The Multimodal Effects of Voice-Based Asynchronous Technology-Mediated Communication on EAP Speaking Performance

Soroush Sabbaghan, Murray Peglar, & M. Gregory Tweedie

This study investigates how using a voice-based Asynchronous Technology-Mediated Communication framework (ATMC) can affect English for Academic Purposes (EAP) students' fluency, accuracy, and intelligibility. Research has highlighted the benefits of ATMC in improving various elements of spoken communication, but to date, its application to EAP contexts has been little studied. In EAP speaking/listening courses at a large Canadian university, a team of instructor-researchers developed a system to provide learners more speaking and listening practice opportunities and to give individualized video feedback (veedback) on speaking performance. The instructor-researchers used a practitioner enquiry research methodology to investigate the effectiveness of the voice-based ATMC for 14 participants, drawing upon qualitative data from student interviews and the instructor's qualitative and quantitative assessment of the students' responses to tasks. Findings indicate that the system resulted in overall gains in fluency, accuracy, and intelligibility due to a combination of the voice-based ATMC design, repetition and practice, and the veedback. Qualitative comments from participants indicate personalized feedback led to increased motivation. The framework described in this article, therefore, represents several important benefits for the application of voice-based ATMCs in EAP classrooms.

La présente étude examine les façons dont le recours à un cadre de communication asynchrone au moyen de la technologie (ATMC) axé sur la voix peut affecter la maîtrise de la langue, la précision et l'intelligibilité chez les apprenantes et apprenants d'anglais académique. Les recherches ont fait ressortir les avantages de la communication asynchrone au moyen de la technologie pour l'amélioration de divers éléments de la communication orale, mais cette méthode d'enseignement a été peu étudiée jusqu'ici dans des contextes d'anglais académique. Les membres d'une équipe de professeurs-chercheurs actifs dans l'enseignement de l'utilisation et de l'écoute de l'anglais académique dans une grande université canadienne ont élaboré un système permettant de fournir aux apprenantes et apprenants davantage d'occasions d'apprendre à utiliser et à comprendre l'anglais académique en mettant à leur disposition un système de rétroactions personnalisées par vidéo (veedback) qui leur permet de constater la justesse de leur expression orale. L'équipe de professeurs-chercheurs a utilisé une méthodologie de recherche basée sur l'interrogation des praticiens pour examiner l'efficacité de la communication asynchrone au moyen de la technologie axée sur la voix chez huit participantes et

six participants à l'aide de données qualitatives obtenues lors d'entrevues avec ces derniers et d'évaluations qualitatives et quantitatives fournies par les enseignantes et enseignants relativement à la réponse des étudiantes et étudiants aux tâches qui leur étaient assignées. L'étude a permis de constater que le système produisait des avantages globaux dans les domaines de la maîtrise, de la précision et de l'intelligibilité de l'anglais académique grâce aux effets combinés de la conception du système de communication asynchrone au moyen de la technologie axé sur la voix, à la répétition, aux exercices et au veedback. Les commentaires qualitatifs fournis par les participantes et les participants indiquent que la rétroaction personnalisée faite par le système a eu un effet positif sur leur motivation. C'est donc dire que le cadre décrit dans le présent article représente plusieurs avantages importants relativement à l'utilisation d'un système de communication asynchrone au moyen de la technologie axé sur la voix dans les classes d'anglais académique.

Introduction

To achieve academic success in English-medium universities, having a good command of English is critical. Both intellectual development and social participation are part of academic life and require excellent oral communication skills. In the Canadian context, examples include participating in and contributing to discussions, asking and responding to questions in lectures, presenting formal oral presentations on academic topics, and completing problem and project-based learning tasks (Cheng et al., 2004). These needs provide strong reasons for English for Academic Purposes (EAP) programs to offer a range of opportunities for students to develop oral communication skills.

Although many language programs are well-intentioned, international English as a second language (ESL) students' learning experiences are not always positive. For instance, research at one Canadian university has noted "lost opportunities for incidental and independent acquisition of English" as an inadequacy of EAP programs (Cheng et al., 2004, p. 51). An internal analysis of student evaluations of EAP speaking/listening courses at a large Canadian university revealed that one common difficulty for students is the lack of opportunities to employ the language knowledge acquired in practice, which is a challenge echoed in other EAP contexts (Ahn & Lee, 2016; Chen, 2011; Dooey, 2010; Hsu, 2016; Huang, 2015; Sun, 2012). Some students also indicated that even when opportunities arise, they might also hesitate to practice in class due to feelings of anxiety and inadequacy (see Bernales, 2016; Gregersen & Horwitz, 2002; Tatar, 2005). Our internal analysis revealed that students were unwilling to speak the target language due to instructor-

related factors such as restricted timing and lack of rehearsal opportunities, as reiterated by English as a foreign language (EFL) students in other studies (Pawlak & Mystkowska-Wiertelak, 2015; Zarrinabadi, 2014). Finally, some students noted that owing to the large number of students in their class, feedback offered by the instructor was sometimes "improper," inefficient, demotivating, or anxiety-provoking.

The current study attempts to enhance the student learning experience, develop student oral abilities, and offer a holistic solution to remedy the issues raised by students in our EAP listening/speaking classes. The first step was to find or modify a framework that gave our EAP instructors the opportunity to create and develop modes of learning that offer students opportunities to enrich their speaking/listening opportunities both inside and outside of class. In our review of the literature, we noted that technology-mediated asynchronous communication offers some interesting affordances. Because students work in isolation and self-manage their usage, asynchronous computer-mediated communication (CMC) can reduce levels of anxiety and fear of inadequacy compared with a synchronous communication mode (Buckingham & Alpaslan, 2017). Asynchronous CMC permits more planning time, which can positively influence the quality and quantity of oral production (Li et al., 2015). With asynchronous CMC, students can also make as many recordings as they like, and when reviewing their voice recordings, they benefit from extra listening practice (Buckingham & Alpaslan, 2017). Making multiple voice recordings fosters the development of learning strategies such as monitoring evaluation of content, organization, and language-usage (Sun, 2009). Pairing asynchronous CMC with a learning management system allows the students to store evidence of their speaking abilities, giving the instructor opportunities to offer more individualized feedback, which in turn seems to increase learners' speaking confidence (Hsu et al., 2008).

Owing to its many advantages, we developed a framework based on asynchronous CMC. The framework involves multiple iterations of students providing comprehensible output in response to comprehensible input and receiving multimodal feedback from their instructor (see Figure 1). As the medium of student output and feedback does not always require a computer, we labelled our framework voice-based Asynchronous Technology-Mediated Communication (ATMC).
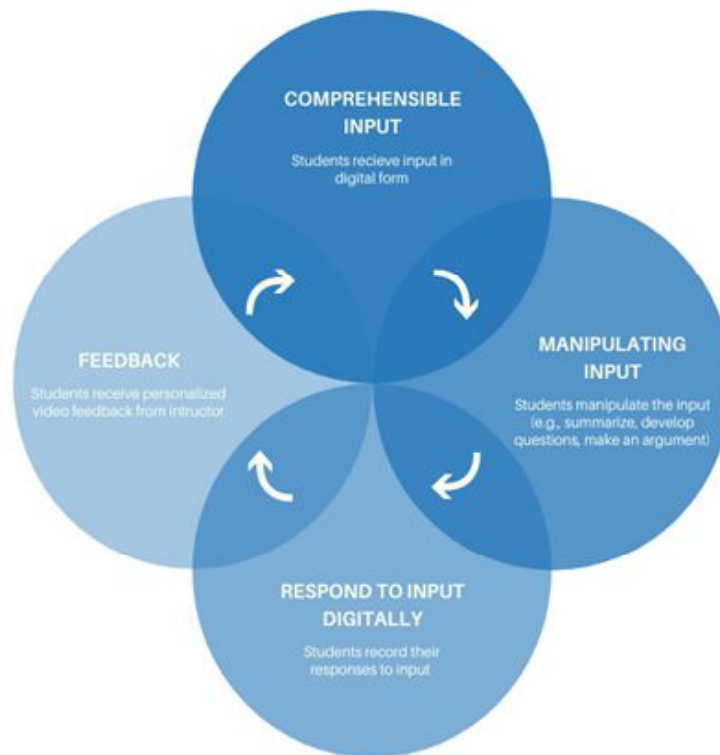
Figure 1. ATMC conceptual framework.
Note. ATMC = Asynchronous Technology-Mediated Communication.

Theoretical Underpinnings

Voice-based ATMC is grounded in Second Language Acquisition (SLA) theory, and its instructional techniques draw on strong- (see DeKeyser, 2007), non- (see Krashen, 1982), and weak- (see Ellis, 2008) interface positions. One of the primary benefits of voice-based ATMC is that it offers students further opportunities for practice, which is an important component of the strong interface position. As exemplified in our voice-based ATMC, practice involves students providing some form of output related to the input they receive (Leow, 2007); furthermore, DeKeyser (2007) has collected an array of articles discussing and empirically supporting the importance and merits of such practice in applied linguistics and cognitive psychology. As a condition of language acquisition in EAP contexts, Krashen (1982) has argued that the input students receive should be comprehensible to bridge the gap between the language class and the academic mainstream. Comprehensible input is a requirement of the noninterface position, and our voice-based ATMC framework contains this component. A second condition is that the output also needs to be comprehensible. Swain (1985) has maintained that when students are obliged to use their knowledge in a productive way, such as explaining a concept, they might modify a previous utterance or test new forms. In other

words, comprehensible output facilitates hypothesis testing about the structures and meanings of the second language (L2). However, if the students do not receive feedback to verify these hypotheses, they are unlikely to develop accurate and fluent production (Swain, 1995). In our framework, students are consistently required to manipulate the comprehensible input they receive in particular ways and produce output, on which they receive feedback. We believe this aspect of our framework is a component of the weak interface position.

Literature Review

Similar to the ATMC design of our study, previous studies on the development of language proficiency commonly used asynchronous CMC, such as voice-blogs or podcasts. Students were asked to create and post voice recordings as a form of extra practice, particularly in contexts where they had limited opportunities for oral practice in their classrooms. Moreover, while none of the literature specifically and only addressed fluency, accuracy, and intelligibility, these studies treated aspects of spoken output that are complementary, overlapping, or analogous to these focuses of our study.

Sun (2012) and Hsu (2016) investigated whether asynchronous CMC was effective in enhancing oral complexity, accuracy, and fluency among university-level EFL students. Neither study reported significant overall gains in oral skills; however, Hsu reported some improvement in students' oral complexity. Both studies noted that the informal, personal nature of voice blogging might be a factor in why overall improvement was not observed. In both studies, students were not encouraged to monitor their output for accuracy, nor was feedback offered to students. Moreover, Sun did not require her participants to submit multiple voice recordings intermittently over a predetermined period. Therefore, some of her participants submitted many of their recordings at the end of the semester over a short period of time, thus reducing the opportunity for practice and remediation between outputs.

Engin (2014) conducted a study measuring spoken accuracy using asynchronous CMC, wherein students were asked to create English writing explanations (tutorials) for other students in video format. Due to the shift in teaching responsibility, Engin hypothesized that this practice would compel students to produce linguistically accurate explanations. At the end of the 13-week course, she interviewed participants and administered questionnaires focusing on students' perception of their ability to do so. Engin reported that, overall, students believed that the video activity helped improve their linguistic accuracy. However, the study does not clarify which aspects of the students' accuracy improved. Moreover, as measures of accuracy were self-reported by the students, more reliable and trustworthy sources are needed to determine how speech became more accurate through video-recording.

Ducate and Lomicka (2009) investigated the effect of podcasts, as asynchronous CMC, on the intelligibility of German and French. In their study, students produced both scripted and improvised voice recordings at regular intervals, and overall, they reported positive effects in both languages. In the study, students received corrective feedback on some submitted recordings. However, a significant improvement in pronunciation ratings was not noted between the beginning and end or (in most cases) between the intervals during the semester. The authors attribute this lack of improvement to a ceiling effect. Specifically, they note participating students had relatively high ratings of their pronunciation prior to the study, and, therefore, posit that practice and corrective feedback no longer significantly influenced pronunciation.

Lepore (2014) linked employing asynchronous CMC to positive effects on pronunciation. In this study, 39 foreign language students of French used VoiceThread to produce three audio recordings in response to instructor-created prompts. In addition, they were also asked to comment on one another's recordings. Students also received feedback from the instructor, who commented on students' accuracy, fluency, and comprehensibility. Students completed a self-evaluation survey after reviewing the instructor's feedback and rated their pronunciation abilities in each performance against those of classmates. They were also allowed to gauge how difficult or easy they perceived the pronunciation demands to be and identify which aspects of their pronunciation were strongest and weakest. Unfortunately, as Lepore's results are solely based on untrained student perception, the findings are questionable. The rating system Lepore used did provide guidance on which aspects of pronunciation should be rated. Therefore, changes in pronunciation might have been constructed (in the minds of the students) as improvements because no formal measure of pronunciation was offered.

ATMC similar to that in our study has been shown to improve fluency. Gromik (2012) conducted a case study requiring participants to produce weekly 30-s videos on their cell phones on teacher-selected topics over 13 weeks. He reported a 37% increase in fluency when comparing the average speech production in the first and final weeks. Gromik attributed the general increase in speech rate to the repeated production of asynchronous videos. However, as Gromik only considered the speech rate of very short videos (30-s clips), it is unclear whether producing longer video output would result in a similar outcome.

In sum, ATMC and asynchronous CMC have been shown to promote oral language gains. However, as discussed above, the studies that have reported these gains leave some questions unanswered. Therefore, further investigation with more rigorous research methods are required to confirm claims offered by previous studies, and provide focus on ATMC rather than asynchronous CMC. The purpose of the present study is to investigate the effects of voice-based ATMC on fluency, accuracy, and intelligibility in a Canadian EAP context. In SLA literature, accuracy is usually associated with

the avoidance of error during production or avoidance of challenging structures that might provoke error, and is an indicator of the increase of control over newly acquired language, while fluency (Riggenback, 1991) is an indicator of advanced control of L2 language structures to emphasize meanings (Skehan, 1991; Skehan & Foster, 1999). Unfortunately, intelligibility is not always consistently defined in the literature. Nonetheless, seminal research in the field points to it being the degree to which an utterance is understood by the listener (Derwing, 2010; Murphy, 2014). These conceptualizations of target language production underlie the present study, and give rise to three research questions:

Research Question 1 (RQ1): Does voice-based ATMC improve fluency?

Research Question 2 (RQ2): Does voice-based ATMC improve accuracy?

Research Question 3 (RQ3): Does voice-based ATMC improve intelligibility?

Methodology

Practitioner Enquiry

Our study required a pragmatic research approach, where the instructor and researchers could work together such that the instructor could be fully involved in the research. An approach was needed where the teacher's reflections could be accommodated and integrated while data were being collected to inform ongoing practice. Practitioner enquiry (Reid, 2004) is such an approach, and is positioned between teacher reflection and action research. It places more emphasis on bringing extant research to bear on teaching practice than some of the more inward-looking approaches of teacher reflection and does not carry with it the same emancipatory expectations associated with some forms of action research. A defining feature of a practitioner enquiry approach includes "systematic inquiry made public" (Baumfield et al., 2013, p. 8), as the circle of data gathering, planning, and action is placed within a "critical community" (p. 6) where practitioners and researchers forge effective partnerships, and boundary lines between theory and practice "are challenged and expertise is distributed as teachers as researchers and researchers as teachers learn together" (p. 11). Practitioner enquiry is further distinguished by its emphasis on student voices as a central element of informing research and reflection for classroom practice. These emphases on the reporting of inquiry, the practitioner-researcher partnership, and the central place of student voice made practitioner enquiry an ideal approach for our particular study context.

Participants

Participants in this study were 14 non-English major ESL learners enrolled in an EAP listening/speaking course at a large, research-intensive Canadian university. All students were nonnative speakers of English who had taken an internal English proficiency test and had obtained a score at the Common European Framework of Reference (CEFR) B2 level. The class in which we conducted our study had 25 students, but only 14 gave consent to use their data for research purposes. Of the 14, eight were female and six were male. The participants were from different parts of the world, including China, Chile, Iran, Mexico, Kuwait, and Saudi Arabia. All were between 18 and 22 years old, and none had experience studying abroad in a target-language embedded context (Graves, 2008).

Context of Study

The focus of the listening/speaking course in which the participants had enrolled is to prepare learners with the oral/aural skills necessary for success in university. The course is offered pre-enrollment, and its successful completion is a prerequisite for entrance to students' chosen degree programs. The course has set outcomes, including presenting appropriate and summative content, asking appropriate questions in seminar/lecture situations, and demonstrating critical thinking. These three outcomes were selected as most suitable for asynchronous CMC, and as relating well to the academic competencies needed by EAP students.

Our voice-based ATMC framework was delivered on the Desire2Learn (D2L) platform, as the students were already familiar with it because it is the official Learning Management System (LMS) of the university. Using D2L is advantageous because students could focus more on the learning task and less on how to use the technology. Furthermore, D2L has a built-in video and audio recording mechanism that saves recorded files within the LMS. The equipment needed for students included a computer or tablet with headset microphones, or a cell phone. Instructors and students were able to independently and freely download both the digital input files and their recorded response output files to their personal media players. D2L also allows the instructor to develop tailor-made rubrics and gives students continued access to written and oral feedback on assignments, which is key to student learning because feedback is specific to the task but can also be contextualized within a progression through task iterations.

Procedure: Cycles of Enquiry

Pre-course preparation.   The teacher-researcher team was guided by a number of questions as they developed a syllabus built around voice-based ATMC: (a) How can technological mediation be used to improve fluency, accuracy, and intelligibility? (b) How will students receive instructor feed-

back? (c) How can fluency, accuracy, and intelligibility be operationally defined? In response to these questions, a plan was developed and undertaken in advance of the course's commencement. The teacher-researcher team began by preparing academic lectures. All lectures are at the higher end of the Vantage or Upper intermediate level (i.e., B2+) on the CEFR, reflecting the admission requirement of the university. In addition, due to class time constraints, lectures needed to be between 4 and 6 min long. We were unable to find online video sources matching our criteria. Therefore, we adapted content from open-source academic textbooks. At the outset of the course, students were surveyed regarding their preferences for lecture topics, drawn from a list of required first-year university courses. Based on the results of this survey, texts were selected from the fields of Astronomy, Biology, Canadian Studies, Chemistry, Economics, Engineering, Geography, and History. We then adapted the lexical profiles of the selected texts to meet the required proficiency level using the Text Inspector online lexical analysis tool (Text Inspector, 2018). Modified texts (see Appendix A) were then recorded as video lectures. The teacher-researcher team also prepared separate guides modelling each of the students' three recording tasks within the voice-based ATMC framework, as well as analytic rubrics to evaluate each (see Appendix B for sample evaluation). Task rubrics were developed by the research team, as no previously validated rubrics for this specific type of ATMC system were readily available. Scores given based on the rubrics were used strictly for course assessment and were not used for research purposes due to ethics limitations.

Task types. Three weekly recording tasks (total of 13 weeks) were assigned to students: Lecture Summary Task, Jeopardy Questions Task, and Debate Task.

The Lecture Summary Task, corresponding with the course outcome of presenting appropriate and summative content, required students to record a short video (the LMS limits recordings to a 3-min maximum) summarizing the content of the video lecture they had viewed. The summary format typically involved defining the topic and recapping two supporting ideas with examples (see Appendix C).

The Jeopardy Questions Task was aligned with the outcome of asking appropriate and relevant questions in seminar/lecture situations. Students were supplied with four answers from the video lecture and required to develop corresponding questions using grammatically accurate question forms and intelligible intonation. Question foci included specification, clarification, exemplification, and application (see Appendix C).

In the Debate Task, students were required to record two responses, defending two opposing sides of an argument, corresponding with the outcome of demonstrating critical thinking. An example of a Debate Task and sample script are included in the Appendix C.

Cycle 1: Weeks 1–7. The first week of the course was used to allow students to familiarize themselves with the course content, the LMS, and record-

ing and uploading process, so no performances were assessed. A member of the research team explained the research project to the students and obtained informed consent from students who volunteered to participate in interviews and complete surveys. The instructor was unaware of which students were participating, as per procedure stipulated by the university's research ethics body, though all students in the class completed all tasks as part of their course work.

In Week 2, the instructor began assessing and giving veedback on student performance (see Appendix B for student view of veedback). After students produced videos each week in response to the three Task types, the instructor selected videos for assessment and veedback—one for each prescribed Task—offering nine veedbacks to each student over the course of the semester. The students did not know which of their three weekly recordings would be graded and from their perspective, assessment seemed random. We did not inform students when they were to be assessed to maintain student motivation to complete each listening/recording Task. The instructor, however, had previously created an assessment schedule to ensure an equal and parsed assessment regimen.

Quantitative data collection (on fluency, accuracy, and intelligibility) began in Week 3, whereas qualitative data, which focused on students' perception of voice-based ATMC, was collected in Week 2. The anonymized qualitative data were shared in research team meetings with the course instructor, who then adjusted accordingly. For example, several students noted that the instructor's veedback tended to focus only on areas for improvement, overlooking student strengths. Upon receiving this feedback, the instructor began each piece of veedback by commenting on what the student had done well before suggesting areas that needed improvement.

Cycle 2: Weeks 8–13.   Week 8 marked the beginning of the second enquiry cycle. Based on student feedback, instructor input, and assessment scores, changes were made to the Lecture Summary Task design. First, students were given longer lecture videos. Second, students were now required to add an additional element in the Lecture Summary Task, wherein they were asked to compare and contrast the subtopics offered in each input video. The second round of interviews with students started in Week 11. Students' comments were given in an anonymized form to the course instructor.

Measures

In our study, we employed a mixed-methodology (Tashakkoir & Teddlie, 2010) approach, as both quantitative (fluency, accuracy, and intelligibility) and qualitative (two semistructured interviews) measures of data were collected. Operational definitions of measures of fluency, accuracy, and intelligibility were developed to evaluate students' recorded output. The measures of fluency and accuracy are mostly the same as those used in other SLA studies

(see Ellis, 2009), and the measure for intelligibility was adapted from studies on this attribute (Derwing, 2010; Murphy, 2014).

Fluency. We defined fluency as the number of syllables per the last 1 min of speech produced by students in their recordings. We selected only the last 1 min of speech to include as data because we noticed most students began their recording using preformulations, thus including utterances made in the starting moments of recording would have skewed our results. This measure was chosen as it was used in other studies (e.g., Yuan & Ellis, 2003).

Accuracy. One global accuracy measure was adopted. Accuracy was operationalized as the percentage of lexical and grammatical errors in student-recorded productions, as used in a number of other studies (e.g., Knoch et al., 2015; McCormick & Vercellotti, 2013). This percentage was calculated by dividing the total number of grammatical and lexical errors per 100 words and extrapolating for the entire length of the recording.

Intelligibility. Intelligibility is the extent to which a listener comprehends the speaker (Derwing, 2010; Murphy, 2014). Based on this definition, and for our study, we operationalized intelligibility as the number of instances the instructor did not understand utterances when reviewing a student's recording.

Student-recorded entries for each of the three Tasks that were assessed in the third week of the course (A1), the beginning of Week 8 (A2), and at the end of the term (A3), were transcribed verbatim. In addition, instructor veedbacks for these responses were transcribed word-by-word (see Appendix B for example). The transcripts of student responses were input into the Text Inspector online lexical analysis tool (Text Inspector, 2018) to determine indexes of fluency based on the operational definition. The research team also used student transcripts to measure accuracy. For intelligibility, however, the transcripts of the instructor's veedback were used to count the number of instances in each veedback the instructor indicated that he did not understand the student's utterance.

We also conducted student interviews, aiming to gather qualitative data regarding student perceptions on engaging with our voice-based ATMC framework and the veedback they had received. Two 20-min audio-recorded semistructured interviews were conducted with seven volunteer students in Weeks 2 and 11. The first set of interview questions (see Appendix D) was focused on student perceptions of their oral abilities. The second set (see Appendix D) was focused on student perceptions of the effectiveness of the voice-based ATMC (veedback included) on their speaking performance. All interviews were recorded and transcribed. Guiding our analysis of the interview responses was the analysis model of Miles et al. (2019), which discusses transcription, theme identification, and results revision.

Data Analysis

Due to the small sample size, nonparametric procedures were used. Kruskal-Wallis, the nonparametric equivalent to the one-way ANOVA, was calculated to compare the three assessment instances (A1, A2, A3) for each of the three measures (fluency, accuracy, intelligibly) for each of the three Tasks (Lecture Summary, Jeopardy Questions, Debate). The alpha level for all group comparisons was set at .05. In case of significant differences, follow-up Dunn's pairwise tests were conducted.

Results

Results used for quantitative analysis were taken from student recordings and instructor evaluations, which focus on student performance. Qualitative data were gathered from transcripts of recorded interviews with students, which focused on student experience of using voice-based ATMC.

Quantitative Analysis

Fluency.

Lecture Summary Task. The Kruskal-Wallis test did not reveal a statistically significant difference between the number of syllables produced in the last 1 min of speech among the A1, A2, A3 student recordings, $H(2, N = 14) = 4.39$, $p = .11$.

Jeopardy Questions Task. We did not measure fluency for this Task as responses were almost always less than 1 min. Furthermore, as students recorded the question (though sometimes including context of the question), the productions contained chunks of speech. Therefore, the recording of this task could not offer any meaningful measure of fluency.

Debate Task. The Kruskal-Wallis test of the data collected about this Task indicated a statistically significant difference between the number of syllables produced in the last 1 min of speech among A1, A2, and A3, $H(2, N = 14) = 14.06$, $p < .00$. Dunn's pairwise tests were carried out for the three pairs of groups. There was evidence ($p = .001$, adjusted using the Bonferroni correction) of a difference between the number of syllables produced between A1 and A3, as well as between A2 and A3 ($p = .045$, adjusted using the Bonferroni correction) in the last 1 min of speech. There was no evidence of a difference between A1 and A2.

Accuracy.

Lecture Summary Task. The Kruskal-Wallis test of the data collected in this Task indicated a statistically significant difference in the percentage of lexical or grammatical errors among A1, A2, and A3, $H(2, N = 14) = 18.21$, $p < .00$. Dunn's pairwise tests were carried out for the three pairs of groups.

There was evidence (p = .001, adjusted using the Bonferroni correction) of a difference between the number of lexical and grammatical errors between A1 and A3. There was no evidence of a difference between the other pairs.

Jeopardy Questions Task. The Kruskal-Wallis test of the data collected in the Jeopardy Questions Task indicated a statistically significant difference between the percentages of lexical or grammatical errors among A1, A2, and A3, $H(2, N = 14) = 22.22$, $p < .00$. Subsequently, Dunn's pairwise tests were carried out for the three pairs of groups. There was evidence (p = .001, adjusted using the Bonferroni correction) of a difference between the percentage of lexical and grammatical errors produced between A1 and A3, as well as between the A2 and A3 assessments (p = .006, adjusted using the Bonferroni correction). There was no evidence of a significant difference between the percentage of errors made between A1 and A2.

Debate Task. The Kruskal-Wallis test of the data collected about the Jeopardy Questions Task indicated a statistically significant difference between the percentage of lexical or grammatical errors among the first, second, and third assessed recorded productions, $H(2, N = 14) = 20.46$, $p < .00$. Dunn's pairwise tests were carried out for the three pairs. There was evidence (p = .001, adjusted using the Bonferroni correction) of a difference between the number of lexical and grammatical errors between A1 and A3. There was no evidence of a difference between the other pairs.

Intelligibility.

Lecture Summary Task. The Kruskal-Wallis test of the data collected in this task indicated a statistically significant difference in the number of instances the instructor indicated that he did not understand an utterance in the recordings among A1, A2, and A3 recorded productions, $H(2, N = 14) = 9.55$, p = .008. Dunn's pairwise tests were carried out for the three pairs of groups. There was evidence (p = .006, adjusted using the Bonferroni correction) of a difference between the number of instances the instructor indicated that he did not understand an utterance between A1 and A3. There was no evidence of a difference between the other pairs.

Jeopardy Questions Task. The Kruskal-Wallis test did not reveal a statistically significant difference in the number of instances the instructor indicated that he did not understand an utterance in the recordings among A1, A2, and A3 recorded productions, $H(2, N = 14) = 3.57$, p = .168.

Debate Task. The Kruskal-Wallis test of the data indicated a statistically significant difference in the number of instances the instructor indicated that he did not understand an utterance in the recordings among A1, A2, and A3 student recordings, $H(2, N = 14) = 13.78$, p = .001. Dunn's pairwise tests were carried out for the three pairs of groups. There was evidence (p = .001, adjusted using the Bonferroni correction) of a difference between the number of instances the instructor indicated that he did not understand an utterance

between A1 and A3. There was no evidence of a difference between the other pairs.

Quantitative analysis summary.   In answer to RQ1 (focus on fluency), two Kruskal-Wallis tests were conducted. Data gathered from the Lecture Summary Task did not indicate a statistically significant difference between the number of syllables produced in the last 1 min of speech among the A1, A2, and A3 student recordings. However, statistical analysis of data in the Debate Task revealed a significant difference between the number of syllables produced between A1 and A3 as well as between A2 and A3. Table 1 displays the results of tests conducted to answer RQ1.

Table 1
Kruskal-Wallis Tests for Fluency

Tasks	Chi square	*df*	*p*
Lecture Summary Task	4.39	2	.11
Debate Task	14.06	2	.00*

Note. *$p < .05$

Three Kruskal-Wallis tests were conducted to answer RQ2 (focus on accuracy). The results indicated a statistically significant difference among the three data groups for all three Tasks, as displayed in Table 2.

Table 2
Kruskal-Wallis Tests for Accuracy

Tasks	Chi square	*df*	*p*
Lecture Summary Task	18.21	2	.00*
Jeopardy Questions Task	22.22	2	.00*
Debate Task	20.46	2	.00*

Note. *$p < .05$

Finally, to answer RQ3 (focus on intelligibility), three Kruskal-Wallis tests were performed. The results revealed a significant difference among the three data groups in the Lecture Summary Task and the Debate Task, but not for the Jeopardy Questions Task. Table 3 displays the results of the tests.

Table 3
Kruskal-Wallis Tests for Intelligibility

Tasks	Chi square	*df*	*p*
Lecture Summary Task	9.55	2	.008*
Jeopardy Questions Task	3.57	2	.168
Debate Task	13.78	2	.001*

Note. *$p < .05$

Qualitative Analysis

A review of the qualitative data (inductive coding) from interview transcripts revealed four major themes in students' perceptions of our voice-based ATMC. First, students stated that the framework helped them develop their speaking; in particular, they felt motivated to apply the instructor's feedback because it was specifically tailored to their individual needs. This increase in motivation may be one contributing factor to why the frequency of speaking errors decreased as the term unfolded. Second, students stated that the framework facilitated learning when compared with written feedback alone. They noted that with veedback, they developed a better understanding of their errors because they were able to relate the information to a specific moment in a particular Task and receive targeted, modelled correction. Third, students interviewed noted that voice-based ATMC helped them formulate accurate questions. As a result, students felt more confident about asking more specific and targeted questions and reported that they were now more likely to raise questions during class.

Students also revealed some challenges when engaging with our voice-based ATMC. The first challenge they raised was a technical issue. Students noted that the uploading process on the D2L shell sometimes corrupted their files. The second challenge was specifically directed toward the Debate Task. Some felt uncomfortable arguing for a belief they did not agree with. They noted that as a result, they might have spoken more quickly or used fewer words to complete the Task. The final emergent theme is specific to the feedback. While many students noted veedback improved their pronunciation and intonation, some stated that when the instructor commented on grammatical mistakes, he did not offer references or remedies to help students avoid these mistakes.

Discussion

Our voice-based ATMC collected students' oral responses to input (video and text) chronologically, which allowed the instructor to observe which components of L2 performance were being developed through the system and provide veedback. Results show overall benefits to speaking traits of fluency, accuracy, and intelligibility, as well as generally positive perceptions of voice-based ATMC. The suggested causes for improvement are repetition and practice, and the veedback given by the instructor.

Fluency

Speaking fluency did not show gains in the Lecture Summary Tasks. This may have been because the complexity of the Task design was increased at the beginning of the second cycle, thus, becoming more cognitively demanding. Moreover, there is the possibility that either fluency was not problematic

for our participants at their proficiency level, or that they had hit a ceiling (Ducate & Lomicka, 2009), or that the rehearsal afforded by the opportunity to re-record effectively addressed fluency challenges, as reported by Li et al. (2015). Another possibility is that veedback itself may have inadvertently limited fluency development due to student self-awareness and increased focus on other speaking traits more specifically noted by the instructor.

Fluency did show significant gains in the Debate Task. Participants found the Task relevant to their studies, in that they "can use [it] daily because we are always debating in a positive way or giving our opinion about topics," which may have augmented or reinforced the fluency gains measured for this Task. Participants also tended to find this Task "really hard. Like when you don't support the idea it's really hard for me to prove that idea somehow and you have to struggle with yourself," but that "it makes you more flexible." Such agility and flexibility underpin fluency, are required in the oral communication skills outlined by Cheng et al. (2004), and reduce in-class anxiety over real-time communication (reverberated in Bernales, 2016; Gregersen & Horwitz, 2002, Tatar, 2005). Overall, Swain's (1995) relation of students' hypothesis-testing leading to fluent production, on the condition that crucial feedback is given, appears to have been exhibited in the Debate Task within the voice-based ATMC design.

Accuracy

Grammatical accuracy, however, did show significant gains across all Tasks. Our voice-based ATMC design developed accuracy both through the progression of iterations (i.e., the summaries "didn't challenge me at first, just copying at the beginning . . . now you have to actually focus"), and the targeted veedback given by the instructor. The veedback medium did not preclude specific grammar correction, as one student noted, "if you have a problem in grammar he's [the instructor's] like in 2:36, you said this." Veedback was also supported by textual outlines of the instructor's corrections (see Appendix B), though students indicated that targeted references and remedies to grammatical errors were still lacking. This was likely due to navigating a new multimodal mode of providing feedback, all while under the eye of the researchers, yet does point to a lingering concern about providing timely and detailed veedback to large classes. Nonetheless, that the voice-based ATMC framework, as an asynchronous CMC, did effect gains in student output accuracy aligns with the effects of increased planning time (Li et al., 2015), regardless of feedback. Moreover, the veedback and textual feedback given recorded specific aspects of students' linguistic accuracy, thus, lending more targeted and trustworthy feedback tools to previous research in asynchronous CMC accuracy (Engin, 2014).

Intelligibility

The last academic oral competency measured, intelligibility, also showed gains through the Debate and Lecture Summary Tasks. Students' intelligibility was supported by the Task structures, for example, requiring clear diction of key vocabulary. Perhaps most importantly, veedback gave students effective correction through modelling and comparison with their output, and both could be reviewed and targeted in audio, video, and text media. Although textual feedback complemented veedback, students noted, "It's easier to understand when you see the expression of the face and the tone" through veedback, and highlighted the role of instructor modelling, as his "English is good so I like that. To repeat it where you have to stress each and every word, where he is stressing that." Students' increased intelligibility could also be attributed to facility with recurring or targeted academic language; nonetheless, this does not detract from the gains made, which focus specifically on such needs required by EAP students. In addition, it did not appear that the intelligibility of students was mitigated or reduced by the medium, or the Lombard effect, whereby ambient noise affects pronunciation (Cooke & Lecumberri, 2012; Lecumberri et al., 2010). Either of these could have confounded findings if intelligibility were reduced due to technology-generated or ambient noise. Student intelligibility gains were not limited by the ceiling effect discussed by Ducate and Lomicka (2009), and benefitted from verification by the instructor, rather than peers (Lepore, 2014). Thus, as the culmination of each week's tasks within the voice-based ATMC framework, veedback provided useful, targeted feedback on students' intelligibility throughout the Tasks and did not appear to be negatively affected by the system design or medium.

Limitations

As the current study is exploratory, we advise caution with the interpretation of our findings. Although the class in which we conducted our study contained 25 students, only 14 gave us consent to use their recordings for research purposes. Therefore, statistical power is low, and there is a probability of making a Type II error. For instance, the results did not indicate a significant improvement in fluency in the Lecture Summary Task, but in the Debate Task, gains were recorded. Time was another limiting factor. Data were gathered over 11 weeks, which may not have been a sufficient amount of time for students to develop fluency because it is generally considered to improve as a result of automation (De Bot, 1996). Finally, a control group was not included in our study. Although improvements were observed in fluency, accuracy, and intelligibility, factors other than our design, repetition, and practice (e.g., a learning effect from other courses students were studying), probably played a role. Adding a control group with similar language

SOROUSH SABBAGHAN, MURRAY PEGLAR, & M. GREGORY TWEEDIE

background and proficiency would likely offer better conditions to control hidden variables.

Implications for Next Steps

Fluency.   As no statistically significant fluency gains were found in the Lecture Summary Task, modifications to the design are warranted. For example, reducing the time for preparation or increasing the amount or complexity of content may counter a ceiling effect, a connection noted by Robinson (2007). To increase complexity, this Task could align with more relevant functions and content in the students' fields of study.

The Jeopardy Questions Tasks were not assessed for fluency, as the utterances were too short; however, qualitative feedback noted affective gains in students' willingness and ability to ask questions in class. Altering the Task could assess a different metric of fluency, such as the time taken to generate a response.

More targeted attention to fluency could improve instructor feedback. Veedback allows the instructor to note stalled output in phrasing while modelling more fluent delivery. Overall, as fluency gains were likely the result of automaticity and practice acquired through voice-based ATMC, students could be further challenged by increasing the Task difficulty and/or decreasing the time given.

Accuracy.   While significant gains in accuracy were recorded, modifications to the voice-based ATMC design might increase accuracy gains even further. Students relied more heavily on textual feedback of accuracy, while veedback was seen as complementary. Our framework afforded opportunities to assess accuracy, but more complex and specific rubrics could tease apart grammatical or lexical issues and focus on a greater range of errors. Instructors could target specific errors such as verb endings, which have been shown to benefit from pushed oral output (Mamaghani & Birjandi, 2017), or complexity, as treated by Hsu (2016), and communicate feedback through a more specific rubric. Moreover, students (McCormick & Vercellotti, 2013) and instructors could note areas for improvement through textual self-reflection and feedback, while still taking advantage of the personalization and affective gains offered through veedback.

There were several ancillary benefits to accuracy supported by the framework. The design afforded ample opportunities to re-record, which is a limitation of traditional classroom instruction. Even in large classes, all students are given the opportunity to speak, practice, self-assess, and receive rich feedback, thus, moving beyond the limitations of textual approaches to grammar and lexis. Students appreciated this approach, and although they focused on textual feedback for accuracy, they reported that the complementary veedback was more personalized and considerate. Therefore, veedback on accuracy must be personal, targeted, and actionable.

Intelligibility.   As pronunciation gains are one of the central outcomes of voice-based ATMC, the framework would benefit from more specific, targeted attention in this realm, beyond intelligibility. For example, the role of comprehensibility and accentedness also has effects on listener ratings, and Munro and Derwing (1995) have shown skilled assessors can distinguish between all three. Although EAP instructors have been found to be more sympathetic to nonnative English speakers' intelligibility, they may be less biased than content faculty in assessing comprehensibility (Sheppard et al., 2017). Because of this, future veedback could evolve from focused attention on basic intelligibility into assessing more complex communication of meaning and comprehensibility, separate from accentedness. Moreover, intelligibility and comprehensibility could be integrated into a more robust pronunciation rubric, including suprasegmental phrasing, and physical delivery issues. Delivery would benefit from larger camera pan focus to give the opportunity for feedback on stance, gesture, and enthusiasm, mimicking classroom presentations more closely. There is also room to develop more structured feedback on the rhetorical tone of the students' recordings. For example, the Debate Task could invoke convincing, inviting, and disagreeing tones and expressions, which would employ the advantages of rich modelling and targeted review that veedback affords.

Due to the power of voice-based ATMC in supporting pronunciation, a trained instructor is required, though accentedness should not be an issue (Munro & Derwing, 1995). Specifically, veedback offers a strong medium for instructors to recognize and model intelligible, comprehensible, and animated speech while also timing, replaying, and modelling feedback for student utterances. These benefits are not limited to instructor feedback, and self- and peer-evaluation is recommended for even more reflective and aural practice.

Embedding contextualized input.   The opportunity to develop additional, and more targeted, tasks may also allow voice-based ATMC to cross curricula and to target specific skills and subjects in context. Perhaps the most useful extension would be the integration of other academic skills, content, and courses. Students in English for Specific Purposes (ESP), Pathways, or other content-based courses could benefit from the integration of language and content support. For example, while some students noted that veedback on the "pronunciation of technical words" was not useful because "I have never heard it and heard it for the first time," comprehensible input could include contextualized vocabulary instruction so as to blend content and language.

Students might also benefit when the comprehensible input in a voice-based ATMC design is used in university and course orientation presentations. Course faculty could develop videos introducing themselves and their research areas, explain institution infrastructure and policy, and even initiate academic advisory and counselling interactions to create both meaning and

inclusion within the class. The EAP instructor could then use these videos in a voice-based ATMC framework, in which the task could include summarizing and asking questions about the materials presented, or extending and applying the material in moderated online discussion and other tasks. In such a setup, students will be given the opportunity to develop their language skills while also learning important information about course structure and campus life.

Beyond the classroom. The voice-based ATMC framework outlined in this study gives instructors a powerful tool to push and assess language remotely. This design is advantageous in EFL contexts where pronunciation instruction may be lacking. Given that textual instruction is ubiquitous and more easily taught by local instructors, the framework can complement such instruction by enabling access to native-speaking instructors remotely. Moreover, the system effectively democratizes communication, allowing equal access for all students with access to online communication. In particular, this system can be useful for marginalized students who feel unable to speak openly or equally due to personal or cultural barriers, and it may serve to share a culture of learning beyond the language itself.

Overall, voice-based ATMC can easily be modified to allow for a range of applications. New tasks and more nuanced rubrics could identify discrete elements of speaking beyond the fluency, accuracy, and intelligibility assessed in this study. Furthermore, tasks can easily be adapted to different language levels and outcomes. In our context, summarizing, questioning, and debating skills are determined to be critical to postsecondary studies, as noted by the students themselves: "I see like other programs . . . they listen more than they speak, and they just give presentations, but they didn't do Lecture Summary, Jeopardy Questions or Debate so this thing I think is very useful academically in university."

Conclusion

The current study explored the effectiveness of a voice-based ATMC framework on students' fluency, accuracy, and intelligibility in an academic context. Our findings indicated that engagement with voice-based ATMC resulted in overall improvement in these language traits. The causes for these improvements are due to the task design, repetition, and practice, and the veedback given by the instructor. The framework can be modified in terms of task design, feedback, and the medium used for different academic skills, content, and contexts. We believe the technology will continue to play an important role in EAP development, and that this study provides new insight into the practicality and affordances of voice-based ATMC frameworks in the EAP context, and, indeed, other learning contexts that will benefit from rich, democratized, and formalized communication and feedback.

The Authors

Soroush Sabbaghan is a senior instructor in the Learning Sciences specialization at the Werklund School of Education, University of Calgary. His research foci in English as a second language (ESL) are mainly in the impact of various awareness and attention levels on the quality of second language (L2) writing and the nature of complexity in task design.

Murray Peglar is an instructor in the Language and Literacy specialization at the Werklund School of Education, University of Calgary. His research interests include internationalization; teacher training and development; clear pronunciation; and writing focus and development.

M. Gregory Tweedie is an associate professor in the Language and Literacy specialization area at the Werklund School of Education, University of Calgary. His teaching and research draw heavily upon his experiences as a language teacher and language teacher trainer in East, Southeast, and Central Asia; the Middle East; Canada; and his native Australia.

References

Ahn, T., & Lee, S.-M. (2016). User experience of a mobile speaking application with automatic speech recognition for EFL learning. *British Journal of Educational Technology, 47*(4), 778–786. https://doi.org/10.1111/bjet.12354

Baumfield, V., Hall, E., & Wall, K. (2013). *Action research in education: Learning through practitioner enquiry* (2nd ed.). London: Sage. https://doi.org/10.4135/9781526402240

Bernales, C. (2016). Towards a comprehensive concept of Willingness to Communicate: Learners' predicted and self-reported participation in the foreign language classroom. *System, 56*, 1–12. https://doi.org/10.1016/j.system.2015.11.002

Buckingham, L., & Alpaslan, R. S. (2017). Promoting speaking proficiency and willingness to communicate in Turkish young learners of English through asynchronous computer-mediated practice. *System, 65*, 25–37. https://doi.org/10.1016/j.system.2016.12.016

Chen, H. H.-J. (2011). Developing and evaluating an oral skills training website supported by automatic speech recognition technology. *ReCALL, 23*(1), 59–78. https://doi.org/10.1017/S0958344010000285

Cheng, L., Myles, J., & Curtis, A. (2004). Targeting language support for non-native English-speaking graduate students at a Canadian university. *TESL Canada Journal, 21*(2), 50. https://doi.org/10.18806/tesl.v21i2.174

Cooke, M., & Lecumberri, M. L. G. (2012). The intelligibility of Lombard speech for non-native listeners. *The Journal of the Acoustical Society of America, 132*(2), 1120–1129. https://doi.org/10.1121/1.4732062

De Bot, K. (1996). The psycholinguistics of the output hypothesis. *Language Learning, 46*(3), 529–555. https://doi.org/10.1111/j.1467-1770.1996.tb01246.x

DeKeyser, R. M. (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge: Cambridge University Press.

Derwing, T. (2010, January). Utopian goals for pronunciation teaching. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 1st Pronunciation in Second Language Learning and Teaching Conference* (pp. 24–37). Ames: Iowa State University.

Dooey, P. (2010). Students' perspectives of an EAP pathway program. *Journal of English for Academic Purposes, 9*, 184–197. https://doi.org/10.1016/j.jeap.2010.02.013

Ducate, L., & Lomicka, L. (2009). Podcasting: An effective tool for honing language students' pronunciation? *Language Learning and Technology, 13*(3), 66–86. Retrieved from http://llt.msu.edu/vol13num3/ducatelomicka.pdf

Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford: Oxford University Press.

Ellis, R. (2009). The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied Linguistics, 30*(4), 474–509. https://doi.org/10.1093/applin/amp042

Engin, M. (2014). Extending the flipped classroom model: Developing second language writing skills through student-created digital videos. *Journal of the Scholarship of Teaching and Learning, 14*(5), 12–26. https://doi.org/10.14434/josotlv14i5.12829

Graves, K. (2008). The language curriculum: A social contextual perspective. *Language Teaching, 41*(2), 147–181.

Gregersen, T., & Horwitz, E. K. (2002). Language learning and perfectionism: Anxious and non-anxious language learners' reactions to their own oral performance. *The Modern Language Journal, 86*(4), 562–570. https://doi.org/10.1111/1540-4781.00161

Gromik, N. A. (2012). Cell phone video recording feature as a language learning tool: A case study. *Computers & Education, 58*(1), 223–230. https://doi.org/10.1016/j.compedu.2011.06.013

Hsu, H.-C. (2016). Voice blogging and L2 speaking performance. *Computer Assisted Language Learning, 29*(5), 968–983. https://doi.org/10.1080/09588221.2015.1113185

Hsu, H.-Y., Wang, S.-K., & Comac, L. (2008). Using audioblogs to assist English-language learning: An investigation into student perception. *Computer Assisted Language Learning, 21*(2), 181–198. https://doi.org/10.1080/09588220801943775

Huang, H. (2015). From web-based readers to voice bloggers: EFL learners' perspectives. *Computer Assisted Language Learning, 28*(2), 145–170. https://doi.org/10.1080/09588221.2013.803983

Knoch, U., Rouhshad, A., Oon, S. P., & Storch, N. (2015). What happens to ESL students' writing after three years of study at an English medium university? *Journal of Second Language Writing, 28*, 39–52. https://doi.org/10.1016/j.jslw.2015.02.005

Krashen, S. D. (1982). *Principles and practice in second language acquisition.* Oxford: Pergamon.

Lecumberri, M. L. G., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication, 52*(11–12), 864–886. https://doi.org/10.1016/j.specom.2010.08.014

Leow, R. P. (2007). Input in the L2 classroom: An attentional perspective on receptive practice. In R. DeKeyser (Ed.), *Practice in a Second Language* (pp. 21–50). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511667275.004

Lepore, C. E. (2014). Influencing students' pronunciation and willingness to communicate through interpersonal audio discussions. *Dimension*, 73–96. Retrieved from https://eric.ed.gov/?id=EJ1080228

Li, L., Chen, J., & Sun, L. (2015). The effects of different lengths of pretask planning time on L2 learners' oral test performance. *TESOL Quarterly, 49*(1), 38–66. https://doi.org/10.1002/tesq.159

Mamaghani, H. J., & Birjandi, P. (2017). Oral pushed output: The route to long-term grammatical accuracy. *Journal of Teaching Language Skills, 36*(1), 57–84. https://doi.org/10.22099/jtls.2017.4044

McCormick, D. E., & Vercellotti, M. Lou. (2013). Examining the impact of self-correction notes on grammatical accuracy in speaking. *TESOL Quarterly, 47*(2), 410–420. https://doi.org/10.1002/tesq.92

Miles, M. B., Huberman, M., & Saldaña, J. (2019). *Qualitative data analysis: A methods sourcebook* (4th ed.). Thousand Oaks: Sage.

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45*(1), 73–97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Murphy, J. M. (2014). Intelligible, comprehensible, non-native models in ESL/EFL pronunciation teaching. *System, 42*(1), 258–269. https://doi.org/10.1016/j.system.2013.12.007

Pawlak, M., & Mystkowska-Wiertelak, A. (2015). Investigating the dynamic nature of L2 willingness to communicate. *System, 50*, 1–9. https://doi.org/10.1016/j.system.2015.02.001

Reid, A. (2004). *Towards a culture of inquiry in DECS*. Occasional Paper Series. Available at http://www.decs.sa.gov.au/learnerwellbeing/files/links/link_72576.pdf.

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse processes, 14*(4), 423–441.

Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty. *IRAL-International Review of Applied Linguistics in Language Teaching, 45*(3), 193–213.

Sheppard, B. E., Elliott, N. C., & Baese-Berk, M. M. (2017). Comprehensibility and intelligibility of international student speech: Comparing perceptions of university EAP instructors and content faculty. *Journal of English for Academic Purposes, 26*, 42–51. https://doi.org/10.1016/j.jeap.2017.01.006

Skehan, P. (1991). Individual differences in Second Language Learning. *Studies in Second Language Acquisition, 13*(2), 275-298. https://doi.org/10.1017/S0272263100009979.

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning, 49*(1), 93–120. https://doi.org/10.1111/1467-9922.00071

Sun, Y.-C. (2009). Voice blog: An exploratory study of language learning. *Language Learning and Technology, 13*(2), 88–103. Retrieved from https://doaj.org/article/1d8a6a5e81c6473ab409e87f4f26653d

Sun, Y.-C. (2012). Examining the effectiveness of extensive speaking-practice via voice blogs in a foreign language learning context. *CALICO Journal, 29*(3), 494–506. Retrieved from https://www.jstor.org/stable/calicojournal.29.3.494

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Rowley: Newbury House.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 125–144). Oxford: Oxford University Press.

Tashakkori, A., & Teddlie, C. B. (2010). *SAGE handbook of mixed methods in social and behavioral research* (2nd ed.). Thousand Oaks: Sage.

Tatar, S. (2005). Why keep silent? The classroom participation experiences of non-native-English-speaking students. *Language and Intercultural Communication, 5*(3–4), 284–293. https://doi.org/10.1080/14708470508668902

Text Inspector. (2018). Text Inspector: Online lexis analysis tool. Retrieved from https://textinspector.com/

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics, 24*(1), 1–27. https://doi.org/10.1093/applin/24.1.1

Zarrinabadi, N. (2014). Communicating in a second language: Investigating the effect of teacher on learners' willingness to communicate. *System, 42*, 288–295. https://doi.org/10.1016/j.system.2013.12.014

Appendix A

Sample Modified Text

In this lesson, we'll talk about two different types of lines.

Take a look at your surroundings. Are you close to a window with blinds? If you look out that window, can you see the next street or a highway? If you answered yes to any of these questions, then you are surrounded by lines, which are everywhere!

In this lesson, we are going to take a closer look at parallel lines, perpendicular lines. Each of these types of lines are classified as *coplanar* lines, meaning that they are located on the same plane, which is a flat, two-dimensional surface.

Parallel lines are defined as coplanar lines that do not intersect. They have the same slope and, just as the definition states, will never, ever meet at any point. Think about it: since slope is referred to as rise over run, having the same slope means that two lines will rise and run at the exact same rate, ensuring that they will never intersect each other. Let's take a look at real-life examples of parallel lines.

First, we have a window with blinds. Here, you can see that each blind is moving in the same direction and never touches another blind. Next, we have a parking lot. Notice that all of the lines are going in the same direction.

Perpendicular lines are coplanar lines that intersect and form a 90-degree angle. So, any time you have perpendicular lines, you will also have right angles and vice versa.

The slopes of perpendicular lines are *opposite reciprocals* of each other. Being opposite means that one slope will be positive and the other will be negative. Being reciprocals means that one slope will be the upside down or flipped version of the other.

Perpendicular lines are also visible in the real world. Take a look at a desk. Can you see how the top of it lays flat on all the legs? This means that the top of the desk is perpendicular to the legs and forms 90-degree angles, which keeps things from sliding off of it.

Appendix B

Student View of Instructor Veedback

Grammar

- *the professor talk about* (*the professor talked about*)

- *each of this line* (*each of these lines*)

- *Second line is ...* (*The second line is...*)

Pronunciation:

Please listen to the pronunciation of these two words at cambridge.dictionary.org:

- coplanar (especially note syllable stress)

- parallel (especially pronunciation of -r)

- intersect (I think you said intercept)

Be sure to sound the -s clearly on plurals (I heard you say *two line* - should be *two lines*)

Sample Instructor Veedback (Transcription for image on previous page)
Hi Anna (pseudonym). A couple of points of grammar and pronunciation for you to look at. First of all, remember that when you are speaking about a lecture that happened in the past you want past tense so, you said the professor talk about—the professor talked, ed, talked about. Also, with pronoun reference I think a problem with plurals maybe. [you said] Each of this line—Each of these lines were plural. Then I'd like you to go to dictionary.cambridge.org because you can listen to the pronunciation of coplanar especially the syllable stress of parallel, which is a hard word to say, the R followed by L and intersect, I think I hear you say intercept, which is a different word. So, intersect, listen to that one at dictionary.cambridge.org as well be sure that you are sounding the S in plurals clearly so two line should be two lines. Keep trying, your fluency was much better this week and you're looking into the camera continue to speak loud and clearly.

Appendix C

Lecture Summary Task Sample Script
Use the following script to help you summarize the lecture.
> The professor discusses (topic), which s/he defines as . . .
> S/he describes two types of . . .
> First, (definition) . . .
> The professor provides an example of . . . to illustrate this
> Second, (definition) . . .
> The professor provides an example of . . . to show this
> That is how the professor describes (topic) . . .

Remember that your summary should have a similar outline to that of the lecture. Use the following outline as a guide:
1. Topic
   a. Definition
2. Subtopic
   a. Definition
   b. Example
3. Subtopic
   a. Definition
   b. Example

Jeopardy Questions Task Script
Use the following examples to help you create Jeopardy Questions based on the lecture.
To reference the lecture:
> You mentioned that . . .
> When you were talking about _____, you referred to _____.
> In the first/second/last part of the lecture, you talked about/referred to _____.
> Could you repeat what you said about _____?
> You mentioned previously that . . .

To ask for specific information from the lecture (Specify):
> How much/How many/How does . . .?
> Why does/doesn't . . .?
> When is/are/do/does . . .?
> Where is/are/do/does . . .?
> What's the difference/connection between _____ and _____?

To ask for something in the lecture to be made more clear (Clarify):
> What do you mean by . . .?/What does _____ mean?
> What does _____ stand for?
> Did I hear you say that . . .?
> Did I understand you when you said . . .?
> Did I hear you correctly when you said . . .?
> What criteria did you use to . . .?

To ask for an example of something in the lecture (Exemplify):
> Can you give/Would you mind giving an example of . . .?
> Can you say more/explain a bit more about . . .?
> I'm not sure about _____. Would you mind explaining _____ in a different/in another way?
> Can you elaborate/Would you mind elaborating on . . .?
> I'm not sure I understand/I'm clear on/about _____. Can you give an illustration?

To take something from the lecture and make application to a different context or field of study (Apply):
> What do you think would happen if . . .?
> What other . . .?
> Are there other . . .?
> Do we know of any other . . .?

Sample Debate Task

You are the human resources manager and have the responsibility of filling a position for your company. Your friend Paul, whom you have known for 20 years, has applied and is qualified, but someone else seems even more qualified. Who should you hire?
> Solution 1: Hire Paul
> Solution 2: Hire the more qualified person

Debate Task Sample Script
> If I were you, I would choose [Solution 1 or Solution 2]
> Logically, [argument against the solution you did not choose]
> Plus, [argument against the solution you did not choose]
> [Argument for the solution you chose]
> Not only that, but [argument for the solution you chose]
> That is why I think you should [Solution you chose]

Appendix D

Interview Questions—First Set

1. Let's talk about your listening and speaking ability *in English*, in academic situations.
   a. Do you feel that you are a good note-taker? Why/why not?
   b. Do you feel that you can summarize key information for a lecture or presentation?

Why/why not?
   c. Do you usually ask questions in class? Do you feel that you can ask questions correctly so that your instructor understands? Why/why not?
   d. Do you feel that you can listen actively in class? Why do you say that?
   e. Do you feel that you can offer your opinion clearly so that others understand you? Why or why not?
2. Let's talk about your listening and speaking development in English.
   a. Do you feel that you can use the new vocabulary you learn when you talk? How do you know?
   b. Do you feel you're able to explain a lecture you attended to your friend by looking at your notes? Why/why not?
   c. In class, were there any instances where you felt you'd like to say something, but you didn't? If yes, when and why?
3. Let's now talk just about your English-speaking skills.
   a. How hard is it for you to express yourself fluently, with little hesitation and pauses? How do you know this?
   b. How hard is it for you to talk in a clear and easily understandable manner? How do you know this?
4. Let's talk about your strengths and weaknesses in English listening and speaking in an academic context.
   a. What kind of challenges or difficulties in listening and speaking do you think you'll face during the term?
   b. In relation to your speaking skills specifically, what is, in your opinion, your weak point?
   c. What about your strengths?

Interview Questions—Second Set
1. Let's talk about the listening speaking tasks in your class this term.
   a. How do you feel about these tasks?
   b. Which tasks are easiest for you to do? Which are more difficult?
   c. How hard was it to express yourself fluently, with little hesitation and pauses in the tasks that you did? Why do you think that was?
   d. Do you think these tasks have helped you in your listening and speaking? Why not?/ In what ways?
2. Let's talk about the grades and feedback you received on these tasks.
   a. Do you think the feedback you received has helped you? Why?/Why not?/In what ways?
   b. Do you think the grades and feedback are accurate? Why?/Why not?/In what ways?
   c. Do you listen to the feedback from your instructor during class or after? Why?
   d. How have you tried to incorporate your instructor's feedback in your video/audio responses?
   e. Is there any part of the feedback that was not useful for you? Why?/Why not?
3. (At the end of the term) How would you evaluate your listening and speaking ability?
   a. Do you have the same challenges after taking the course? What are they?
   b. What has helped you improve your listening and speaking?