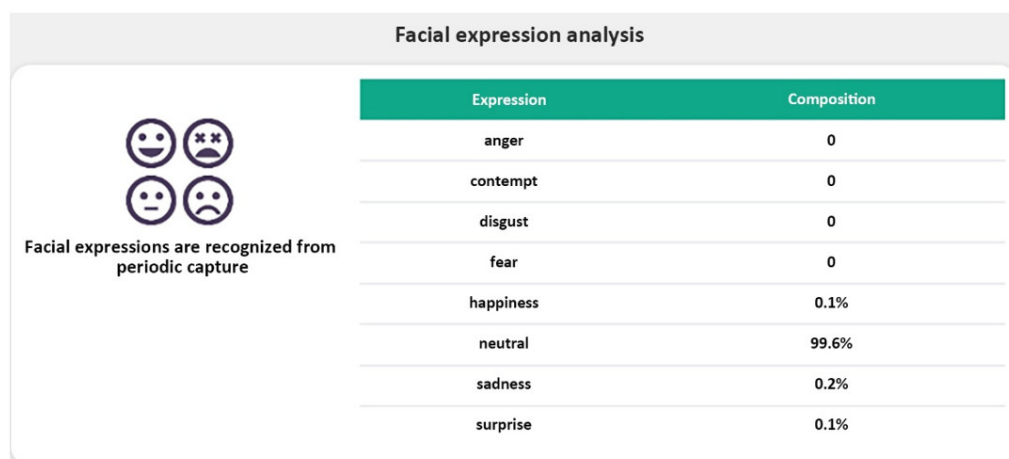


Figure 9. Results of facial expression analysis.

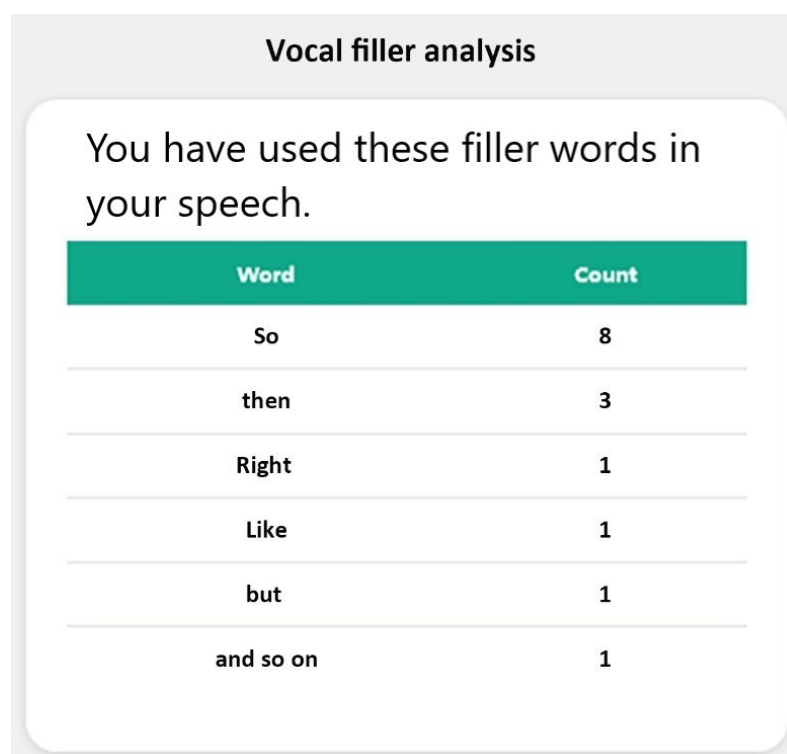


Figure 10. Result of vocal filler analysis.

For pronunciation assessments, we use an AI tool to determine accuracy and fluency. Pronunciation accuracy is evaluated based on phonemes, while fluency is evaluated based on silent breaks between words. The overall score is an average score derived from the above two scores. The result displays the accuracy score, the fluency score, the overall score, and the AI-generated script, as shown in Figure 11.

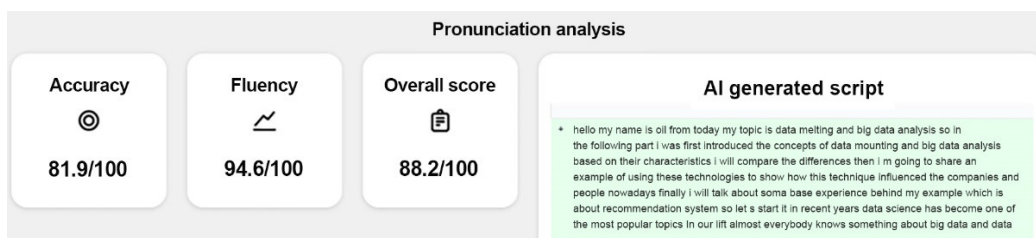


Figure 11. Result of pronunciation analysis.

Numerous challenges were encountered in developing the platform. Since this is a collaborative project across two universities, we had to overcome the challenge of cross-university account logins. To allow students to use their university email accounts to log in and protect their privacy by not storing their passwords in the system, we integrated the authentication service offered by an external provider. Another problem concerns AI tools. We used existing web services, which should have been maturely trained by machine learning. However, most existing AI services could not precisely fulfill the needs of our approach to presentation analysis. We needed to develop a customized algorithm on top of existing AI services to create AI tools specific to our platform. In addition, some AI web services take only images as input, so we were obliged to pre-process the videos to extract useful assets for AI analysis. Considering the challenges of a long waiting time due to video pre-processing and AI analysis, unclear processing status, and data loss due to accidental termination of the AI analyzing process, we created an AI job scheduling subsystem to solve the problem. Each video submission is now generated as an individual task in the system and is automatically stacked in a queue for background processing. This subsystem allows multiple students to submit assignments simultaneously without congestion.

4.3. Beta Test of the Platform

Following the pilot build of the platform, we conducted a beta test to examine the reliability of the system. Students were invited to trial the pilot platform and twenty-four joined the beta test. User responses were very encouraging, as the vast majority of them found the learning units on content and structure (91.7%), delivery (87.5%), and pronunciation (87.5%) helpful for improving their oral presentation skills. Many students found AI tools for pronunciation (79.2%), facial expression (78.3%), and delivery (87.6%) to be helpful as well. The machine-generated scores presented in the form of a scorecard, in particular, invited more interest in fine-tuning specific skills, with students suggesting that the system could list not only “mispronounced words and the corresponding correct sound with IPA” but also “the reasons for their current score and how they can improve”.

Linking these responses to some of the needs identified from the baseline study, we see the potential of both the learning units and the AI tools as useful self-directed learning tools for students. As no humans are involved in the AI scoring, and since students can test their performance as many times as desired without tiring a human assessor, they may find it less uncomfortable having a machine help them polish the skills that they might otherwise be too shy to practice. AI scoring also guarantees that assessment is strictly based on quantifiable parameters, thus minimizing any bias perceived by students, an issue that was revealed as a concern in the needs analysis.

Respondents also indicated that they would welcome more sophisticated and “clever” functions, with some suggesting additional tools for tracking eye contact, body language, and speech pace. The valuable feedback and suggestions illustrate a strong interest in AI and students’ faith in more impartial machine assessment.

4.4. Comparing AI Scoring with Human Scoring

To further ascertain the reliability of AI grading and the areas of the platform that may need improvement, the team decided to invite an external rater who is a highly experienced English language teacher (that does not teach in the universities that developed

the platform) to participate in a comparison of the autogenerated AI and human grading. Thirty-six video presentation samples from an engineering class were selected by the subject teacher, including high, medium, and low performances. As a reliability test of the grading component of the pilot platform, these videos were first scored by AI and then coded for anonymity before being graded by the external rater. The rater was tasked to provide qualitative feedback and/or a numerical score for the following metrics: fluency, accuracy, perceived facial expressions (emotions), and vocal fillers. The grading was done according to the rubrics in Appendix A.

4.4.1. Fluency

The autogenerated fluency scores were compared with the human rater's and a correlation coefficient of 0.136 was obtained. According to Microsoft, this value falls into the category of low association, which means the autogenerated scores are not significantly aligned with human ratings [45]. It is also noted that the human fluency score, which spans a 46-point range, is much broader than the AI range of 13.7 points. This weak correlation may be attributed to the use of "silent breaks between words" as the main measurement in the autogenerated fluency score. As illustrated in the rater's qualitative comments on samples with the biggest (51.4) and smallest (1.3) scoring differences below, the human perception of poor fluency can come from inefficient use of word linking, chunking, sentence stress, and rhythm, all of which cannot be measured by silence breaks alone (see Table 2).

Table 2. Qualitative comments from human rater on fluency.

	Fluency Scoring with Biggest Difference between Machine and Human Rater (51.4)	Smallest Fluency Scoring Difference (1.3)
Human rater comments	<i>The speech was extremely unnatural and choppy due to a complete lack of the speaker's ability to maneuver linking of words, chunking, sentence stress and rhythm; although not many "long" awkward breaks were noted, the overall fluency suffered greatly because of the above factors.</i>	<i>Very fluent and natural with little hesitation only.</i>

In assessing fluent speakers, however, the silence break measurement is more accepted by the human rater. The human rater's qualitative comments on video samples that score well on both sides consist of descriptors such as "very little hesitation", "some stumbling but otherwise fluent throughout", and "fluent with minimal distracting breaks." Therefore, we contend that the machine-generated feedback on the frequency of silent breaks as the main indicator of fluency may only be reliable for fluent speakers. Yet, the same mechanism would fail to screen out poor performers, who may confuse the tool by using less obvious tricks between words—for example, the autogenerated fluency score would not be affected even when meaningless utterances are made. As a result, the score may fail to reflect poor fluency, making it difficult to provide actionable feedback for self-directed language learning.

4.4.2. Accuracy

The autogenerated accuracy score, which is based on phonemic closeness to a native speaker's pronunciation, is more aligned with the human rater, as indicated by the correlation coefficient value of 0.450. This is considered a medium-strength correlation, meaning that while autogenerated scores are generally aligned with human perception, there are differences and subjective human ratings may differ from AI-generated results. The qualitative comments on the samples with the biggest (36.8) and smallest (0.7) scoring differences are listed in Table 3.

Table 3. Qualitative comments from human rater on accuracy.

	Accuracy Scoring with Biggest Difference between Machine and Human Rater (36.8)	Smallest Accuracy Scoring Difference (0.7)
Human rater comments	<i>The presenter had many systematic errors in individual sounds, strong and weak forms, word pronunciation, and sentence stress. His grammatical accuracy was also inadequate. If the listener had no slides to refer to, then most probably the speech could not be understood at all (e.g., a keyword “patients” was mispronounced vaguely as “parent” or “parents” repeatedly).</i>	<i>Only very few words (e.g., products or produces) were pronounced very unclearly.</i>

Similar to our observations of fluency, the ranges of human (50) and AI (11.5) accuracy scores exhibited a marked difference, with the former more than four times broader than the latter. We believe this is a result of using phoneme-level accuracy as the basis for scoring. As noted in the human rater’s qualitative comments, human perception and cognition are highly sensitive to inaccurate grammar and mispronunciation. Incorrect stress in a word such as “produce” (noun: /^lprɒdʒ.u:s/, verb /prə^ldʒu:s/), for example, was picked up by humans but not by the AI tool. In fact, 15 out of 36 video samples were scored more than 20 points higher by the AI tool. Based on these observations, we propose that an AI accuracy score based on phonemes can only be useful when it is combined with other metrics and perhaps requires a different label in a self-directed language learning tool.

4.4.3. Emotion

The AI emotion scoring tool, which focuses on facial expression, has a very strong advantage: it does not become fatigued from continuously looking at students’ faces; it can perform under poor lighting conditions and even when the speaker window is small; it is more impartial than any human rater; it can also provide many more details on eight different emotions, namely, anger, contempt, disgust, fear, happiness, neutrality, sadness, and surprise. The human rater’s scores focused primarily on happy and neutral emotions, despite all eight options being provided on the mark sheet. That said, the rater did provide comments beyond facial expressions, unlike the AI scoring tool, such as “engaging/distractive hand gestures”, “display of interest”, “uplifting/happy voice”, “loud and clear sound”, “enthusiastic”, “stilted”, “self-absorbed and unaware of an audience”, “energetic”, and “emotionless.” Self-directed learners may find these qualitative comments more useful than a percentage value for each of the eight facial expressions.

4.4.4. Vocal Fillers

As our sample videos are collected from non-native English speakers, and since vocal fillers are known to be a prominent problem for many EAL learners during oral presentations, an AI tool was developed to identify a list of fillers. While the AI tool can count most of the listed filler words, it does not identify non-word fillers such as “hmm” and “err”. We consider this a largely effective function, as it frees a human grader from the tedious task of counting unwanted words in a presentation and offers concrete feedback to students who may be unaware of their overuse of a particular filler. On the other hand, we need to fine-tune the current tool so that it can identify and mark non-word utterances for students’ reference. An additional list of qualitative feedback can also help students understand the importance of reducing fillers and learn techniques for that purpose.

4.4.5. Overall Observations from the Comparison between AI and Human Scoring

In terms of reliability, the AI vocal fillers identification function is most useful, as it can automatically report repeated utterances, which are usually undesirable. Students can easily eliminate them as they continue to practice with the platform. The AI-generated accuracy score, which has a medium-strength correlation with a human score, can become more useful if it is designed to show exactly which utterance is considered inaccurate.

The AI fluency score and the emotion identifier are both weaker performers among the tested features, therefore their use in the platform must be further investigated and fine-tuned. For example, the current AI-generated fluency score may be integrated with another metric such as words per minute, so that platform users can act on the given scores in a more meaningful way.

Similar inter-rater comparisons must be carried out for new platform features in the future, firstly as a reliability test, and secondly to explore opportunities to adjust and consolidate into more meaningful feedback to support users' self-assessed learning.

5. Conclusions

Our AI-assisted presentation training platform was created based on three observations: (i) EAL students need more practice giving oral presentations in English; (ii) discipline faculty (such as engineering teachers) often teach large classes with oral presentation assignments and have time to focus on the content of students' oral presentations but not their communicative aspects; and (iii) technological advancements have raised the possibility of AI scoring as a means of giving feedback to students about certain aspects of their oral presentations. The speaking practice needs of EAL students were corroborated via two surveys for students and teachers, the results of which then formed the design of our AI training platform. The first build of the platform revealed numerous challenges that compelled us to evaluate the preliminary platform via two channels: a beta test with 24 students and a comparison of AI and human scoring of the presentation performances by 36 students. While the former yielded highly encouraging results of students finding the AI training useful in numerous aspects of communication (such as structure, delivery, and pronunciation), the latter showed a weak to medium association between AI and human raters in terms of fluency and accuracy, respectively, with the AI tool having the distinct advantage of continuous capture of facial emotions—even in low-light situations—and the counting of vocal fillers. The main differences between AI and human ratings stem from the (in)ability of AI to comprehend nuances. Not all abstract concepts (e.g., fluency) can be measured through AI (e.g., silence breaks between words, as pausing can be a purposeful act for dramatic effect rather than an indication of hesitancy or disfluency). In contrast, the human scorer cannot match the ability of AI to register larger numerical occurrences but can detect the multiple facets of communication and their interplay.

As presented in the Section 4 above, AI evaluation of oral presentations has limitations in its current stage of development and cannot be exclusively relied upon as an assessment tool. Communication is multifaceted and immensely complex; to become more aligned with human rating and to supplement human raters, the training platform should include more AI tools to strengthen the rating of the important concepts of fluency, accuracy, facial expressions, and to examine the possibility of expanding the scope of AI evaluation to include other key aspects of presentations, such as eye contact and helpful signposting. Future work is also warranted in fine-tuning the definition of the criteria for assessing presentations to improve the reliability of AI evaluation. The potential of artificial and human intelligence working in tandem should continue to be explored in the ongoing search for solutions that can enhance learning experiences and improve learning outcomes.

Author Contributions: Conceptualization, J.C., P.L. and V.M.; Data curation, J.C., P.L., A.C., V.M. and C.-H.C.; Formal analysis, P.L., A.C. and C.-H.C.; Funding acquisition, J.C., P.L. and V.M.; Investigation, A.C. and C.-H.C.; Methodology, J.C., P.L., A.C. and V.M.; Project administration, P.L. and V.M.; Visualization, A.C. and C.-H.C.; Writing—original draft, J.C., P.L., A.C., V.M. and C.-H.C.; Writing—review & editing, J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Hong Kong University Grants Committee's Special Grant for Strategic Development of Virtual Teaching and Learning, project number LTC19-22/VTL-IICA/EIE.

Institutional Review Board Statement: The project has been approved by the Human Subjects Ethics Sub-committee (HSESC) (or its Delegate) of The Hong Kong Polytechnic University (HSESC Reference Number: HSEARS20220712003).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: As the research is ongoing, the data is not publicly archived. Please contact the authors for any enquiries about the data reported in this article.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Grading rubrics for fluency and accuracy adopted by human rater.

Grade	Fluency	Accuracy
A (85–100%) A– (80–84%)	The speech is presented very fluently. Any silent breaks are simply occasional lapses.	The pronunciation of vowels, consonants, and consonant clusters is always accurate and intelligible. Any errors are simply occasional lapses.
B+ (75–79%) B (70–74%) B– (65–69%)	The speech is presented fluently most of the time. A few awkward silent breaks exist. The speech is presented fluently sometimes. Some awkward silent breaks exist.	The pronunciation of vowels, consonants, and consonant clusters is mostly accurate and intelligible. A few (systematic) errors exist. The pronunciation of vowels and consonants is generally intelligible. Some (systematic) errors exist.
C+ (60–64%) C (55–59%) C– (50–54%)	The speech is NOT presented fluently most of the time. A large number of awkward silent breaks exist.	The pronunciation of vowels, consonants, and consonant clusters is often inaccurate and unintelligible. A large number of (systematic) errors exist.
D (>49%)	The speech is NOT presented fluently all the time. A very large number of awkward silent breaks exist.	The pronunciation of vowels, consonants, and consonant clusters is always inaccurate and unintelligible. A very large number of (systematic) errors exist.

References

- Morton, J.; Rosse, M. Persuasive Presentations in Engineering Spoken Discourse. *Australas. J. Eng. Educ.* **2011**, *17*, 55–66. [\[CrossRef\]](#)
- Grieve, R.; Woodley, J.; Hunt, S.E.; McKay, A. Student fears of oral presentations and public speaking in higher education: A qualitative survey. *J. Furth. High. Educ.* **2021**, *45*, 1281–1293. [\[CrossRef\]](#)
- Marinho, A.C.F.; de Medeiros, A.M.; Gama, A.C.C.; Teixeira, L.C. Fear of Public Speaking: Perception of College Students and Correlates. *J. Voice* **2016**, *31*, 127.e7–127.e11. [\[CrossRef\]](#)
- McDougall, J.; Holden, H. The silence about oral presentation skills in distance and online education: New perspectives from an Australian university preparatory programme. *Open Learn. J. Open Distance E-Learn.* **2017**, *32*, 163–176. [\[CrossRef\]](#)
- Bhandari, B.; Chopra, D.; Singh, K. Self-directed learning: Assessment of students' abilities and their perspective. *Adv. Physiol. Educ.* **2020**, *44*, 383–386. [\[CrossRef\]](#)
- Al Harun, M.O.F.; Islam, K.M.A.; Rahman, M.A. Challenges in oral presentation in English for the freshers at tertiary level. *Green Univ. Rev. Soc. Sci.* **2016**, *3*, 137–157.
- Al-Nouh, N.A.; Abdul-Kareem, M.M.; Taqi, H.A. EFL College Students' Perceptions of the Difficulties in Oral Presentation as a Form of Assessment. *Int. J. High. Educ.* **2014**, *4*, 136. [\[CrossRef\]](#)
- Mohamed, A.A.; Asmawi, A. Understanding Engineering Undergraduates' Technical Oral Presentation: Challenges and Perspectives. *Int. J. Lang. Educ. Appl. Linguist.* **2018**, *1*, 41–53. [\[CrossRef\]](#)
- Riaz, M.; Riaz, M.R. Causes of Anxiety among Engineering Students While Making Oral Presentations in English. *Pak. J. Psychol. Res.* **2022**, *37*, 205–218. [\[CrossRef\]](#)
- Soomro, M.A.; Siming, I.A.; Shah, S.H.R.; Rajper, M.A.; Naz, S.; Channa, M.A. An Investigation of Anxiety Factors During English Oral Presentation Skills of Engineering Undergraduates in Pakistan. *Int. J. Engl. Linguist.* **2019**, *9*, 203. [\[CrossRef\]](#)
- Kim, E.G.; Shin, A. Seeking an Effective Program to Improve Communication Skills of Non-English-Speaking Graduate Engineering Students: The Case of a Korean Engineering School. *IEEE Trans. Dependable Secur. Comput.* **2014**, *57*, 41–55. [\[CrossRef\]](#)

12. Chen, J.; Chan, C.; Man, V.; Tsang, E. Helping students from different disciplines with their final year/capstone project: Supervisors' and students' needs and requests. In *English across the curriculum: Voices from around the world*; Morrison, B., Chen, J., Lin, L., Urmston, A., Eds.; WAC Clearinghouse: Fort Collins, CO, USA, 2021; pp. 91–106. [CrossRef]
13. Karpovich, I.; Sheredekina, O.; Krepkai, T.; Voronova, L. The Use of Monologue Speaking Tasks to Improve First-Year Students' English-Speaking Skills. *Educ. Sci.* **2021**, *11*, 298. [CrossRef]
14. Pathak, A.; Le Vasan, M. Developing oral presentation competence in professional contexts: A design-based collaborative approach. *Int. J. Eval. Res. Educ.* **2015**, *4*, 179–184. [CrossRef]
15. Akobirov, F. *The Influence of Technology on Language Learning and Motivation with Uzbek EFL and United States ESL Students*; University of Kansas: Lawrence, KS, USA, 2017.
16. Stefanovic, S.; Klochkova, E. Digitalisation of Teaching and Learning as a Tool for Increasing Students' Satisfaction and Educational Efficiency: Using Smart Platforms in EFL. *Sustainability* **2021**, *13*, 4892. [CrossRef]
17. Bohát, R.; Rödlingová, B.; Horáková, N. Applied linguistics project: Student-led computer assisted research in high school EAL/EAP. In Proceedings of the 2015 EUROCALL Conference, Padova, Italy, 26–29 August 2015; pp. 65–70. [CrossRef]
18. Rojabi, A.R. Exploring EFL Students' Perception of Online Learning via Microsoft Teams: University Level in Indonesia. *Engl. Lang. Teach. Educ. J.* **2020**, *3*, 163–173. [CrossRef]
19. AlAdwani, A.; AlFadley, A. Online Learning via Microsoft TEAMS During the COVID-19 Pandemic as Perceived by Kuwaiti EFL Learners. *J. Educ. Learn.* **2022**, *11*, 132. [CrossRef]
20. Sülter, R.E.; Ketelaar, P.E.; Lange, W.-G. SpeakApp-Kids! Virtual reality training to reduce fear of public speaking in children—A proof of concept. *Comput. Educ.* **2021**, *178*, 104384. [CrossRef]
21. Yu, D.; Deng, L. *Automatic Speech Recognition: A Deep Learning Approach*; Springer: London, UK, 2015.
22. Junaidi, H.B.; Julita, K.; Rahman, F.; Derin, T. Artificial intelligence in EFL context: Rising students' speaking performance with Lyra Virtual Assistance. *Int. J. Adv. Sci. Technol.* **2020**, *29*, 6735–6741. Available online: <http://repository.unhas.ac.id/id/eprint/11102> (accessed on 6 December 2022).
23. Yang, Z. Application and Exploration of VR and AI Technology in College English Teaching. *Adv. Multimed.* **2022**, *2022*, 1810177. [CrossRef]
24. Muwanga-Zake, J.W.F. Applications of computer-aided assessment in the diagnosis of science learning and teaching. *Int. J. Educ. Dev. Using Inf. Commun. Technol.* **2006**, *2*, 44–66.
25. Chen, C.-H.; Koong, C.-S.; Liao, C. Influences of Integrating Dynamic Assessment into a Speech Recognition Learning Design to Support Students' English Speaking Skills, Learning Anxiety and Cognitive Load. *Educ. Technol. Soc.* **2022**, *25*, 1–14.
26. Yu, D.; Deng, L. Hidden Markov models and the variants. In *Automatic Speech Recognition: A Deep Learning Approach*; Springer: London, UK, 2014; pp. 23–54. [CrossRef]
27. Baby, A.; Prakash, J.J.; Subramanian, A.S.; Murthy, H.A. Significance of spectral cues in automatic speech segmentation for Indian language speech synthesizers. *Speech Commun.* **2020**, *123*, 10–25. [CrossRef]
28. Kong, Q.; Xu, Y.; Sobieraj, I.; Wang, W.; Plumbley, M.D. Sound Event Detection and Time–Frequency Segmentation from Weakly Labelled Data. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2019**, *27*, 777–787. [CrossRef]
29. Lee, K.A.; Wang, Q.; Koshinaka, T. The CORAL+ algorithm for unsupervised domain adaptation of PLDA. In Proceedings of the ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 5821–5825. [CrossRef]
30. Li, J. Design, Implementation, and Evaluation of Online English Learning Platforms. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 5549782. [CrossRef]
31. Oliwa, R. The process of designing the functionalities of an online learning platform—A case study. *Teach. Engl. Technol.* **2021**, *21*, 101–120.
32. Watanobe, Y.; Intisar, C.; Cortez, R.; Vazhenin, A. Next-Generation Programming Learning Platform: Architecture and Challenges. *SHS Web Conf.* **2020**, *77*, 01004. [CrossRef]
33. Wang, Q.; Okabe, K.; Lee, K.A.; Koshinaka, T. A Generalized Framework for Domain Adaptation of PLDA in Speaker Recognition. In Proceedings of the ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, Barcelona, Spain, 4–8 May 2020; pp. 6619–6623. [CrossRef]
34. Jiao, Y.; LaCross, A.; Berisha, V.; Liss, J. Objective Intelligibility Assessment by Automated Segmental and Suprasegmental Listening Error Analysis. *J. Speech Lang. Hear. Res.* **2019**, *62*, 3359–3366. [CrossRef]
35. Iancu, B. Evaluating Google Speech-to-Text API's Performance for Romanian e-Learning Resources. *Inform. Econ.* **2019**, *23*, 17–25. [CrossRef]
36. HireVue Discontinues Facial Analysis Screening: Decision Reflects Re-Examination of Ai Hiring Tools. HRNews. 2021. Available online: <http://ezproxy.lb.polyu.edu.hk/login?url=https://www.proquest.com/trade-journals/hirevue-discontinues-facial-analysis-screening/docview/2486138234/se-2> (accessed on 20 September 2022).
37. Isbell, D.R.; Kremmel, B. Test Review: Current options in at-home language proficiency tests for making high-stakes decisions. *Lang. Test.* **2020**, *37*, 600–619. [CrossRef]
38. Richardson, M.; Clesham, R. Rise of the machines? The evolving role of AI technologies in high-stakes assessment. *Lond. Rev. Educ.* **2021**, *19*, 1–13. [CrossRef]
39. Brooks, V. Marking as judgment. *Res. Pap. Educ.* **2012**, *27*, 63–80. [CrossRef]

40. Rhead, S.; Black, B.; Pinot de Moira, A. Marking Consistency Metrics. 14 November 2016. Available online: <http://dera.ioe.ac.uk/id/eprint/27827> (accessed on 13 September 2022).
41. Moreno-Murcia, J.A.; Torregrosa, Y.S.; Pedreño, N.B. Questionnaire evaluating teaching competencies in the university environment. Evaluation of teaching competencies in the university. *J. New Approaches Educ. Res.* **2015**, *4*, 54–61. [[CrossRef](#)]
42. Yong, Q. Application Analysis of Artificial Intelligence in Oral English Assessment. *J. Phys. Conf. Ser.* **2020**, *1533*, 032028. [[CrossRef](#)]
43. Hadi, M.A.; Alldred, D.P.; Closs, S.J.; Briggs, M. Mixed-methods research in pharmacy practice: Recommendations for quality reporting (part 2). *Int. J. Pharm. Pract.* **2013**, *22*, 96–100. [[CrossRef](#)] [[PubMed](#)]
44. Compeau, L.D.; Franke, G.R. Book Review: Handbook of Mixed Methods in Social & Behavioral Research. *J. Mark. Res.* **2003**, *40*, 244–245. [[CrossRef](#)]
45. Microsoft. Characteristics and Limitations of Pronunciation Assessment. 6 September 2022. Available online: <https://learn.microsoft.com/en-us/legal/cognitive-services/speech-service/pronunciation-assessment/characteristics-and-limitations-pronunciation-assessment#comparing-pronunciation-assessment-to-human-judges> (accessed on 6 September 2022).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.