Skilled orators and their public speaking skills have been held in high esteem since antiquity. Artificial intelligence-powered presentation platforms (AI-PPPs) are an emerging educational technology that provides a virtual venue and AI audience for users to practice a presentation or pitch and receive feedback to improve their public speaking skills. AI-PPPs allow users to customize their context (e.g., audience type, setting, type of presentation) and make their presentation to a virtual audience comprised of avatars. AI-PPPs use algorithms to interpret the user’s public speaking skills and provide an analysis of their performance, often including metrics on pitch, eye contact, pace, volume, distractors, pauses, etc. This research analyzes one such AI-PPP to determine its effectiveness utilizing the text complexity framework. This mixed methods case study reports graduate educational
innovation program participants’ AI-PPP-assigned public speaking scores, experiences, and concerns with using AI-PPPs as a learning tool, and data was collected from survey items, performance scores, and focus group interviews. The researchers identified that while the participants found initial value with the AI-PPP, there was growing frustration with its feedback and functionality. Implications based on these findings include recommendations for instructors for improving public speaking instruction with AI-PPPs, as well as developers for refining AI-PPP tools. In addition, limitations that may have impacted this study’s results are discussed.

Keywords: Artificial Intelligence; Instructional Technology; Emerging Technology; Avatars; Artificial Intelligence-Powered Presentation Platforms; AI-PPPs; Public Speaking

INTRODUCTION

From Socrates to present-day, public speaking skills have been and are valued throughout history and cultures. Across industries, being able to effectively generate and then communicate a message to an intended audience in a way that resonates is an essential skill set, and educators at all levels are working to develop their students’ public speaking abilities (Haunts, 2022; Parvis, 2001). Public speaking involves verbal communication – pitch, pace, cadence, volume – as well as body language – eye contact, interpretive gestures, – to make a presentation to an audience (Mehrabian, 2017). Researchers have identified preparation, practice, and feedback as being important elements for making an impactful public speaking presentation (Boyce et al., 2007; Menzel & Carrell, 1994). Educators have long focused on improving students’ public speaking skills (Winans, 1915). Now, developers are combining virtual venues and audiences with artificial intelligence (AI) to provide feedback on students’ public speaking skills.

AI is an emerging technology impacting education at all levels, and its impact is expected to grow significantly in the coming years (HoloniQ, 2021; Pelletier et al., 2021; World Economic Forum, 2022). For example, AI is already being used in self-paced early literacy programs along with math programs at all levels. At its core, AI is a series of algorithms and statistical logic models that analyze data by following specific, preprogrammed rules for making evaluations before coming to conclusions, and
AI commonly reports those conclusions as an output of some kind, such as a decision, feedback, or strategy (Eger et al., 2017; Mahesh, 2020). The goal of AI is to replicate human thinking and reasoning through voice recognition, language processing, analytical and predictive statistics, neural networking, deep learning, and cognitive computing (Zanetti et al., 2019).

The Current Study

Artificial intelligence-powered presentation platforms (AI-PPPs) are a type of emerging technology that is disrupting traditional approaches to teaching public speaking. AI-PPPs are innovative because of their use of AI for increasing the quality of presentations, and large companies including Microsoft are competing with smaller startups such as Virtual Orator, Pitch-Vantage, and Orai to develop them.

In an AI-PPP, users can practice public speaking in a low-stakes simulated environment by making presentations to audiences of AI avatars. When operating an AI-PPP, users will appear in the location they selected – such as a meeting room or lecture hall – along with the number of audience members, which take the form of avatars. They will also see their presentation deck, and a timer based on the presentation they selected. As users present, the platform will use AI to analyze the patterns in their public speaking that can include metrics about their rate of speech, eye contact, use of filler language, and other indicators. Afterwards, the AI-PPP analyzes different attributes of users’ public speaking and provides feedback.

Generally, this study adds to the emerging literature on the effectiveness and concerns related to using AI in education. Because AI-powered platforms are an emerging educational technology that has the potential to be applied across fields including education, corporate settings, and entrepreneurship, among others, it is important to examine AI’s interpretation processes for when clear-cut answers do not exist, such as in public speaking. Specifically, AI-PPPs have not been deeply analyzed, nor have they been analyzed using the text complexity framework. This study investigates the impact one AI-PPP platform had on graduate students’ public speaking skills. This study posed two research questions:

1. How do the quantitative public speaking scores students receive based on their presentation in an AI-PPP change over time?
2. How useful is AI-PPPs’ feedback for improving students’ public speaking skills?
To investigate those questions, a case study bound to 20 graduate students enrolled in an educational innovation program in a large research-based university in the southeastern United States was conducted.

**REVIEW OF LITERATURE**

**Public Speaking Practice**

From healthcare and education to sales and marketing, professionals must have the ability to effectively communicate with their stakeholders (Chevalier et al., 2017; DeKoven et al., 2008; Lundy & Stephens, 2015). In the context of public speaking, there is an interrelationship between preparation and practice, in that practice can be a form of preparation (Menzel et al., 1994). *Preparation* includes the collection, analysis, and structuring of information to be presented along with the creation of materials to be used during the presentation (Lucas, 2009). *Practice* is then the act of rehearsing the actual presentation, such as presenting to a peer-group, family members, or in another low-stakes manner, so long as formative feedback is provided (Berkelaar, 2019). For example, when preparing for a presentation, students may write note cards with key points and practice by giving that presentation in front of a mirror or video camera.

Across fields and contexts, practice that is sustained, rigorous, and focused has been identified as a key element for success, which extends to public speaking (Davis et al., 2020; Raja, 2017). The ways students can practice varies. For example, researchers have found that allocating class time to public speaking practice is a form of practical support that benefits students (Grieve et al., 2021; Stewart & Tassie, 2011).

Along with practicing presentations, researchers have identified high-quality feedback as being integral to developing students’ public speaking. Regarding feedback, Nicolini and Cole (2019) explain peer feedback as “an opportunity for students to engage in a reciprocal process of both producing and receiving feedback from peers” (p. 81), and they found that feedback from the speakers’ peers was most useful when it focused on the source credibility and quality of speech along with providing feedback that was both supportive and critical. Feedback in other areas, such as the delivery of the speech, was less useful. Saidalvi and Samad (2019) found that feedback was motivational and helpful to students when it pointed out positives and strengths of the speech, like Nicolini and Cole’s (2019) findings.
AI in Education

In education, AI is being used to personalize problems to students’ out-of-school interests (Walkington & Bernacki, 2019), individualize learning analytics’ feedback (Lim et al., 2021), fill in as teaching assistants (Kim et al., 2020), and offer conversational practice to foreign language learners (Fryer et al., 2017). AI is often used in dynamic quizzing programs that present students with a series of problems that become either more or less rigorous depending on students’ achievement. As Reich (2020) explains, AI works particularly well in these types of programs where there are objectively correct or incorrect answers because the programs are designed so students first take a diagnostic test that AI uses to gauge student performance and ability levels. After analyzing the results, AI provides a series of scaled problems based on them. AI then increases or decreases the problems’ rigor and types of supports it offers based on students’ responses. Reich explains that using AI for these purposes is reasonable because psychometricians can identify the rigor of a problem and plot it in a sequence, and AI draws from that sequence when assigning students new problems based on their responses to previous ones. Or, in other words, early literacy and mathematics programs tend to have a “correct” answer that can be proven, such as identifying the numerical value of a variable or correct definition for a word. However, using AI to support student learning when there is not a clearly identifiable correct answer becomes more challenging.

When a situation arises that has no clear answer, AI struggles because it lacks the objective information needed to quantify a response as correct or incorrect. In these situations, AI relies on formulas and theories to quasi-quantify the information using the best available data, and Lin and Zhu’s (2021) work with the “Fuzzy-AI Model” provides a relevant example. In a leadup to their discussion, they explain that there are many contexts in which subjective information is analyzed by humans to make decisions in qualitative contexts, with organizational leadership, equity-driven initiatives, and management practices being examples. These are like open-ended questions that do not have a correct answer, as they are each in their own contexts. For AI to be helpful, it must convert the subjective information into an objective form, so the AI can analyze it and make recommendations. As such, the human experience becomes “hardened” in a way that allows for a phenomenon under discussion to be analyzed using the Fuzzy-AI model (Lin & Zhao, 2023). In the model, the data is categorized to make an “if-then” statement, which is “a conditional sentence where the sentence following IF is called antecedent, and the sentence after THEN is called con-
sequent” (Maathuis et al., 2021, p. 356-357). This logic allows for the “if” to be a scenario and the “then” to be an outcome. As an example, the “if” might be a comment made by a manager to discipline an employee and the “then” would be the employee’s reaction, and AI would categorize this scenario, so it could draw on it in the future when providing feedback and responses that are relative to that scenario. While imprecision and uncertainty are part of this model – as not all employees are going to react in the same way in response to a manager’s disciplinary comments – those elements are represented in the “fuzziness” reflected in the model’s name, and there are implications for education.

When asking students to respond to an open-ended question or a multiple-choice question that requires interpretation, there are multiple ways students may respond based on their background knowledge, prior experiences, beliefs, and other qualitative factors (Rosenblatt, 1982). The result is that AI is not able to effectively compute those variables to determine if the student’s response was correct based on a binary “yes/no” system. This is where the fuzzy AI model works to categorize students’ responses using the Fuzzy AI model, so it can be “trained” to interpret the qualitative information in a way that mimics human’s thinking and reasoning. At this point in AI’s development, this represents an area that researchers are working to continually improve by adding their own ideas (e.g., Hong Yun et al., 2022; Krizhevsky et al., 2017) to improve AI’s interpretation skills for when clear-cut answers do not exist.

**Using AI with Simulations to Support Students’ Public Speaking Development**

Simulations have been used for training purposes in the medical and military fields for decades in situations where human interaction or feedback is necessary (Delingette et al., 1999; Page & Smith, 1998). Recently, simulations have started being used in combination with AI to create digital environments that replicate experiences in the real world, and researchers have leveraged these technologies to develop students’ public speaking skills. However, AI-PPPs are an emerging technology, thus, only a limited number of studies have been conducted about their effectiveness. The vast majority of these case studies tested one VR application or simulation as being an effective response for developing students’ public speaking skills in a controlled environment, such as in the confines of a laboratory or classroom setting.
Of particular interest was Palmas et al.’s (2021) study that analyzed the type of feedback offered by a virtual reality speech training program and if that feedback resulted in the participants having decreased levels of anxiety when making presentations. Palmas et al. (2021) had their participants, comprised of 200 undergraduate students, make presentations to avatars in a virtual environment. As they presented, AI and gamification principles were used to provide students with direct feedback, such as icons appearing above avatars’ heads in their digital audience to indicate if their presentation was impactful, and this feedback was based on the participants rate of speech, eye contact, and other attributes. Overall, they found that the feedback was beneficial and making multiple virtual presentations of this kind decreased their participants’ public speaking anxiety.

In another study, Boetje and Ginkel’s (2021) used a pre-test that required their 35 graduate student participants to complete a three-item questionnaire about their views of public speaking as well as make an oral presentation that was scored using a validated rubric. They then divided the participants into two groups, with one group making a practice presentation to avatars twice in a virtual environment, once with feedback being provided and once without receiving it. The other group made three presentations to the avatars in the virtual environment, and they received AI feedback for one of their presentations but not for the other two. Overall, they found that the extra session resulted in increased performance and lessened public speaking anxiety. They also found that using VR in this capacity resulted in decreased levels of anxiety for speaking in public, based on the questionnaire data, and that phenomenon was found to be true for both the AI and non-AI participant groups without any statistical significance between them.

El-Yamri et al. (2019) sought to improve AI’s interpretation and feedback in a VR tool they had developed for practicing public speaking skills with a simulated reactive audience. The AI interpreted inputs of the speaker’s voice, content, body movements, and gaze, with the main goal being to develop AI that interpreted the speaker’s emotions (e.g., tone, message/body language coherence). Five second iterative recordings of the speaker’s voice were analyzed for emotional indicators by voice recognition software and were assigned confidence scores between 0 and 100. In their experiment, El-Yamri et al. (2019) aimed to adjust the subjective weight values given to each emotion by the researchers to better align the AI with human interpretations. Three actors performed speeches for the 16 human audience members who rated the speeches’ emotional values. The AI’s ratings and humans’ ratings were then compared. Results indicated that the voice analysis software only interpreted the emotions successfully around 55% of the time.
Further, results indicated that the human audience was significantly engaged with the emotions of stress, calmness, and happiness, which the AI’s algorithms did not heavily weigh.

While the studies by Palmas et al. (2021), Boetje and Ginkel (2021), and El-Yamri et al. (2019) are representative of current efforts to research the impacts of AI in simulations for public speaking, the contemporary literature remains inconclusive and focused on metrics of public speaking anxiety. More investigation is needed into the effectiveness of contemporary AI-PPP interpretation processes and the resulting feedback they offer as technology evolves. For AI-PPPs to reach their potential, it is critical for researchers to assess the efficacy of these technologies.

THEORETICAL FRAMEWORK

To guide this study, the researchers adopted the text complexity framework used by the Common Core State Standards (National Governors Association Center for Best Practices, 2010). As shown in Figure 1, the framework for text complexity consists of a Quantitative, Qualitative, and “Reader and Task” components, and it was developed to determine the rigor of a text.

Figure 1

*The Common Core State Standards Framework for Text Complexity*
The quantitative component determines the complexity of a text based on the frequency of polysyllabic words, the length and variety of sentences, and the use of transition words and phrases to build text cohesion. There are multiple reading formulas that can be used to assess the quantitative rigor of texts (e.g., Dale-Chall, Lexile Framework), and websites have been developed, so individuals can upload a text into one of them to determine the text’s quantitative complexity, such as the Dale-Chall Readability Calculator (https://charactercalculator.com/dale-chall-readability/) and the Lexile Text Analyzer (https://hub.lexile.com/analyzer).

Whereas the quantitative component is evaluated by a website, the qualitative component requires a human to make “an informed decision about the difficulty of a text in terms of one or more factors,” notes the National Governors Association Center for Best Practices (2010, p. 5). Because there are differences in the qualitative component based on if the text is narrative or information, main considerations used to determine a text’s qualitative complexity includes its level of meaning (e.g., if the text is satirical or literal), structure (e.g., if the text’s meaning is explicit or implicit), language conventionality and clarity (e.g., if the text uses conversational language or academic/professional language), and knowledge demands (e.g., if the text only requires “common” knowledge or domain-specific knowledge). Because these considerations go beyond formulas and require inferential interpretation, the text complexity positions humans as being capable of making these determinations, as mathematical formula are not designed to accommodate the nuanced language usage needed for this work.

Finally, the “Reader and Task” component focuses on individuals’ motivations for engaging a text and the actions they are intended to complete during and after that engagement. Considerations for this component relate to individuals’ intrinsic motivations to engage with a text, including if individuals are interested in its topic and can foresee how it will advance their knowledge base or ability level. Extrinsic motivation includes individuals reading the text to complete a class or work assignment, without having an internal desire for doing so. The “Task” element then extends to what individuals are expected to do after reading the text, and examples include answering questions about the text, debating the text’s meaning with peers, and using the information from the text to complete a project.
METHODS

Setting and Participants

Case studies are intensive, holistic, in-depth investigations of an individual or group (Heale & Twycross, 2018). Meyer (2007) explains that case studies must be contextually “bound” to a specific group of participants and the context in which those participants engaged a phenomenon must be clearly explained. By providing that description along with findings, it allows for other researchers to replicate the study and validate its findings, which makes it possible for generalizations to be developed (Kumar & Ormiston, 2012).

This study is bound to participants (N = 20) who were graduate students enrolled in an educational innovation program at a large research-based university in the southeastern United States. Housed in the university’s School of Education, the program prepares its students to work in the field of educational innovation. All participants were enrolled in the required “Seminar in Educational Innovation” course, and its main focal points include lessons about emerging technologies, instructional design, and educational entrepreneurism. The course met once a week for three hours.

To create an experience with an AI-PPP as part of the course, a series of asynchronous modules were designed to develop the students’ understanding of educational entrepreneurism. For authenticity and to generate interest, the students first independently brainstormed an idea for a new venture in the field of educational innovation prior to starting the modules. Example ideas included consultancies focused on hybrid teaching, a digital tool for high school students to study historical battles, and a method for using popular American culture to teach foreign language. Next, the participants completed one asynchronous module per week for six weeks, and each module addressed a foundational element related to the business planning needed to launch their venture idea, and the elements included (1) value propositions, (2) customer segments, (3) channels, (4) customer relationships, (5) cost structures, and (6) revenue streams.

Of the participants, 12 identified as White, four as Black, and four as Asian. Five identified as males, 14 as females, and one as non-binary. One participant’s highest degree earned was a doctorate, one a master’s degree, and 18 held a bachelor’s degree. Four of the 20 participants were not native English speakers who all tested as “proficient” in English as required for admissions into the program. Table 1 provides further participant demographic data.
### Table 1

**Participant Demographics**

<table>
<thead>
<tr>
<th>Age</th>
<th># of Participants</th>
<th>Years of Professional Work Experience</th>
<th># of Participants</th>
<th>Professional Field Prior to the Program</th>
<th># of Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>21-25</td>
<td>6</td>
<td>0</td>
<td>3</td>
<td>Education</td>
<td>10</td>
</tr>
<tr>
<td>26-30</td>
<td>6</td>
<td>1-3</td>
<td>6</td>
<td>Public Policy</td>
<td>2</td>
</tr>
<tr>
<td>31-35</td>
<td>5</td>
<td>4-7</td>
<td>6</td>
<td>Neurology</td>
<td>1</td>
</tr>
<tr>
<td>36-40</td>
<td>0</td>
<td>8-10</td>
<td>1</td>
<td>Computer Sci. and Tech.</td>
<td>3</td>
</tr>
<tr>
<td>41-45</td>
<td>0</td>
<td>11-15</td>
<td>0</td>
<td>Economics</td>
<td>2</td>
</tr>
<tr>
<td>46-50</td>
<td>1</td>
<td>16+</td>
<td>4</td>
<td>Humanities</td>
<td>2</td>
</tr>
<tr>
<td>51-55</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>56-60</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>20</td>
<td>Total 20</td>
<td>Total 20</td>
<td>Total 20</td>
<td></td>
</tr>
</tbody>
</table>

Based on the demographics, the participants in this study were mostly recent college graduates who are native English speakers, they all had at least a bachelor’s degree, and the majority had at least some professional, post-college work experience.

### The AI-PPP

PitchVantage was the AI-PPP used by participants in this research. PitchVantage was selected for this study because the participants were studying emerging technologies developed by startup companies, and it provided students with an authentic example of that type of technology. In this AI-PPP, the user first adjusts a series of settings to select a certain room size, audience type, focus of presentation (e.g., informative or persuasive), and the length of presentation. Once the settings are selected and confirmed, the user has the option to upload presentation slides. Next, PitchVantage generates the digital environment, including the avatars, and the user begins their presentation. While they present, the platform records them using the built-in camera on users’ device. In addition, the user can toggle between viewing the avatars and their presentation slides, and a timer appears in the corner of their screen. During the presentation, the avatars make movements.
to show their interest (e.g., nodding their heads) and disinterest (e.g., looking around the room), and these movements are responsive to users’ presentation.

As described in PitchVantage’s patent (Gupta et al., 2022), microphone, camera, and presentation file inputs are run through six different analysis engines whose results feed into an overall presentation analysis engine. The microphone inputs are analyzed by (1) a speech-to-text engine, (2) a speech text analysis engine, and (3) a vocalics analysis engine. The speech-to-text engine translates the audio input from the microphone into textual representations of the words being vocalized by the user. The text is then analyzed using natural language processing to assess the linguistic complexity of the speech, word choice, and number of verbal distractors (e.g., “um” or “uh”) with the speech text analysis engine. The vocalics analysis engine analyzes the non-language aspects of the user’s vocalizations (pace, pitch, volume, pauses, rhythm, intonation, and intensity of voice). Camera inputs are run through a behavior analysis engine and a biometrics analysis (e.g., facial recognition) engine. The behavior analysis engine takes the camera’s video and analyzes the user’s movement, position, gestures, facial changes, and eye contact. Eye tracking is calibrated through an introductory screen where the AI-PPP is trained to recognize specifically where a user is looking.

Notably, third-party software is employed by PitchVantage for eye tracking and partially for body tracking. A user’s uploaded presentation file is analyzed for the number of bullets per slide, the amount of time a user spends on each bullet, as well as the presentation file’s design, including use of text, images, and overall organization through a materials analysis engine. An overall presentation analysis takes the metrics from the various engines and runs them through a supervised machine learning algorithm. This algorithm is trained by (a) interpreting thousands of prerecorded presentations with its engines, and (b) contrasting them against ratings of six human public speaking experts who have also reviewed the prerecorded presentations. Machine learning produces a predictive model that merges the ratings of the presentations from the analysis engines and the six public speaking experts.

When finished, the platform analyzes a user’s public speaking and reports feedback to them based on eight variables that include (1) pitch variability, (2) pace variability, (3) volume variability, (4) verbal distractors (5) total pause percentage, (6) pace, (7) long pauses, and (8) eye contact. As a user continues making presentations, PitchVantage tracks their data, so the user can see their progress over time, and PitchVantage calculates the scores from the eight attributes into a total score out of 100. For the class, the
participants’ final assignment was to pitch their venture idea to their classmates, and PitchVantage was used to give students a way to practice and prepare for that final presentation.

**Data Collection and Analysis**

For this study, the researchers collected performance scores, surveys, and focus group interview data, so they could employ a mixed methods approach. Specifically, this study utilized a concurrent triangulation mixed methods design (Almeidai, 2018) because the researchers collected three sets of data simultaneously, analyzed each data set independent of the others, and validated their findings by cross-referencing the data sets against themselves. While the quantitative data in this study were employed to point toward the effect of the AI-PPP on students’ public speaking scores, qualitative data were harnessed to report the experiences and opinions of the participants on the AI-PPP’s interpretations of their public speaking and the quality of the resulting feedback. The researchers selected mixed methods due to the holistic nature of case study research and because the merging of analyses from two different forms of data leads to greater credibility in a study’s findings by showing the full picture of the phenomena under examination (Creswell, 2014; Mertler, 2017).

The quantitative data consisted of the participants’ performance scores reported by PitchVantage. The range of these scores was 0-100, and PitchVantage used its proprietary formula (not available to the researchers) to calculate them based on the eight attributes it tracked. For this study, all 20 participants completed five presentations, and the researchers documented their individual performance scores for each one. They next summarized the data by plotting the performance scores as a function of time for each participant. Due to the data’s normality, paired-samples $t$-test was used to determine if there were differences between the first and last sessions, and a linear regression model was used to test whether there were mean differences among any of the five points of time, not just the first and last sessions. These analyses allowed the researchers to look across the data for growth over time and determine if there was a particular session in which no further growth was found, as recommended by Boetje and Ginkel (2021).

The qualitative data included two open-ended items from a survey and three focus group interviews. The survey items asked the participants what they would change about the modules and the second one asked if they would recommend the use of the AI-PPP as a learning tool. The focus
groups each lasted for approximately 60 minutes, and the participants were asked what they liked about using the AI-PPP and what challenges they had with it. They were also asked about their emotions when using the AI-PPP, specifically if there were times when they felt happy, sad, anxious, frustrated, excited, or another way.

To analyze these data, the researchers first transcribed the focus groups’ transcripts. Two researchers independently analyzed both the responses to the survey items and the transcripts for themes and patterns. For their analysis, the researchers identified in vivo and sociologically constructed codes in the data. The in vivo codes consisted of the participants’ exact words and phrases that the researchers saw as communicating the participants’ perception of the platform and its impact on their abilities as a presenter (Strauss, 1987). The sociologically constructed codes were the meaning the researchers attached to the participants’ words and phrases in relation to their perceptions of the platform and its impact on them (Strauss, 1987). The researchers categorized their codes into themes, which they operationalized.

To add validity, Anfara et al. (2002) recommend code mapping as a way “to bring meaning, structure, and order to data” (p. 31). Further, according to Anfara et al. (2002), there are three levels of iterations in a code map. The first iteration is initial coding, which consisted of the in vivo and sociologically constructed codes. The second iteration is patterns in the data that the researchers developed into themes, which they did by categorizing their initial coding. The third iteration is application that is the creation of a hypothesis or theory in response to the research question(s), and the researchers worked to achieve application by explaining how their first two iterations of coding respond to the research question. Because Anfara et al. (2002) stress the importance of making coding visible, the researchers included their code map in the Appendix.

The act of qualitative research positions researchers as tools of research (Poggenpoel & Myburgh, 2003), which means that researchers carry their own background knowledge, biases, and beliefs with them as they make meaning from the data. To solidify the credibility of the qualitative findings of this study, peer debriefing (Lincoln & Guba, 1985; Shenton, 2004) was performed by the researchers to align codes and generate themes. While each researcher brings their own background with them to this study, the researchers held meetings where they shared their coding and their underlying perspectives on which those codes were built. This process, outlined below, helped to limit personal biases inherent to qualitative methodology and bridge incongruent views between the researchers to build consensus in the meaning-making process.
Merging Quantitative and Qualitative Data

To arrive at their findings, the researchers met and reviewed the data. For the review, the researcher who analyzed the quantitative data shared the results. Next, the two researchers who completed the code maps presented their work. At this point, the researchers discussed the connections they identified between the quantitative data and the first two iterations from the code maps. They asked questions about the basis for the qualitative themes in relation to the quantitative data. After this meeting, the researchers who coded the qualitative data analyzed it again to identify the specific codes that served as the foundation for their themes, so to make clear connections between the qualitative and quantitative data in response to the research questions. After completing that process, the researchers met a second time to review the data and discuss the connections. From this conversation, the researchers came to a consensus about the connections and how they evidence both the impact of PitchVantage on the participants’ scores and their perceptions. By solidifying those connections, the researchers agreed they had collected and analyzed the data needed for this study.

FINDINGS

In this section, the researchers will report their quantitative and then qualitative findings. The researchers chose this order because the quantitative data shows the scores the participants received from the AI-PPP. The qualitative data then is intended to focus on the participants’ experiences as users of the AI-PPP. In this way, the quantitative data evaluated the participants’ performance based on their use of the platform that is next explained by the qualitative data.

Quantitative Findings

For each session, the AI-PPP provides participants with a “performance” score that is a composite score based on the eight attributes it analyzes from the presentation (pitch variability, pace variability, volume variability, verbal distractors, total pauses, pace, long pause, and eye contact), which it calculates from 0-100 using the developers’ proprietary formula. For this study, the researchers collected the first five performance scores from each participant to determine if there was a statistically significant
change of the scores over time. As shown in Figure 2, there was not a significant or consistent pattern of change over the five sessions. From the first time using the AI-PPP to the last time, the analysis shows that there was no improvement, on average ($t(19) = -0.137, p = .89$), which is shown in the regression model based on the average scores across the five presentations ($F(4, 76) = .49, p = .74$).

**Figure 2**

*Participant Performance Scores Over the Five Sessions*

In addition, when analyzing the presentations – from the first presentation to the second, from the second to the third, etc. – no statistical significance between presentations was found.

**Qualitative Findings**

Based on the survey items and focus group interviews, the researchers identified four main qualitative themes. In them, the participants identified the benefits and frustration points for using the AI-PPP and the ways they manipulated the AI’s interpretation processes to earn higher scores. All
quotes in the sections below are attributed to gender neutral pronouns in order to support participants’ confidentiality.

**Benefits of Using PitchVantage.** Across the responses from the survey items and interview prompts, the participants identified multiple benefits for using AI-PPPs, with a main focus being improved communication skills. One participant commented that “I think it’s [the AI-PPP] nice to work on stating things clearly,” and another noted it being “a useful platform to practice articulating ideas.” Both comments are in response to the participants using the AI-PPP to complete the class assignments. Outside of articulating their ideas, one participant said that the AI-PPP “helps to a limited extent in getting students to think about how they are presenting.” This comment differs in that the earlier ones focused on the verbalization of words. Further, it was unique in that the participant used the word “presenting” in a way the researchers understood to mean a method for combining spoken words, visual aids, and the act of communicating to an audience. This insight is key because AI-PPPs are platforms for practicing presentations, which is separate from making a speech, and these benefits extend to its functionalities.

When using the AI-PPP, the participants were able to switch between viewing their avatar audience and their presentation slides. Multiple participants noted that functionality in their comments, with one stating that “I like that you can… toggle like between the audience and like what your purpose is and the time and you can have slides.” Using those functionalities while making presentations resulted in some participants reporting an increased level of comfort for when they had to present to a live audience. For example, “It [the AI-PPP] helped me feel more comfortable in front of a real audience,” mentioned one participant. About the experience of presenting to a virtual audience, a participant explained that when entering the digital environment for the presentation:

“I didn’t expect to have that much of a reaction to that one. I was like, ‘oh, I’m still speaking to the screen,’ but it was like that you’re talking to the room of people versus like Zoom, which is like ‘Oh, I am kind of now staring at squares again. That’s [the presenting in a digital room] a completely different kind of way of coming about it.’”

As described by this participant, it was the experience of being in a virtual presentation setting and not speaking using Zoom, a video-conferencing platform, that made a significant impact. The avatars acted more like an audience than the commonly used displays of video-conferencing platforms, with each person appearing in a square. In all, the main benefit of using the AI-PPP was that it provided a virtual space to make practice presentations.
Then, when they made their in-person presentations, they reported feeling more comfortable and prepared.

**Frustrated with Feedback.** A significant concern of the participants was the type of feedback the AI-PPP provided them with after completing their presentations, and they were largely underwhelmed with it. Regarding the types of quantitative feedback, the participants were confused about its meaning and ways to improve based on it.

About the feedback, a participant said “It did not provide clear reasons as to why it gave us the scores that it did. I was not sure how to reasonably improve my pitch.” Another participant added that “I don’t think that the way that PitchVantage graded the video was providing helpful feedback. As much as I initially tried to implement the changes, my score didn’t really change ... I think the standards they have around pitch, volume, etc. are unrealistic in execution.” A third participant stated that “The feedback at the end wasn’t as useful as I would have hoped because it is too general. I also conscientiously tried to fix my problem areas over the course of the sessions, but I continuously received the same feedback and low scores in those areas.” This frustration was captured by a participant, who wrote: “I want to get real feedback from real people.” Overall, the participants were invested in using the AI-PPP, but they grew frustrated due to its lack of clarity about the criteria used to assess them, and they felt that they could not use the feedback to improve their score. This resulted in them desiring feedback from real people, and it led to some participants experimenting with the platform to improve their score.

**“Gaming” the Platform.** The AI-PPP’s algorithm evaluates speech by measuring the verbally quantifiable patterns in the speaker’s language. While this approach works well to identify specific patterns, the actual content and meaning of the speaker’s presentation is not measured, and the participants identified this area as being manipulable.

When explaining their experience using the AI-PPP, one participant shared that they made multiple presentations to the platform and received average scores, but then explained that:

“At some point I was like, I would try to cheat the platform and see where the gaps are at this point… One of those times... I stared at the ceiling and just spoke gibberish, and it told me my eye contact was good, my tone and content were all outstanding and I was like, ‘OK, so speaking at the ceiling and speaking gibberish gets me a near-perfect score while talking about my context and staring at the screen gets me mediocre results.’"
For this participant, they grew frustrated when following the procedures and assumptions for presenting, meaning that they followed the AI-PPP’s recommendations for looking into the camera when presenting, adjusting their pitch and volume to engage the avatar audience, and ensuring their content was clear. However, when they experimented by not abiding by those norms, their score increased, and another participant shared a similar story:

“I always got rated really low on eye contact… so I tried staring straight at the webcam the whole time to get a higher eye contact score… I tried looking at the people [avatars] on the screen to get a higher eye contact score… I tried switching between the two and that didn’t work. So, I’m like, this dang eye contact score is making my whole average go down. And then one time I was like, I’m going to speak like a literal cartoon character with like crazy like pauses and voice inflection. Like, I sounded crazy, basically, and I got the highest score that I got.”

In both examples, the participants grew frustrated with their failed attempts to increase their scores. In response, they each did a “mock” presentation where they altered their speech – one using gibberish and another speaking like a cartoon character – while not making eye contact with the avatars. After making their mock presentations, both participants received some of their highest scores. The result, as one of these participants explained, was that they liked AI-PPPs “in theory” but not in practice, and that sentiment also appeared in another finding.

**Expectation for the AI-PPP.** Throughout the data, the researchers identified an implicit expectation for the AI-PPP’s benefits that went unfulfilled. To explain, the participants’ responses to the survey items that asked if they would recommend the AI-PPP as an effective teaching tool (the majority of them would not) were structured in a way that indicated they thought they would enjoy and recommend it, but their experiences changed their viewpoints and recommendations. For example, a participant wrote, “Initially I thought PitchVantage would be a great way to practice my public speaking, but I always felt super awkward completing the videos.” Another one shared that the AI-PPP “was a good tool to be shown, but it was frustrating to use.” Other participants referred to the AI-PPP as being a “great tool” and “interesting tool” before sharing a frustration they had with it, most commonly that frustration was due to the feedback it provided. The researchers interpreted this pattern to mean that the participant saw potential value in the conceptual utility of an AI-PPP; however, their experiences with this particular AI-PPP’s interpretation processes and feedback resulted in them not wanting to recommend it.
DISCUSSION AND IMPLICATIONS

Based on the findings from this study, the AI-PPP did not result in improved public speaking skills for students over time, which differs from previous studies that demonstrated growth in students’ public speaking skills (Boetje & Ginkel, 2021; Palmas et al., 2021), and students did not perceive the platform as being helpful to improving their skills. In its current form, the AI fails in the context of evaluating public speaking skills, a context in which there are no clear-cut answers. This study’s findings imply that further research into AI-PPPs is needed. Future research may expand upon the case study design for evaluating multiple AI-PPPs to analyze patterns in strengths and shortcomings. In addition, future researchers may utilize a true experimental design to provide context regarding the effectiveness of AI-PPPs.

The participants did express an expectation that AI-PPP platforms can be useful by providing specific recommendations for them to improve their public speaking skills. For example, the findings indicate that PitchVantage’s third-party eye tracking software has difficulty identifying the user’s eye contact with the audience, leading to skewed results. Therefore, AI-PPP developers must ensure their AI is calibrated to better mimic a human’s interpretations of public speaking in a systematic process like that of El-Yamri et al. (2019). Through a scientific and systematic process, the AI’s final scores can be compared to that of a panel of humans for the developers to then fine-tune their algorithms by adding greater weights to aspects of public speaking that align with a human’s interpretation.

While the findings of this study had negative results that run contrary to the extant literature, new insights can be gathered by comparing these findings to the text complexity framework. For this study, the three-pronged approach to evaluate the AI-PPP’s text complexity aligns tightly because the presentation platforms use AI to provide quantitative feedback to users, and that feedback includes a numerical score based on different aspects of the presentation. During the presentation, users are provided “real-time” qualitative feedback based on the avatar’s reaction to the speech, such as them nodding their heads to show interest or checking their watches and rolling their eyes to express disinterest. For the “Reader and Task” dimension, it focused on the students’ motivation to use the platform during their courses and how they used the feedback to improve their performance. To support AI-PPPs and instructors interested in using them, this study’s implications will draw from the text complexity framework to provide further analysis, insight, and recommendations for both instructors and developers.
Recommendations for Instructors

Instructors need to prepare students to use an AI-PPP by providing them with information about what to expect while engaging the platform. From this study, it was clear that the participants’ expectations for using the AI-PPP did not align to the experiences they had when engaging the platform. This gap is problematic because it resulted in the participants not wanting to use the platform. In response, instructors can first explain the current status of AI, meaning that they highlight what AI can do well and where it is still being developed. They can also explain the fuzzy AI model and its use of “if-then” statements to make evaluations about their performance. Instructors can then have students practice making presentations with the platform in class. That way, instructors can make certain with a larger sample size of students that the AI-PPP cannot be creatively “gamed,” and students will have time to experience the platform with their classmates. Afterwards, the instructor can debrief the students about the AI-PPP, which will help level-set expectations.

Next, instructors should not solely rely on the reports from the AI-PPP as the only piece of feedback students receive about their public speaking. While watching each student’s presentation can be time intensive, there are multiple ways instructors can supply students with feedback about their presentation. For example, students can do a self-reflection on their performance after making a presentation in the platform based on the feedback they received, students can view their classmates’ recorded presentations that they made to the AI-PPP and offer feedback, and instructors can view short clips of their students’ recorded presentations to offer feedback. In addition, students can use multiple AI-PPPs and compare their results. These methods can be used individually or in combination to provide students with multiple forms of feedback.

Finally, instructors need to fully explain the purpose for developing their students’ public speaking skills. Resigning public speaking to one class in students’ post-secondary coursework but requiring them to make multiple presentations is not logical, though full programs of studies make this a reality. Instructors who require students to make presentations, speeches, pitches, and other forms of public speaking in their classes also need to be teachers of public speaking. Before using any AI-PPP, it is recommended that instructors take time to explain the importance of public speaking and show high-quality examples of it from professionals in their field.
Recommendations for Developers

First, the quantitative feedback provided by AI-PPPs needs to be rigorously tested and improved, and developers need to be cognizant of their AI-PPPs’ limitations. The participants in this study found the feedback to be unreliable, and when they took steps to improve based on the AI-PPP’s feedback, they did not see their efforts reflected in higher scores. This area of frustration is noteworthy because readability formulas have been used successfully to assess the rigor of written texts. With AI-PPPs, the text takes a verbal form, which makes it more complicated to interpret into text to analyze. For example, with an audio recording, AI-PPPs must analyze the spoken words and body language, which have unique challenges as compared to analyzing written text. In addition, whereas punctuation marks can be used to help interpret pauses, moments of excitement, and length of sentences of written texts, AI-PPPs must identify specific instances of voice inflection, utterances, and other audio markers coupled with a person’s body language to analyze their speech and presentation. While technology is improving in these areas (Gallino, 2022; Zhang et al., 2021), AI-PPPs may not be providing accurate, usable feedback to users, or, if that feedback is used, they are not seeing improvements based on their scores. In response, AI-PPPs need to invest in their voice analysis software (e.g., speech-to-text, speech text analysis, and vocalics engines) and algorithms to ensure their feedback is consistent and reliable. Furthermore, when users adjust their presentations based on the feedback, they need to see improvements in their scores. Score ranges as currently constructed may not be helpful to students, and AI-PPPs need to be able to provide fine-grain scores that more clearly indicate change (e.g., updating measurements to be more sensitive). If instructors are going to use AI-PPPs with their students for practicing presentations, this information is crucial to demonstrate incremental improvement. If AI-PPPs are not able to deliver that feedback, they should not be adopted or marketed in that way.

Second, regarding the qualitative element, the AI-PPP did not provide any insight into these aspects of their presentations, according to the participants. For example, when the two participants shared about how they talked “gibberish” and as a “cartoon character,” they had higher scores. The researchers interpreted these higher scores as the platform’s speech-to-text, speech text analysis, and vocalics analysis engines becoming confused by the unusual vocalizations they were not trained to understand, resulting in misidentifying the meaning of the presentation. Therefore, like Reich (2020) and the text complexity framework noted, a human is still needed to discern the meaning of a presentation. While AI-PPPs may have the func-
tionality for their users to record themselves, they do not yet have the technological capability to determine if the presentation is on topic nor if the information being presented is meaningful. Rather, a human is still needed to make that distinction.

To improve, AI-PPPs can continue developing their speech-to-text, speech text, and vocalics features and algorithms to further address this area. Developers may choose to expand their human-labeled sound and image classifications or use deeper, more expansive machine learning with human refinement to correctly detect and classify eye contact and appropriate cadence and pitch changes. Siri, Alexa, and Google Assistant, along with similar voice recognition tools, offer significantly complex and robust voice recognition and interpretation technologies. As the price point for AI falls, developers may be able to purchase significantly advanced third-party AI voice interpretation kits from Apple, Amazon, Google, or similar providers.

Alternatively, AI-PPPs developers can design features to better utilize humans as part of their platform. For example, AI-PPPs can create a user-to-user interface where to get human feedback about their presentation, users would have to submit feedback to one or more users about their presentation. AI-PPPs could allow course instructors to load a rubric for the presentation into the feedback process and have users self-assess their presentation or use the rubric to assess their classmates. For AI-PPPs to be better used in higher education, it is encouraged that they utilize one of the human-centered recommendations, which will help course instructors use them with their students.

Third, for the “Reader and Task” element, the AI-PPP did not include any functionalities that could be utilized for this element, but there are options to improve it as related to intrinsic and extrinsic motivation. To increase their intrinsic motivation, AI-PPPs could have students complete a self-assessment about their public speaking and presentation skills, and that self-assessment can include areas that they perceive as being their strengths along with those for improvement. That way, after students make a presentation, they can view their assessment results from the original self-assessment and use the feedback from their presentation to see if they are improving in areas they identified as weaknesses. In addition, AI-PPPs can offer them strategies for improving by providing users with examples and non-examples of different areas. For extrinsic motivation, AI-PPPs can create digital badges, certificates, or other micro-credentials that students earn as they reach certain public speaking milestones measured by the platform. There is a significant amount of research on building student motivation in the literature, and it is recommended that developers integrate research-based strategies for increasing student motivation into AI-PPPs.
Finally, AI-PPPs are often standalone platforms, meaning that they do not yet integrate with learning management systems. Creating a plug-in, extension, or application that can fully embed an AI-PPP into a learning management system will likely help instructors better integrate the platform into their instruction and make it more accessible to students.

**LIMITATIONS**

This study is limited by a few factors. First, the sample of participants in this case study was small ($N = 20$). Second, only the PitchVantage AI-PPP was analyzed. Because PitchVantage was an essential component for the asynchronous modules, it needed to be used by the course instructor, and it was unrealistic to use multiple AI-PPPs. As a result, PitchVantage was the focus technology used in this study and case studies of other AI-PPPs may result in different findings and recommendations. Third, PitchVantage’s developers do not share their proprietary formulas for composite performance scores (0-100 based on eight dimensions), which would provide the researchers with the opportunity for more fine-grained analyses. Fourth, the context of the graduate program the participants were completing at the time of this study is also important. This program has an intentional focus on evaluating educational technologies, and the participants may have an advanced keenness in that area due to the program’s focus. For example, if this case study was conducted with students in a business or fine arts program, the results may have differed. In addition, the varied professional and educational experiences along with the students’ ages may have impacted this study. The final limitation addresses the intervention used by the researchers of this study.

The findings showed that there were no significant improvements in the participants’ scores, from their first score through their final one, and there are potential reasons that may have impacted that finding. While the PitchVantage itself may not have worked well, it is also possible that the researcher who prepared the participants to use the tool did not give them enough time to practice with it. Other possibilities include that the participants needed more information about AI, time to explore the tool’s functionalities, and the expectations for their performance using the tool, among other reasons. With AI and AI-PPPs becoming more readily available, this presents future researchers with an opportunity to further study this area to identify best practices for using AI-PPPs.
CONCLUSION

PitchVantage is representative of AI-PPPs, an emerging technology with the potential to significantly improve students’ public speaking and presentations skills. However, that potential is yet to be realized. This study’s participants voiced concerns regarding the feedback provided to them by the AI-PPP along with the ability to “game” it. They also expressed expectations for AI-PPPs, which were not met, and this study’s researchers used the text complexity framework to guide their recommendations for meeting instructors and developers to meet those expectations. Currently, PitchVantage specifically and AI-PPPs generally are still in their technological infancy. Yet, with large investments being made to develop AI for educational purposes (HolonIQ, 2021), these platforms stand to greatly benefit by advancements in their algorithms, improved feedback to users, and better integrate into learning management systems. In turn, it provides researchers with the opportunity to study AI-PPPs as both a learning tool and potentially useful educational application of AI. By doing that research, it can inform the way other educational technologies are designed and developed along with how instructors can use them with their students.

REFERENCES


APPENDIX: THREE ITERATIONS OF ANALYSIS

*Note: should be read from bottom up*

**Qualitative Research Question:**

How useful is AI-PPPs’ feedback for improving students’ public speaking skills?

**Third Iteration: Application of Data**

There are distinct benefits and potential benefits for using AI-PPPs, but the limitations with its feedback and algorithms are problematic at this point in its development.

**Second Iteration: Categories**

A. Benefits of using the platform  
B. Frustrated with Feedback  
C. “Gaming” the Platform  
D. Expectations for AI-PPP

**First Iteration: Initial Codes**

A. “Great tool for practicing presentation”  
C. “Stare at the ceiling and just speak gibberish”  
A. “Practice presenting information without heavily preparing”  
C. “Speak like a literal cartoon character”  
A. Good for stating things clearly  
C. No focus on meaning  
B. “As much as I tried to change, my score wouldn’t change”  
D. “AI-PPP is good in theory, but not practice”  
B. “I want to get real feedback from real people”  
D. “Initially I thought…”  
B. Feedback did not support improvement  
D. A.I. will improve in this area

---

1 Quotation marks denote in vivo codes, no quotation marks denote sociologically-constructed codes