Using Technology to Support English Language Learners in Higher Education

A Study of Voxy's Effect on English Language Proficiency

January 2019

Ann-Marie Faria | Rebecca Bergey | Ashley Simpson Baird | Alex Lishinski

AIR®
AMERICAN INSTITUTES FOR RESEARCH®

Contents

Tables

Figures

Executive Summary

Although English language learners (ELLs) have an increasing presence in postsecondary education, their unique needs are often unmet by institutions of higher education (Institute of International Education, 2018a; Kanno & Varghese, 2010; Roberge, Siegal, & Harklau, 2009; Sheppard et al., 2015). This challenge partly results from the diversity of ELLs enrolled in postsecondary education—they vary in their educational backgrounds, English proficiency levels, career goals, and reasons for learning English (Espinoza, 2010; Kanno & Varghese, 2010; Roberge et al., 2009; Sheppard et al., 2015). Yet, many institutions are unable to offer courses in English as a Second Language (ESL) that take into account this heterogeneity. As a result, many ELLs spend valuable time and money on remedial courses that do not meet their educational needs, which makes them less likely to persist in their coursework or enter a degree track (Mamiseishvili, 2012; Razfar & Simon, 2011). Given the broad range of backgrounds and experiences of ELLs in higher education, it is a challenging task for programs and instructors to provide the differentiated support that is required to address their needs. Technology-mediated English language instruction may be a solution to the pressing challenges that postsecondary institutions face in providing personalized instruction to ELLs, because it allows instruction to be tailored to an individual's English proficiency, academic needs, and professional interests.

This study specifically tested the efficacy of one such educational technology—Voxy—for improving ELLs' English language proficiency. Voxy (https://voxy.com) is a web- and mobile-based application that supports English language acquisition. It includes a comprehensive suite of adaptive lessons based on authentic English-language resources from a range of media (including articles, academic texts, emails, tweets, videos, audio recordings, and images). The lessons target a range of skills, including reading, writing, listening, grammar, spelling, pronunciation, and vocabulary development. Voxy uses an algorithm to match online content with students' proficiency levels, interests, and academic goals, making differentiated instruction through the online platform seamless and efficient.

The purpose of this quasi-experimental study was to understand whether access to Voxy's online platform as part of the regular coursework in a community college language laboratory improved ELLs' English language acquisition after one semester (16 weeks). The two primary research questions were these: *Does access to Voxy's web and mobile platforms increase ELLs' English language proficiency after one semester? How is student use of the Voxy platform related to English language proficiency?* The study also examined implementation and documented when and how long students typically used Voxy in the lab based on back-end usage data from the online platform.

The study was conducted at Miami Dade College (MDC). Voxy was implemented in nine different sections for the intermediate language labs on campus. The student sample included 317 students enrolled in MDC's language labs: 156 students in the treatment group who had access to Voxy in the lab and 161 students who had enrolled the semester before Voxy was introduced and did not have access to it. Nearly all students (98%) were Latino and Spanish speakers, and 93% received federal financial aid through Pell Grants.

After one semester of use, Voxy improved English language skills for students enrolled in MDC:

■ Students with access to Voxy had larger English language gains than their peers in the comparison group. Adjusting for background characteristics, students with access to Voxy gained an average of 92 points on the Voxy Proficiency Assessment (VPA) between pretest and posttest, compared to only 19 points for students in the comparison group. This statistically significant gain in English language

skills translates to an effect size of 0.41 and corresponds to about 20% of a proficiency level.

- Platform usage was lower than recommended, but students used the platform outside of regular lab times.

  Students, on average, logged in 18 times and spent 6 hours (20.8 minutes a week) in the Voxy online platform throughout the semester.

  However, students consistently used Voxy outside of class time. Three quarters of all time spent in the Voxy platform occurred outside of regularly scheduled language lab times, suggesting that students used the system at times that were convenient for them.

The findings provide promising evidence that Voxy can support English language learning in community college language labs. Even in the context of low implementation, students with access to Voxy learned more English throughout one semester than their peers in the comparison group. This study suggests that educational technologies that provide personalized support for English language learning can improve student language proficiency outcomes for ELLs enrolled in community college. Given that overall usage of the platform was low, more research is necessary to understand how to improve implementation and to track longer term outcomes of technology-mediated language learning in institutions of higher education, especially persistence and progress in community college.

Why This Study?

English language learners (ELLs) are a growing presence in institutions of higher education

Educational institutions in the United States face a growing number of ELLs in prekindergarten through postsecondary classrooms. ELLs account for 1 in 10 students in K–12 public schools nationally, and the percentage is even higher in states with large immigrant enclaves. For example, 22% of students are designated as ELLs in California, 17% in Nevada, and 15% in Texas (National Center for Education Statistics, 2017). In higher education, the number of ELLs is growing steadily, in part because of increases in immigration and in the number of international students coming to study on U.S. campuses. From 1990 to 2014, the number of international students enrolled in U.S. institutions of higher education more than doubled, reaching a total of almost 1.1 million students in the 2017–18 academic year (Institute of International Education, 2018a). Although not all immigrant and international students are considered ELLs, these trends represent a shift toward more linguistically diverse classrooms across the PK–20 grade span.

ELLs have diverse backgrounds, needs, and motivations for learning English

ELLs enrolled in postsecondary education are highly diverse: They vary in their educational backgrounds, English proficiency levels, career goals, and reasons for learning English (Espinoza, 2010; Kanno & Varghese, 2010; Roberge et al., 2009). ELLs in postsecondary education include the broad categories of recent immigrants and refugees, students born in the United States in households speaking a language other than English, and international students. Although international students often come with high levels of formal training in their home country and thus strong literacy skills, they may need specialized support with speaking and listening activities (Sheppard et

al., 2015). On the other hand, students who have been living in the United States for several years have strong oral proficiency but need targeted instruction to support reading and writing development (Kibler, 2014). In addition to having diverse academic experiences, ELLs represent diverse background characteristics and needs. Many immigrant students who enroll in English as a second language (ESL) courses, for example, have responsibilities outside of school, such as full- or part-time work, care for dependents, and financial obligations to their families (Espinoza, 2010; Kanno & Varghese, 2010).

Meeting students' myriad needs creates both programmatic and instructional challenges

Working with limited material resources and staffing, many institutions of higher education uniformly assess, place, and instruct ELLs, resulting in scenarios such as highly educated and undereducated students being placed in classes together. When students arrive in the classroom with such diverse needs, instructors are tasked with the challenge of developing a single curriculum and identifying materials that meet the needs of—and are appropriate for—all students in their classrooms (Mathews-Aydinli & Van Horne, 2006). Many teachers ultimately choose to use traditional materials because these materials are easily accessible or historically have been used in such programs; however, the materials do not meet the individual needs of students (Ignatius, 2016) and require teachers to do significant additional work adapting the materials for their students.

When ELL needs are not adequately addressed, students have limited success

When ELLs are placed in one program or coursework strand, their individual needs may not be adequately addressed. For example, Peyton and Schaetzel (2016) found that many ESL teachers tend not to teach academic writing despite the necessity of learning how to write for

academic or professional audiences. Moreover, teaching writing skills in isolation does not sufficiently prepare students for academic or professional writing (Grubb et al., 2011). Also, when ESL classes are taught in isolation from academic content, students do not acquire the discipline-specific terminology of their future coursework (Parrish, 2015).

When ELLs' needs are not adequately addressed, ELLs show lower rates of persistence and progress than the general population of students in higher education.

For example, one study (Razfar & Simon, 2011) used longitudinal data to analyze the course-taking patterns and language outcomes of Latino ELLs in one large California community college system. Despite students' reported intentions of developing basic academic and career skills (65%), most never advanced beyond the proficiency level at which they started (63%). After just two semesters the majority of ELL students (62%) had dropped out, and after five semesters 85% had dropped out. Similarly, Government Accountability Office data show that, on average, only 40% of postsecondary ELLs advance in English proficiency levels each year. The remaining 60% either drop out or continue to attend classes but do not advance (Kennedy & Walters, 2013).

ELLs are an increasing presence in higher education, especially in community colleges (Rodriguez & Cruz, 2009). As a result, these systems have both the opportunity and the obligation to understand and implement practices that can better support ELLs' language and academic development.

Technology can provide personalized instruction to ELLs

Technology-mediated English language instruction may be a solution to the pressing challenges that postsecondary institutions face in providing personalized instruction to ELLs. In October 2018, the U.S. Department of Education released a report that outlines the ways in which

technology can be used to support ELLs in K–12 settings, and many of the key suggestions apply to higher education. First, technology can be used to assess students for the purposes of programmatic decision making and instructional planning. Educational technology in the classroom also can offer real-time feedback, which can then be used to personalize instruction for each student's English proficiency level, academic needs, and personal interests, with little burden on the teacher. In contrast to a typical curriculum or textbook, which offers a uniform approach for all students, technology can improve learning by offering access to a broad array of authentic real-world content and engaging materials. Furthermore, technology gives students an opportunity to access high-quality instructional materials without restrictions on time and location, thus offering greater flexibility for students with competing priorities.

To date, there is little evidence on how technology-mediated instruction can help ELLs enrolled in postsecondary education

Currently, only a few studies address the ways in which technology can be used effectively to meet the diverse needs of ELLs in postsecondary education, particularly ways to increase their English language proficiency and academic skills. One quasi-experimental study of a web-based supplemental writing program for postsecondary ELLs found that those who received the intervention improved their writing abilities more than those who did not (Al-Jarf, 2004). Specifically, ELLs with access to the online writing program made fewer errors and produced writing samples that were more fluid and proficient compared with peers who did not receive the intervention (Al-Jarf, 2004). More research is necessary, however, to understand how technology can be used to develop ELLs' language skills in reading, writing, listening, and speaking. The current study contributes to the field by examining the impact of Voxy, an

> **Voxy is one educational technology that holds promise for ELLs enrolled in postsecondary education.**
>
> Voxy is a web- and mobile-based technology that tailors English language instruction to individual learners' needs. Before using Voxy, students take a short needs assessment in which they are asked about their interests and academic goals as well as a 30-minute proficiency assessment, and Voxy employs an algorithm to assign content to the students based on these data. Students then begin to engage with Voxy's comprehensive suite of adaptive online lessons. The lessons are based on authentic (i.e., not created for language learners) English oral and written texts from a variety of media, including articles, emails, tweets, videos, audio recordings, and images. Lessons are centered on topics and themes ranging from academic tasks (e.g., report writing and test preparation) to business (e.g., job interviews and presentations). The lessons are organized into thematic units (e.g., entertainment, science, and world news) and as complete courses on specific subjects, like nursing. The platform contains more than 20,000 different lessons, with more added every day. Lessons target a range of skills, including reading, writing, listening, grammar, spelling, pronunciation, and vocabulary development. Voxy lessons are adaptive: They target instruction in the skills that students need to practice the most. Although Voxy also offers virtual live instruction tailored to learners' needs, including writing and other academic skills, these components were not part of the intervention used in this study. Finally, the Voxy platform includes a personalized glossary of key lexical items that students encounter during lessons and a complete guide to English grammar, with references to authentic language in context.

educational technology that can be used to supplement classroom instruction in postsecondary settings.

What This Study Examined

The American Institutes for Research (AIR) conducted a quasi-experimental study of Voxy during the 2017–18 school year to examine the following research questions: *Does access to Voxy's web and mobile platforms increase ELLs' English language proficiency after one semester? How is student use of the Voxy platform related to English language proficiency?*

The study compared the outcomes of students who had access to the Voxy platform with those of students with similar background characteristics who were enrolled in the same coursework but had access to another software program: Tell Me More. The project granted free access to the Voxy platform to a group of intermediate-level ELLs (levels 3 and 4 in a 6-level system) enrolled at Miami Dade College (MDC). The study focused on intermediate learners, because increasing English skills for

these learners is challenging given the plateau in progress they often experience in English language development (Richards, 2008). The AIR team presumed that if gains could be seen among this group of students, then beginning and advanced ELLs likely would show progress if they were granted access to Voxy.

The theory of change posits that Voxy will improve students' English language proficiency

This study is grounded in the theory that using Voxy to supplement instruction for ELLs in English for academic purposes (EAP) classrooms can improve students' English language proficiency skills, ultimately leading to improvements in English language acquisition. This improvement will, in turn, support ELLs' long-term persistence and progress in 2- and 4-year postsecondary institutions (Figure 1). (Note: Measuring this persistence and progress is beyond the scope of this study.)

Instruction incorporating Voxy contrasts with traditional face-to-face EAP instruction, in which providing personalized instruction for a diverse group of ELLs is solely the instructor's

responsibility. For this initial study, the AIR team focused on the short-term educational outcome of English language proficiency (yellow box in Figure 1). However, this work is situated within a more global theory of change and acknowledges the need for additional research to examine intermediate and long-term educational outcomes.

Figure 1. Theory of Change for How Voxy Improves English Language Proficiency, Progress, and Persistence for ELLs in EAP Courses in Community College
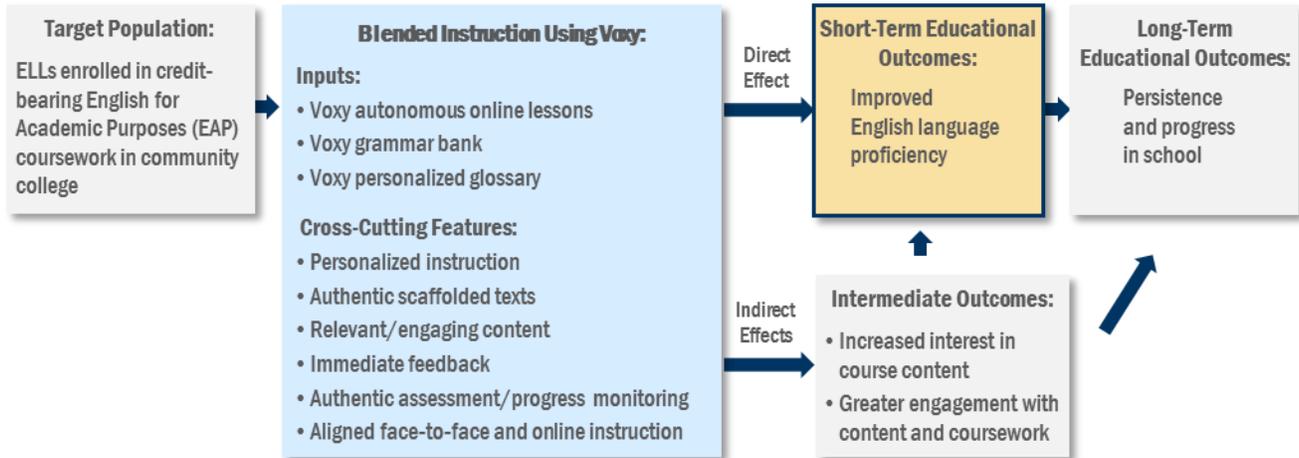
The study was conducted at MDC, where Voxy was introduced for all intermediate students in the spring 2018 semester (January 2018 to May 2018). Using a matched comparison design and a difference-in-differences analytic approach, the AIR team compared the English language gains among intermediate students with access to Voxy in the spring semester to the gains of similar students enrolled the semester before Voxy was introduced. All students in both the fall (comparison) and spring (treatment) semesters completed the Voxy Proficiency Assessment (VPA) at the beginning and end of the semester. This enabled the study team to compare the pre- and posttest English language gains for both groups.

A brief description of the data sources, study sample, measures, and study methodology is presented in Box 1; a detailed discussion is provided in Appendix C.

Box 1. Study Sample, Data, and Methods

The student sample included 317 students enrolled in Miami Dade College's language laboratories for a full semester (16 weeks); 156 students had access to Voxy in the lab (treatment group), and 161 did not (comparison group). All students had 100 minutes per week of lab time, half of which was used for conversation practice and half for self-study using online language-learning tools.

Nearly all students (98%) were Latino Spanish speakers. Twelve percent were U.S. citizens, and the majority of students were from Cuba (61%).

Nearly all students received Pell Grants (93%), and the sample was split about 50/50 between full- and part-time students. Students in the treatment and comparison groups were similar on most background characteristics (e.g., ethnicity, race, immigration status, financial aid, and grade point average) except for full-time status; more students in the treatment group were enrolled part time. See Table A.1 in Appendix A for more details on the student sample.

Treatment Conditions

The students in the treatment group had access to Voxy in the lab, while the students in the comparison group used Tell Me More. All students were enrolled in face-to-face EAP classes with an instructor.

Tell Me More is an English-learning software program that provides lessons in vocabulary, grammar, listening, reading, and pronunciation. Learners move at their own pace through a set curriculum and can access multiple-choice questions related to reading and listening comprehension and grammar. Tell Me More's content is not adaptive, and the delivery of instruction is not personalized. The platform contains some genuine content (i.e., not modified for ELLs), but much of it is scripted and written especially for ELLs.

Data Collection

During the study, the American Institutes for Research (AIR) collected administrative data from MDC, student English language proficiency on the Voxy Proficiency Assessment (VPA), and Voxy platform usage data. Administrative data included student demographics, course enrollment, English language proficiency, and previous education level. The VPA measured the main outcome of interest: students' English language proficiency. The VPA is a valid, reliable English proficiency assessment (Voxy, 2016). Although it is embedded in the Voxy platform, the VPA is a broad measure of English proficiency and is not related to any specific course content. The VPA, which takes between

30 and 60 minutes to complete, tests students' reading, grammar, and listening skills. The VPA was designed by Voxy, and in the development process, Voxy conducted an item-response theory analysis. The VPA aligns with global standards (e.g., Global Scale of English) and has predictive validity with the Test of English as a Foreign Language, the International English Language Testing System, the Test of English for International Communication, and the Pearson Test of English (Voxy, 2016). Voxy usage data included information on students' frequency, duration, and depth of use. See Appendix B for a detailed description of all the data sources.

Methods

To compare the English language gains between students in the treatment and comparison groups, the AIR study team used a difference-in-differences regression model with random effects for labs because students were nested in language labs. To ensure that students in the comparison group were as similar as possible to those in the treatment group, the study team used a matched comparison design. Specifically, propensity score weights were included in the statistical models to adjust for pretreatment differences in student background characteristics. Student covariates were included in the statistical models to improve comparability between groups and increase the precision of the impact estimates. Potential bias from missing outcome data was addressed using multiple imputation by chained equations; see Appendix C for a summary of the missing data rates.

What This Study Found

Students with access to Voxy had larger English language proficiency gains than their peers in the comparison group.

After one semester, students with access to the Voxy platform increased their pre- to posttest gains on the VPA English language proficiency exam more than peers who did not have access to the platform during the fall 2017 semester. The difference-in-differences model showed that the mean proficiency change from pre- to posttest was significantly greater for the Voxy group than for the comparison group (see Figure 2). On average, students with access to Voxy gained 92 points (a 15% increase) on the VPA in one semester, while students in the comparison group gained, on average, only 19 points (a 3.5% increase). As Figure 2 shows, students with access to Voxy (the blue line) started the semester with an average score of 516 and ended the semester with an average score of 608. Students in the comparison group (the gray dotted line) had a pretest average of 528 points on the VPA (a little higher than the students in the treatment group at pretest) but scored only 547 points on average by the end of the semester.

Figure 2. VPA Score Growth: Voxy Versus Comparison
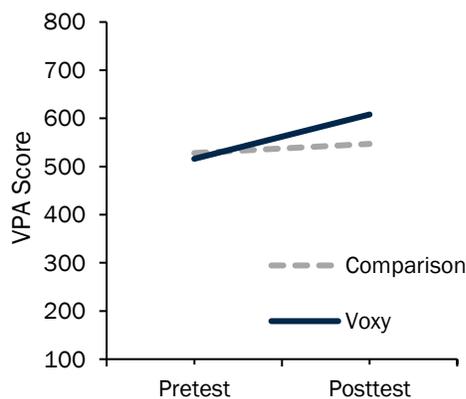
This statistically significant difference ($\beta = 73$, $p < 0.01$) translates to a Hedges' g effect size of 0.41, suggesting that Voxy had a moderate and educationally meaningful impact on ELLs' English language learning. An effect size of 0.25 or larger is considered educationally relevant (What Works Clearinghouse, 2017). The impact of Voxy is nearly double this effect size, suggesting that Voxy had an educationally meaningful impact on English language learning skills for ELLs when used in language labs in community college. An effect size of 0.41 translates to roughly 20% of a proficiency level. While all treatment and control students started the semester at the low end of the Common European Framework of References for Languages (CEFR) entry level language band (B1, sublevel 1), Voxy students gained enough English language to be at the top end of that band by the end of the semester.[1] Sensitivity analyses confirm that the findings are robust and consistent across different analytic approaches (e.g., models with and without covariates, missing data adjustments, and propensity score weights). See Appendix C for a detailed explanation of these analyses and findings.

On average, students used the platform with less frequency, duration, and depth than recommended

AIR analyzed Voxy's back-end usage data to understand how often students interacted with the online platform in language labs at MDC. Specifically, AIR examined the frequency with which students accessed the platform, the duration of their usage, and the depth of their access into the platform (i.e., the different types of activities that they completed and the amount of usage outside of class).

- *Frequency.* At the outset of the study, Voxy recommended that students access the platform

---

[1] The Common European Framework of Reference for Languages, often referred to as CEFR or CEFRL, is an international standard for working out your ability within a language. The six CEFR are A1, A2, B1, B2, C1 and C2. VPA scores between 491 and 630 are equivalent to the B1 sublevel 1 band, scores between 631 and 755 are within the B1 sublevel 2 band, and VPA scores between 756 and 880 are within the B1 sublevel 3 band.

each week during the 16-week semester. However, analysis revealed low to moderate usage levels. Students in the treatment group accessed the platform an average of 18 times throughout the semester. About 75% of treatment students logged in at least 8 weeks (or half of the semester) and less than 1% of treatment students logged in weekly during the 16-week semester.

- *Duration.* At the outset of the study, Voxy recommended that students access the Voxy online platform for 50 minutes during language labs and 40 minutes outside of class each week. In practice, students spent significantly less time in the platform each week. On average, students used Voxy for 362 minutes throughout the semester, or about 23 minutes per week.

- *Depth.* Students in the treatment group had access to Voxy lessons, word banks, and grammar guides. Students spent most of their time (99.8%) completing lessons in the platform. Only 12% of the students used the word bank and grammar guide tools in the platform, and they used these tools for a total of only 4 minutes on average.

In addition, students consistently used Voxy outside of their regularly scheduled language labs. On average, 75 percent of students' time using Voxy occurred outside of class time, suggesting that students used Voxy when it was convenient for them.

Student age was the only background characteristic associated with usage patterns

The AIR team conducted a series of analyses to understand whether any student background characteristics were associated with higher Voxy usage. These analyses tell us more about which students were more likely to engage with the online platform. It may also help us generate hypotheses about which students may thrive using online platforms and which students may need more support to benefit from tech-mediated instructional platforms like Voxy. We

tested background characteristics like age, native language, country of origin, full- or part-time student status, gender, Pell Grant status, and ethnicity. Of these background characteristics, age was the only background characteristic found to significantly relate to usage: Older students used the platform with more frequency and duration. For each additional year of age, students had an additional 0.4 log-ins and spent an additional 10 minutes in the platform.

Students' usage patterns were not associated with gains in English proficiency

The AIR team also analyzed Voxy's back-end usage data to examine associations between platform usage and gains in English proficiency in the treatment sample. We hypothesized that if students used the platform with more frequency, duration, and depth, this could lead to larger English language proficiency gains. To understand this relationship, the team

1. compared the posttest VPA scores of students with high and low usage,

2. estimated the correlations between VPA posttest scores and back-end usage, and

3. estimated multilevel regression models that adjusted for students nested in different language lab sections to determine the statistical significance of the relationship between VPA scores and usage.

All three sets of analyses revealed a small and nonsignificant relationship between usage and gains on the VPA, suggesting that more usage did not necessarily result in large English language gains. See Appendix C for a detailed description of these analyses.

Implications of the Study Findings

Study findings offer promising evidence that Voxy can support English language learning in instructed ESL settings, like community college. Even in the context of low implementation, students with Voxy access throughout the semester improved their English proficiency more than their peers in the comparison group. This study suggests that educational technologies that offer personalized support for English language learning can improve student language proficiency outcomes for ELLs enrolled in community college.

In practice, Voxy can help boost ELLs' language skills when used in community college language labs

This study revealed that even with somewhat minimal usage, access to Voxy was associated with student gains in English proficiency after only one semester. While reviews of computer-assisted learning have highlighted the use of technology to improve student learning outcomes in mathematics (Escueta, Quan, Nickow, & Oreopoulos, 2017), this study represents an important first step in finding evidence of improved student outcomes in English proficiency through the use of technology. Based on the theory of action, the AIR study team attributes these gains to the personalized nature of the Voxy platform. Specifically, besides assessing students' initial English proficiency, Voxy inventories students' goals and interests and then applies a unique algorithm to assign content by English proficiency level, interests, and goals.

Based on these findings, community colleges may consider using personalized technology platforms like Voxy to meet the diverse needs of ELLs. The current study only examined blended learning and therefore the findings may not hold in a context where technology is used without face-to-face instruction. In short, community colleges should allow live instructors to do what they do best, provide real-time feedback as well as human input

and interaction, while allowing technology to do what it does best, provide students with access to authentic real-world content that is personalized to meet their individual needs and interests.

Platform usage was lower than recommended, but students used the platform outside of regular lab times

Given the lower-than-anticipated fidelity, more needs to be learned about how Voxy can best be implemented in a college EAP setting. A future implementation study might consider the following questions: *What is the optimal amount of training for instructors? To what extent do instructors integrate Voxy content into face-to-face instruction, and what does that integration look like? What is the optimal level of support that instructors should provide? What are the optimal levels of frequency, duration, and depth of usage to improve students' English language proficiency?*

However, it was promising that students used the Voxy platform both during language labs and outside of class time. In fact, the majority of use occurred outside of the scheduled language lab time, perhaps suggesting both that students thought Voxy supported their learning and that they chose to access it at times that were convenient to them.

ELLs enrolled in community colleges often have multiple demands on their time, such as full- or part-time work, care for dependents, and financial obligations to their families (Espinoza, 2010; Kanno & Varghese, 2010). Voxy may be one effective instructional tool that provides increased flexibility for when and how they learn. Additionally, self-study with Voxy may reduce some of the burden of extensive face-to-face course time that is typical in IHEs.

There was no statistically significant relationship between usage and VPA change scores

Three sets of analyses resulted in a similar finding—that the number of log-ins and minutes spent in the Voxy system were not statistically

significantly correlated with larger VPA gains. This was not expected. It was hypothesized that more usage would be related to higher gain scores. The lack of a relationship may be due to the small sample size (n = 157 treatment students). This finding could also support an alternate hypothesis: Perhaps some of the students who spent more time in the online platform had lower English language proficiency or faced barriers to using the educational technology, negating the hypothesized relationship. More research is needed with a larger sample of students to better understand the relationship between active engagement in the Voxy online platform and student gains on the VPA.

Research is needed to understand the longer term impacts of Voxy on persistence and progress in school

While the findings from this study are promising, more research is necessary to better understand the longer term and finer grained impacts of Voxy. The findings in this report document the impact of the use of Voxy on English language proficiency after one semester. This study did not measure English language proficiency gains over a longer period or other academic outcomes. Future research should examine whether and how Voxy impacts English proficiency after multiple semesters of implementation as well as how Voxy relates to long-term outcomes such as persistence and progress in degree-seeking programs.

Specifically, Voxy needs to be tested with a larger, more diverse student sample at multiple institutions to understand the impacts of student and institutional factors. Moreover, Voxy should be tested for a longer period. While it is promising to have outcomes for intermediate-level students after only one semester, greater understanding is needed about outcomes for students at other initial proficiency levels over time as well as any plateau or ceiling effects on student language development. Furthermore, future research should examine the long-term impacts on students' persistence, progress, and degree attainment.

Limitations of the Study Findings

This study has several limitations that should be kept in mind when interpreting the findings. These limitations relate to the quasi-experimental nature of the study design, threats to internal validity, the generalizability of the findings, and the validity of the measures used.

The quasi-experimental nature of this study is a limitation because students were not randomly assigned to treatment and comparison groups. The difference-in-differences design compared students between fall and spring semesters. However, the AIR team controlled for differences between groups by using covariates and propensity scores to adjust for pretreatment differences. These statistical adjustments as well as baseline equivalence on all observed background characteristics, combined with examining change scores, helped to strengthen the causal claims that linked Voxy with improved English language proficiency.

Another potential threat to internal validity is differential attrition: There was a significant difference in the degree of attrition between pretest and posttest. The missingness was addressed by using multiple imputation methods to fill in the missing data, but this solution relies critically on the assumption that the missingness pattern was not related to the outcome of interest.

Another limitation of this study is that it was conducted at a single institution; thus, it is unclear whether the results will be generalizable to other contexts. MDC serves a particular population of ELLs, and future research will be necessary to test the efficacy of the Voxy platform for improving English proficiency in other settings with different ELL populations. A related limitation is the subset of students who formed the study sample. The students examined were students of intermediate English proficiency; therefore, future research will be necessary to determine whether the findings can be generalized to students at other proficiency levels.

The use of the VPA as the English language proficiency measure is another possible limitation of the study. While evidence is strong that the VPA is a valid, reliable measure of English language proficiency, the study conclusions depend critically

on this being the case. Therefore, any issues that may be discovered with the quality of the measure, particularly as a measure of change over time, would undermine the validity of the conclusions drawn. Future research should use multiple measures to ensure more robust conclusions.

In sum, the findings offer promising evidence that Voxy can support English language learning in community college language labs. Even in the context of low implementation, students with Voxy access improved their English proficiency more during one semester than their peers in the comparison group. This study suggests that educational technologies that offer personalized support for English language learning can improve student language proficiency outcomes for ELLs enrolled in community college.

References

Al-Jarf, R. S. (2004). The effects of web-based learning on struggling EFL college writers. *Foreign Language Annals*, *37*, 46–56.

Escueta, M., Quan, V., Nickow, A. J., & Oreopoulos, P. (2017). *Education technology: An evidence-based review* (No. w23744). New York, NY: National Bureau of Economic Research.

Espinoza, R. (2010). The good daughter dilemma: Latinas managing family and school demands. *Journal of Hispanic Higher Education, 9*, 317–330.

Grubb, W. N., Boner, E., Frankel, K., Parker, L., Patterson, D., Gabriner, R., … Wilson, S. (2011). *Basic skills instruction in community colleges: The dominance of remedial pedagogy.* Basic Skills Instruction in California Community Colleges No. 2. Stanford, CA: Policy Analysis for California Education.

Ignatius, A. M. (2016). Challenges (what, why and how) of material production for the ESL learners at the tertiary level. *Journal for English Language and Literary Studies, 4*(1), 1–7. Retrieved from http://www.tjells.com/ article/611_IGNATIUS%20S%20X%20C.pdf

Institute of International Education. (2018a). *2018 fast facts.* Retrieved from http://www.iie.org/opendoors

Institute of International Education. (2018b). *International Students by Institutional Type, 2016/2017–2017/2018.* Retrieved from http://www.iie.org/opendoors

Kanno, Y., & Varghese, M. M. (2010). Immigrant and refugee ESL students' challenges to accessing four-year college education: From language policy to educational policy. *Journal of Language, Identity, and Education, 9*, 310–328.

Kennedy, S., & Walters, J. (2013, July). *Repairing the nation's education system for adult English learners*. Arlington, VA: Lexington Institute. Retrieved from http://www.lexingtoninstitute.org/repairing-the-nations-education-system-for-adult-english-learners/

Kibler, A. K. (2014). From high school to the "noviciado": An adolescent linguistic minority student's multilingual journey in writing. *Modern Language Journal, 98*, 629–651.

Mamiseishvili, K. (2012). International student persistence in U.S. postsecondary institutions. *International Journal of Higher Education and Educational Planning, 64*, 1–17.

Mathews-Aydinli, J., & Van Horne, R. (2006). *Promoting success of multilevel ESL classes: What teachers and administrators can do*. Washington, DC: Center for Adult English Language Acquisition, Center for Applied Linguistics. Retrieved from http://www.cal.org/caela/esl_resources/briefs/multilevel.pdf

Miami Dade College. (n.d.). *Highlights and facts.* Retrieved from http://www.mdc.edu/about/facts.aspx

National Center for Education Statistics. (2017). English language learners in public schools. In *The condition of education 2017* (2017-144). Retrieved from https://nces.ed.gov/programs/coe/indicator_cgf.asp

Parrish, B. (2015). *Meeting the language needs of today's adult English language learner* (Issue Brief). LINCS ESL Pro. Washington, DC: American Institutes for Research. Retrieved from https://lincs.ed.gov/sites/default/files/ELL_Increasing_Rigor_508.pdf

Peyton, J. K., & Schaetzel, K. (2016). Teaching writing to adult English language learners: Lessons from the field. *Journal of Literature and Art Studies, 6*, 1407–1423. Retrieved from http://www.davidpublisher.com/index. php/Home/Article/index?id=28304.html

Razfar, A., & Simon, J. (2011). Course-taking patterns of Latino ESL students: Mobility and mainstreaming in urban community colleges in the United States. *TESOL Quarterly, 45*, 595–627.

Richards, J. C. (2008). *Moving beyond the plateau from intermediate to advanced levels in language learning*. New York, NY: Cambridge University Press. Retrieved from http://www.cambridge.org/elt/passages2e/images/booklet_movingbeyondtheplateau.pdf

Roberge, M., Siegal, M., & Harklau, L. (Eds.). (2009). *Generation 1.5 in college composition: Teaching academic writing to U.S.-educated learners of ESL.* New York, NY: Routledge.

Rodriguez, G. M., & Cruz, L. (2009). The transition to college of English learner and undocumented immigrant students: Resource and policy implications. *Teachers College Record, 111*, 2385–2418.

Sheppard, B., Rice, J., Rice, K., DeCoster, B., Drummond-Sardell, R., & Soelberg, N. (2015). Re-evaluating the speaking and listening demands of university classes for novice international students. *ORTESOL Journal, 32,* 1–12.

U.S. Department of Education. (2018). *Educator toolkit: Using educational technology—21st century supports for English learners*. Washington, DC: Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service, National Study of English Learners and Digital Learning Resources.

What Works Clearinghouse. (2017). *What Works Clearinghouse procedures handbook version 4.0.* Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Van Buuren, S., Groothuis-Oudshoorn, K., Robitzsch, A., Vink, G., Doove, L., Jolani, S., Schouten, R., Gaffert, P., Meinfelder, F., & Gray, B. (2018). *Multivariate imputation by chained equations (R package) v. 3.3.0*, https://cran.r-project.org/web/packages/mice/mice.pdf

Voxy. (2016). *VPA: Voxy proficiency assessment.* New York, NY: Author.

Appendix A. Recruitment and Sample Characteristics

This appendix explains the site recruitment process and provides a detailed description of the site and student characteristics, presents a summary of missing data, and describes study participant characteristics.

Site Recruitment

Site recruitment occurred in the spring and summer of 2017. To identify potential sites, the American Institutes for Research (AIR) created a list of community colleges and 4-year colleges with high English language learner (ELL) enrollment using publicly available, site-based enrollment records and national reports on enrollment (Institute of International Education, 2018a). The AIR team contacted 20 sites with an email that outlined the study goals and site responsibilities. Four sites expressed interest, and AIR held a conference call to describe details of the study and to learn more about the site contexts. AIR chose Miami Dade College (MDC) because it had the desired characteristics: It serves a large number of ELLs and has a laboratory component where Voxy could be systematically implemented.

Site Characteristics

MDC is a 4-year public college in south Florida that awards associate's and bachelor's degrees. The largest college in the Florida college system, MDC enrolls more than 56,000 undergraduate students at eight campuses spanning Miami-Dade County. The study was carried out at its InterAmerican campus.

In 2013, MDC had both the largest undergraduate enrollment and the largest Hispanic student enrollment in the United States (Miami Dade College, n.d.). MDC has the eighth largest number of international students served in all U.S. sites that grant associate's degrees (Institute of International Education, 2018b). MDC's noncitizen students come from more than 150 countries and speak 76 different languages (Miami Dade College, n.d.).

English for Academic Purposes Courses

MDC's English for academic purposes (EAP) courses serve U.S. residents and international ELLs who desire to improve their English. EAP courses prepare students to acquire the language skills necessary to matriculate into an academic program of study. EAP courses are offered at 6 proficiency levels, and students are placed in courses based on their scores on the ACCUPLACER assessment. Students advance through proficiency levels on the basis of successful course completion and instructors' recommendations. Full-time students who enter at level 1 (basic proficiency level) typically take 2 years to complete the EAP track. For levels 1 and 2 students, face-to-face courses are offered in speech/listening, reading, writing, and grammar. For students at levels 3 to 6, grammar and writing are combined into one course; reading, speaking, and listening are combined into another. This structure enables students to co-enroll in academic courses for their program of study.

EAP Lab Courses

In addition to face-to-face EAP courses, all students in the EAP program take two lab classes—one for speaking and another for writing. Speaking labs are scheduled for 1 hour and 40 minutes once a week.

- Students in the treatment group consisted of level 3 and 4 students enrolled in a speaking lab during the spring 2018 semester. These students had access to the Voxy platform during their speaking lab times but also were free to access Voxy through personal devices anytime outside of class.

- Students in the comparison group had access to Tell Me More, a language learning program developed by Auralog. These students could access the Tell Me More program only while they were in the lab.

- During lab time, students spent approximately 50 minutes in conversation practice and 50

minutes on Voxy self-study. In addition, Voxy recommended that students use the platform for an additional 40 minutes per week outside of class time, for a total of 90 minutes of Voxy use per week.

Sample Characteristics

In the study, all students in EAP speaking lab sections for levels 1 through 6 took the Voxy Proficiency Assessment at the beginning and end of both the fall 2017 and spring 2018 semesters. The study tested the impact of Voxy on intermediate students. Therefore, all level 3 and 4 (i.e., intermediate) students enrolled in the EAP speaking lab at MDC in the 2017–18 school year participated in the study (n = 317). All students enrolled in speaking labs in the fall 2017 semester made up the comparison group. The comparison group sample comprised 161 students enrolled in 11 different language lab sections. All level 3 and 4 students enrolled in EAP speaking lab sections in the spring 2018 semester made up the treatment group and received access to the Voxy online platform. The Voxy (treatment) group sample comprised 156 students enrolled in nine different language lab sections.

The majority of the student sample were Hispanic (98%), 64% of whom identified as White and 2% as Black. The most common native language was Spanish (55%), but many students reported "other language" (43%), and 2% reported English as their first language. Most of the students were resident aliens (63%), some were refugees/asylees (17%), and others were U.S. citizens (12%). Most students were from Cuba (61%), less than 3% were from other countries in Latin America (Nicaragua, Colombia, Venezuela, Honduras, Peru, Dominican Republic, or Ecuador), 3% were from "other" countries, and 13% did not report their country of origin. Students ranged in age from 18 to 64 years old; the average age was 30 and the median age was 28. Most students (93%) received financial aid through the Pell Grant program. The sample was split by enrollment status, with 53% enrolled as full-time students and 47% enrolled as part-time students. The average grade point average from the semester before Voxy was implemented (Fall 2017) was 2.85.

Students in the treatment and comparison groups were similar in all but one background characteristic: full- versus part-time enrollment status. More students in the treatment group were enrolled as part-time students than in the comparison group. Follow-up analyses that estimated the impact of Voxy separately for full-time and part-time students revealed that the treatment effect was larger for part-time students. Part-time students gained an average of 71 points on the VPA, while full-time students gained 42 points. This finding warrants further exploration in future studies that focus on whether Voxy is more helpful for part-time students in community colleges.

Table A.1 provides a summary of background characteristics for both groups. All other background characteristics were equivalent across treatment and comparison samples.

Table A.1. Student Characteristics for Treatment and Comparison Groups

Student Characteristic	Control Group Mean	Voxy Group Mean	p	Standardized Mean Difference
Received Pell Grant	93%	94%	0.76	-0.03
Enrolled full time	73%	38%	0.00	0.75
White	60%	62%	0.61	-0.06
Black	4%	5%	0.70	-0.04
Hispanic	99%	97%	0.23	0.15
U.S. citizen	15%	11%	0.27	0.13
Male	33%	35%	0.70	-0.05
Age (years)	29.83	28.97	0.82	-0.03
Voxy pretest scores	528	516	0.80	-0.03
Cumulative grade point average	2.82	2.90	0.53	-0.07

Note. The standardized mean difference is the difference between the control group mean and the treatment group mean divided by the standard deviation of the control group, adjusting for the sample size (Hedge's *g*). This value provides the reader with an estimate of the magnitude of the difference; values above 0.25 are considered nonequivalent.

Appendix B. Data Collection

Data Sources

Data for this study came from three sources: administrative data to help in identifying student background characteristics; the Voxy Proficiency Assessment (VPA), which measured students' English language proficiency; and back-end usage data to help in understanding how students used the Voxy platform during the 2018 spring semester.

Administrative Data

AIR entered into a data-sharing agreement with Miami Dade College to obtain administrative data for all students enrolled in language labs at the InterAmerican campus during the fall 2017 and spring 2018 semesters. Demographic data included age, gender, ethnicity, race, primary language, immigration status, international student status, education level, country of prior education, name of high school, financial aid status, part- or full-time enrollment status, program of study, enrollment data, lab course section, lab meeting day and time, ACCUPLACER scores, ACCUPLACER test dates, course grades, and grade point average.

Voxy Proficiency Assessment

The VPA measured students' English language proficiency at pretest and posttest. The VPA is embedded in the Voxy platform and is an objective and reliable online test that evaluates an ELL's English proficiency level (Voxy, 2016). The VPA was developed by Dr. Katie Nielson and her team at Voxy and was used as the primary outcome in this study. The VPA assesses English proficiency in two skill areas: integrated reading and grammatical competence and listening comprehension. The VPA is scored from 0 to 1,680, and it places learners into seven different proficiency categories, from beginner to advanced. There are multiple forms of the VPA at each proficiency level, and each form has an equal number of questions of varying levels of difficulty. All forms assess the three skill areas.

The VPA is aligned with global standards, such as the Common European Framework of Reference for Languages and the Global Scale of English, and it can predict results on common tests such as the Test of English as a Foreign Language, the International English Language Testing System, the Test of English for International Communication, and the Pearson Test of English (Voxy, 2016).

Learners take the VPA as an initial benchmark assessment before using the platform and then periodically to assess their progress. The VPA is administered through the Voxy platform and takes between 30 and 60 minutes to complete. Students in both the treatment and comparison groups took the VPA at the beginning and end of the fall 2017 and spring 2018 semesters. All VPA data were transmitted to AIR through a secure file transfer protocol site.

Voxy Usage Data

The Voxy platform collects back-end usage data that tracks students' interactions with the platform, recording every log-in session, along with durations of time spent on each of the different types of activities. Back-end usage data from the Voxy platform were analyzed on the dimensions of frequency, duration, and depth. *Frequency* refers to the number of times that each student logged on to the platform and the number of weeks (out of 16 possible weeks) that students logged on during the semester. *Duration* is the aggregate amount of time that students spent in the platform throughout the study, excluding time spent completing the pretest and posttest VPA assessments. *Depth* is the amount of time that students spent using the various components of the platform— lessons, grammar guide, and word bank, as well as the amount of time they spent using the platform outside of the lab meetings. All usage data were transmitted through a secure file transfer protocol site.

Appendix C. Analytic Methods and Detailed Findings

Impact Analyses

The impact analyses were conducted using a difference-in-differences design to determine the difference in pretest to posttest change between the Voxy group and the comparison group. To estimate the difference in change between the treatment and comparison group, we used a multiple regression model that includes a term for pretest or posttest, for treatment or comparison, and for the interaction between the two. The interaction term represents the impact estimate of interest. The basic model specification is as follows:

$$Y_{st} = \beta_0 + \beta_1 Treat_s + \beta_2 Post_t + \beta_3(Treat_s \times Post_t) + \epsilon_{st}$$

Y_{st} is the observed outcome in group s and period t;

$Treat_s$ is a dummy variable set to 1 if the observation is from the "treatment" group in either time period;

$Post_t$ is a dummy variable set to 1 if the observation is from the posttreatment period in either group; and

β_3 is the difference-in-difference estimate of the treatment effect.

The primary impact model included covariates to increase the precision of the estimate, propensity weights to adjust for any pretreatment differences on background characteristics, multiple imputation using chained equations to address missing data, and random effects for the nesting of students within their language labs.

Sensitivity Analyses

Several different iterations of the impact model were run to ensure that results were robust to the analytic choices made. Three primary decision points for the analytic methods were used to generate the impact estimates: (a) how to deal with missing posttest data, (b) how to address the clustering of students within classrooms, and (c) how to address the possibility of pretreatment differences between the Voxy and comparison groups. A total of six different versions of the impact model are presented here, with all six showing statistically and practically significant impacts of using the Voxy platform.

Missingness. Rates of missing data were low for administrative data and the Voxy back-end usage data. However, rates of missing data on the VPA posttest were high, with over a third of students in the comparison group missing the posttest VPA. Of special note, students in the treatment group had lower rates of missingness (16% for level 3 students and 7% for level 4 students) than students in the comparison group (31% for level 3 students and 35% for level 4 students). This outcome is promising in that using Voxy may have positive impacts on students' attendance and long-term persistence and progress. The study team, however, needed to adjust for the differential missingness.

To deal with the missing posttest data, the AIR team employed two different approaches: listwise deletion and multiple imputation using chained equations. Listwise deletion was used in a sensitivity model where the students missing VPA posttest scores were removed from the analysis sample. Multiple imputation using chained equations was used in the primary impact model, using imputed posttest data. This method was implemented using the MICE package in R (van Buuren et al., 2018). The analysis was conducted by first creating 10 imputed datasets (imputed separately for treatment and control) using Bayesian linear regression. Then the impact models were fit for each of the 10 imputed data sets and aggregated to create final impact estimates.

Clustering. To address the clustering of students within the classroom, two different approaches were used: including class section as a fixed effect in the multiple regression model and including class section as a random effect in a hierarchical linear model. Students were enrolled in 18 different language lab sections, and intraclass correlation coefficient values for VPA scores between these sections suggested that this clustering needed to be accounted for (ICC – 0.12). In three of the versions of the impact model, the class section was included as a dummy-coded factor variable in the basic difference-in-differences multiple regression model, alongside the other pretreatment covariates. In the other three versions of the impact model, the class section was included as a random effect in a hierarchical linear model fit using the lme4 package in R.

Matched comparison. To find the comparison group most similar to Voxy, two approaches were used: including pretreatment covariates in the difference-in-differences model, and propensity-score-weighting adjustments based on the pretreatment covariates. All the impact models included the set of pretreatment covariates along with the difference-in-differences terms.

This includes the models using propensity-weighting adjustments, which were "doubly robust" models because they included the weighting adjustments in addition to the covariates being included in the difference-in-differences regression model. Propensity scores were created using the TWANG package in R, which uses generalized boosted regression to create propensity scores from an optimal combination of covariates. These weights were used to fit both the multiple regression and hierarchical linear impact models. The propensity-score-weighting adjustment was used with the imputed data sets but not the listwise deleted data set.

Impact Estimates Across Model Specifications

The six impact models were specified as follows: (a) listwise deleted with fixed effects for classrooms, (b) listwise deleted with random effects for classrooms, (c) multiple imputation with fixed effects for classrooms, (d) multiple imputation with random effects for classrooms, (e) propensity score weights plus multiple imputation with fixed effects for classrooms, and (f) the primary impact model, using propensity score weights plus multiple imputation with random effects for classrooms. All six models were difference-in-differences

Table C.1. Impact Estimates by Outcome Model Specification

	Coef	SE	t	Df	p	Hedges' g change	Hedges' g post
Listwise deleted—fixed classroom effects	52.5	24.0	2.2	541.0	0.029	0.35	0.30
Listwise deleted—random classroom effects	58.1	17.0	3.4	575.0	0.001	0.39	0.33
Multiple imputation—fixed classroom effects	69.8	26.3	2.7	146.8	0.008	0.47	0.40
Multiple imputation—random classroom effects	69.8	25.4	2.7	130.4	0.006	0.47	0.40
Multiple imputation and propensity score weights—fixed classroom effects	73.0	26.1	2.8	176.7	0.005	0.49	0.41
Primary impact model with multiple imputation and propensity score weights—random classroom effects	73.0	25.2	2.9	160.0	0.004	0.49	0.41

multiple regression models that included the set of pretreatment covariates. All six impact estimates were statistically and practically significant, suggesting the findings are robust to changes in model specification. Table C.1 shows the impact estimates for all five sensitivity analyses and the primary impact model.

Voxy Usage Data

The AIR team used back-end usage data from the Voxy platform to summarize the ways in which students interacted with the platform. Back-end usage data from the Voxy platform were analyzed on the dimensions of frequency, duration, and depth. Table C.2 shows summary data on students' frequency and duration of use.

Table C.2. Frequency and Duration of Voxy Use

Voxy Usage	Mean	*SD*	Range	Median
Frequency				
Number of log-ins	8.92	7.97	0–40	7
Weeks logged in	4.83	5.89	0–16	5
Duration				
Total usage time (minutes)	362.15	445.09	0–2,633.28	210.53
Depth				
Percent of Voxy use outside of class time	75.00	28.57	0-100.00	83.16

Relationship Between Voxy Usage and English Language Proficiency Gains

AIR also conducted a series of analyses to

1. compare the posttest VPA scores of students with high and low usage,

2. estimate the correlations between VPA posttest scores and back-end usage, and

3. estimate multilevel regression models that adjusted for students nested in different language lab sections to determine the statistical significance of the relationship between VPA scores and usage.

All three sets of analyses revealed a small and nonsignificant relationship between usage and gains on the VPA.

First, we compared the posttest VPA scores for students that were high or low users of the Voxy platform. We defined high use as above the median of time spent in the platform and low use as below the median of time spent in the platform. There were no differences on VPA posttest (controlling for pretest). We also split the sample into quartiles and again found no relationship. Dividing users into quartiles of time spent on the platform or number of log-ins revealed no differences on posttest (controlling for pretest).

Second, we estimated the raw correlations between the VPA change scores and frequency and duration of Voxy use. We found that the correlations were low ($r = 0.10$) and not statistically significant. See Figure C.1 for a scatterplot of the relationship between usage and VPA change scores. Note that the low correlation might be due to the small sample size (student $n = 156$).

Third, we estimated multilevel regression models that adjusted for students nested in different language lab sections, pretest, and student background characteristics. These regressions also revelated no relationship between backend usage and VPA gains.

The scatterplots also reveal a few outliers, or power users, who used the Voxy platform with much more frequency and duration than other users. When we remove these users, the pattern remains the same—there is little to no relationship between usage and gains in English proficiency in the treatment sample.

Figure C.1. Use of the Voxy Platform (Frequency and Duration) and VPA Change Scores