The impact of the pandemic on student Spanish language proficiency

Jesse Gleason, Southern Connecticut State University
Resha Cardone, San Francisco State University
Andrew Bartlett, Southern Connecticut State University

Abstract

In the aftermath of the COVID-19 pandemic, we continue to take stock of student learning. Although the “crisis-context” (Gacs et al., 2020) move to fully online instruction may be over, a complete understanding of how student outcomes have been impacted remains. The present study focuses on how students’ Spanish language proficiency, as measured by the STAMP test, was affected by moving all on-ground language courses online in March 2020 at a small public university in the northeastern US. Comparing overall student Spanish language proficiency as well as reading, writing, speaking, and listening skills across 30 sections of a third-semester Spanish course before and during the pandemic, we examined student learning outcomes based upon the modality of instruction. Results revealed a significant increase in students’ overall Spanish language proficiency and significant increases in sub-level proficiencies in three out of the four skills in the online modality. Thus, in spite of the many changes that took place as a result of the COVID-19 pandemic, students’ Spanish language proficiencies were either positively impacted or unimpacted. We discuss the implications of these results and pose questions for on-ground and online language courses moving forward.

Keywords: Language Proficiency, Pandemic, Online Learning, On-ground Learning

Language(s) Learned in This Study: Spanish

APA Citation: Gleason, J., Cardone, R., & Bartlett, A. (2024). The impact of the pandemic on student Spanish language proficiency. Language Learning & Technology, 28(1), 1–17. https://hdl.handle.net/10125/73561

Introduction

The COVID-19 pandemic brought to academic contexts innumerable examples of change. Although universities are still coming to grips with the greater impact of the pandemic across all areas of learning, in certain contexts dust has begun to settle, allowing us to take stock of the significance of "crisis-context" pedagogies (Gacs et al., 2020; Ross & DiSalvo, 2020) on student learning outcomes. With the move to fully online instruction in the spring of 2020, different contexts were certainly differentially impacted (Moser et al., 2021). Some disciplines and departments that had already offered courses in fully online or hybrid contexts may have observed less of a change in terms of their students' overall learning experiences. Other disciplines may have had to grapple with pedagogical, logistical, and curricular changes brought about by the swift pivot to fully online instruction (Tao & Gao, 2022). The present study looks at one such context, homing in on a large, lower-division Spanish language program that, prior to the pandemic, offered solely on-ground instruction.

In order to answer questions related to how the pandemic impacted Spanish student language proficiency, we examined STAndards-based Measurement of Proficiency (STAMP) test results from a third-semester Spanish course at a small public university in the northeastern U.S. prior to and during the pandemic. With an otherwise consistent set of curricula, instructors, and pedagogies, we examined student proficiency
outcomes after two-and-a-half semesters of fully online instruction. This unique set of circumstances allowed us to ascertain that students’ overall Spanish language proficiency as well as their sub-skills across reading, listening, writing, and speaking either remained constant or improved with the move to fully online instruction. As such, this study addresses an important gap in the literature on commensurability of student proficiency outcomes in online and on-ground courses, leading to pose additional questions about world language teaching and learning.

Proficiency Development and Technology-Mediated Language Instruction

The American Council on the Teaching of Foreign Languages (ACTFL) defines language proficiency as “what individuals can do with language in terms of speaking, writing, listening, and reading in real-world situations in a spontaneous and non-rehearsed context” (ACTFL, 2012, p. 3). The Language Proficiency Guidelines (ACTFL, 2012) describe the characteristics of individuals able to perform differentially across the four skills in terms of five major levels of proficiency: novice, intermediate, advanced, superior, and distinguished. As a measure of functional language ability, the guidelines have been useful in the research and development of a variety of language proficiency tests, which can now classify learners according to their ability to use language in situations in which they would need to communicate in the target language.

For decades, the field of Computer-Assisted Language Learning (CALL) has been examining the issues of students’ Spanish language proficiency in a variety of technology-mediated language instructional contexts (Bahari, 2021). The effectiveness of online, blended, hybrid, or technology-enhanced language instruction for developing student language proficiency has been assessed in a variety of instructional contexts, including in blended/hybrid courses often with a succession of regular, ongoing technology-mediated tasks, as well as in more traditional courses that have a limited scope for technology or may not use it at all for language instruction (Plonsky & Ziegler, 2016). The field has also seen studies that compare student proficiency outcomes in fully on-ground language courses to those drawn from students in fully online courses (e.g., Aldrich & Moneypenny, 2019), the latter of which have been growing in popularity over the past decade as the technological possibilities expand. As these technologies and tech-enhanced instructional practices evolve, so too have our classrooms and students, such that a gradual incorporation of technology has seemed to become slowly but steadily more and more normalized (Bax, 2003, 2011; Goertler, 2019).

Despite their growing prevalence, online language learning can be evermore elusive to pin down and describe. As Blake (2015) asserts, “online courses vary wildly in terms of their quality, affordances, and practices, anywhere from a self-paced approach that relies on commercial software such as Rosetta Stone to other more responsible online offerings with significant interactive components” (p. 409). This fact makes comparison of modality across institutions or contexts a challenge, as controlling variables in online modalities in order to measure student proficiency is only reliable in similar contexts.

By and large, comparative studies from the better half of a century of comparative CALL research has been limited to intra-institutional research (e.g., Chenoweth et al., 2006) and meta-analyses (e.g., Grgurović et al., 2013). These studies overwhelmingly point to the commensurability of online and on-ground teaching for developing students' language proficiency, with few major observable differences in terms of student proficiency outcomes based upon course modality.

In the early 2000s, for example, several studies evaluated the effectiveness of hybrid/online and on-ground elementary and intermediate French and Spanish courses at Carnegie Mellon University (Chenoweth et al., 2006; Murday et al., 2008). Chenoweth et al. (2006) compared student learning outcomes in hybrid/online and on-ground courses as measured by the various sections of the final exam (writing, reading, listening, grammar, and vocabulary) and oral production, finding only two instances of statistically significant differences between multiple groups: (a) hybrid/online intermediate Spanish students significantly outperformed on-ground students on several measures of oral production (fluency, comprehensibility, syntax, and grammar) and (b) on-ground elementary French students significantly outperformed their hybrid/online peers in several measures of written production (vocabulary and ability to remain on-topic).
All other measures of student language proficiency (i.e., reading, listening, grammar, and vocabulary) across hybrid/online and on-ground courses remained constant.

In the early 2010s, two meta-analyses also revealed comparable outcomes between on-ground and online learning more generally and in particular for language learning. The U.S. Department of Education (2010) surveyed 176 studies of online versus on-ground learning, finding that online learners only performed slightly better than their on-ground peers. A meta-analysis of 37 studies examining the effectiveness of online language instruction showed that world language instruction supported by technology was at least as effective at building students’ language proficiency as instruction without technology (Grgurović et al., 2013).

More recently, advances in language proficiency assessment have enabled the reliable measurement of larger swaths of language learners (Lin & Warshauer, 2015; Tarone, 2015). Proficiency testing is now being used across K16 educational contexts for different purposes, including program accreditation. Common assessments include the Versant Test (Bernstein et al., 2010), the ACTFL Assessment of Performance Toward Proficiency in Languages (AAPPL) (Cox & Malone, 2018), and the STAMP (Santos, 2019).

Several studies have harnessed these assessments in order to reliably compare larger groups of language learners. Blake et al. (2008), for example, used the Versant for Spanish test to compare the oral proficiency outcomes for 334 students in blended/hybrid and on-ground courses, finding no statistical differences in terms of students’ overall levels of oral proficiency or the subconstructs of student speaking proficiency, including pronunciation, fluency, sentence formation, and vocabulary production.

A decade later, three studies (Aldrich & Moneypenny, 2019; Moneypenny & Aldrich, 2016, 2018) in particular supported the findings of Blake et al. (2008) and took them a step further to compare fully online courses to on-ground courses. Not only did these studies also use the Versant test to specifically compare student oral language proficiency outcomes, they also largely concluded that students in fully online Spanish courses achieved equivalent oral proficiency outcomes as their peers in on-ground courses.

Whereas earlier research has compared student learning outcomes in blended/hybrid and on-ground courses using criterion-referenced tools, such as final exams (Chenoweth et al., 2006), and other studies have relied primarily on oral proficiency (e.g., Moneypenny & Aldrich, 2019), additional research is needed to shed light on how overall student proficiency and sub-skill language proficiencies are impacted by the move to a fully online course modality. As mentioned, language proficiency has been defined by ACTFL as “what individuals can do with language in terms of speaking, writing, listening, and reading in real-world situations in a spontaneous and non-rehearsed context” (ACTFL, 2012, p. 3).

Karatay and Hegelheimer (2021) report that the COVID-19 pandemic has pushed even low-resource environments to find ways to integrate technology into their language courses. Certainly, and as González-Lloret (2020), Gacs et al. (2020), and others have asserted, planned, intentional, and “collaborative” language learning will appear distinct from the “crisis-prompted” pedagogies taken up halfway through the spring 2020 semester. In many contexts, online language learning has the potential to take on a greater status of “the new normal” (Egbert, 2020). Whereas blended, hybrid, flipped, and technology-enhanced language instruction has gradually become normalized (Bax, 2003, 2011; Goertler, 2019), little research has compared proficiency outcomes in fully on-ground and fully online courses. Furthermore, as Van Deussen-Scholl (2015) claims, although much of the research points to comparability of student learning outcomes between online/hybrid and on-ground courses, there is a clear lack of research that systematically assesses the effectiveness of online language learning.

This study aims to address this gap, focusing on third-semester Spanish student STAMP scores prior to and during the pandemic using a holistic construct of language proficiency across the four skills as measured by a norm-referenced assessment to ensure validity and reliability of results. In harnessing the unique circumstances of the COVID-19 pandemic to understand its impact on student language proficiency, as measured holistically and across the four skills, a number of questions still remain. In the interest of documenting this unprecedented moment, we pose the following questions:
1. Was there a significant difference in overall Spanish language proficiency scores between on-ground and online students in respective spring 2019 and 2021 semesters?
2. Was there a significant difference in proficiency skill (reading, writing, listening, and speaking) scores between on-ground and online students in respective spring 2019 and 2021 semesters?

Methods

Research Context

The context for the present study was a small public university in the northeastern United States. The university responded to the “crisis phase” of the COVID-19 pandemic, which, for the purposes of this study, constituted six-weeks into the spring 2020 semester, the entire fall 2020 semester and the spring 2021 semester, with temporary modifications to several university policies (e.g., late withdrawal, incomplete grade, and pass/fail policies). The purpose of these temporary revisions was to improve student success and retention, while recognizing the varied impacts of the pandemic on the student population. By the fall 2021 semester, however, the university returned to pre-pandemic policies and procedures and administrators clarified that it would no longer request that faculty exercise greater flexibility due to the pandemic.

At the time of this study, students at this university were required either to pass a third-semester language course in a world language or to demonstrate an ACTFL level of Intermediate-Low language proficiency on an external assessment (AAPPL or STAMP) to fulfill their general education requirement. Up until the spring 2020 semester, all sections of the third-semester Spanish course had been taught in an on-ground course modality, meeting 200 minutes per week face-to-face, 50 minutes of which occurred with the instructor leading the class in the language laboratory space.

At the end of the third semester course, students took a proficiency test known as STAMP 4S from Avant Assessment (Santos, 2019) as their final exam in the course. STAMP measured their reading, writing, listening, and speaking skills and, as shown in Figure 1, test-taker scores on the STAMP corresponded to the sub-levels of the ACTFL proficiency scale. The university's proficiency benchmark was for students to receive a minimum score of 4 (Intermediate-Low) for each of the four skills at the end of the third semester course, ideally receiving an overall score of 16 on the STAMP. As one can see, scores of “1,” “2,” and “3” correspond to ACTFL proficiency scores of Novice-Low, Novice-Mid, and Novice-High, respectively.

Figure 1

STAMP Scores as Related to ACTFL Proficiency Scale (Avant Assessment, 2023)

<table>
<thead>
<tr>
<th>Reading and Listening Level Key</th>
<th>Writing and Speaking Level Key</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Novice</strong></td>
<td><strong>Novice</strong></td>
</tr>
<tr>
<td>1 - Novice-Low</td>
<td>1 - Novice-Low</td>
</tr>
<tr>
<td>2 - Novice-Mid</td>
<td>2 - Novice-Mid</td>
</tr>
<tr>
<td>3 - Novice-High</td>
<td>3 - Novice-High</td>
</tr>
<tr>
<td><strong>Intermediate</strong></td>
<td><strong>Intermediate</strong></td>
</tr>
<tr>
<td>4 - Intermediate-Low</td>
<td>4 - Intermediate-Low</td>
</tr>
<tr>
<td>5 - Intermediate-Mid</td>
<td>5 - Intermediate-Mid</td>
</tr>
<tr>
<td>6 - Intermediate-High</td>
<td>6 - Intermediate-High</td>
</tr>
<tr>
<td><strong>Advanced</strong></td>
<td><strong>Advanced</strong></td>
</tr>
<tr>
<td>7 - Advanced-Low</td>
<td>7 - Advanced-Low</td>
</tr>
<tr>
<td>8 - Advanced-Mid</td>
<td>8 - Advanced-Mid</td>
</tr>
<tr>
<td>9 - Advanced-High</td>
<td>NR - Not Ratable</td>
</tr>
</tbody>
</table>

By using STAMP to assess student proficiency, language department stakeholders and university administrators have been able to appreciate changes in student proficiency outcomes over time. Beginning in the fall semester of 2013 and continuing through the fall semester of 2019, the STAMP test had been consistently administered to students in the third-semester Spanish course in an on-ground, face-to-face modality. Students would take the STAMP in the language laboratory during their final exam period.

However, with the pivot to online instruction in the middle of the spring 2020 semester, STAMP could not be feasibly administered in the third-semester Spanish course for two consecutive semesters (spring of 2020
and fall of 2020) until the spring semester of 2021. However, in spring semester of 2021, Avant Assessment and the Department collaborated to offer the STAMP test remotely to all students completing the third-semester language course.

In terms of instructional changes, when the COVID-19 pandemic hit in the middle of the spring semester of 2020 and all on-ground sections of the third-semester Spanish course moved swiftly to a fully-online modality, the coordinator of the Lower-Division Spanish Program (LDSP) endeavored to keep the curriculum as consistent as possible, including grade allocations, which remained constant. Both before and during the pandemic, instructors were asked to assign students 30 minutes of online homework activities in the online platform accompanying the course textbook (Blanco & Donley, 2019) known as “the Supersite” per hour of class meeting time, for a total of 150 minutes per week.

The most significant change that was made during the move to online instruction was that the 50-minute on-ground lab was replaced with 50 minutes of asynchronous lab activities in the Supersite. After the pandemic hit, the LDSP coordinator urged instructors to continue to assign homework activities that allowed students to work individually, largely focusing on form (e.g., vocabulary, grammar, and reading activities), however, instructors were asked to have their students complete more communicative/interpersonal types of activities in the Supersite, such as partner chats in which students would receive instructions for a task and were guided to complete roleplays which they could then enact and record. Other types of asynchronous lab activities in the Supersite included simulated conversations which prompted students to respond to an avatar, replicating the types of spontaneous interactions that might be required of them in the target language culture.

Participants

The participants in this study included 568 students taking one of 30 sections of a third-semester general Spanish course during the spring 2019 and spring 2021 semesters. Although specific information about the participants is unavailable, anecdotally we can surmise that the majority of students were likely between the ages of 18 and 21 and in their first or second year of university study. There were 9,485 students in spring of 2019 (5,993 females and 3,493 males) and 8,648 students in the spring of 2021 (5,668 females and 2,980 males).

The instructors of the third-semester Spanish courses included two full-time tenured faculty members in the Department of World Languages and Literatures and 17 part-time adjunct faculty. These instructors had a range of digital literacy and familiarity with online teaching; several of them taught online language courses on a regular basis, while others had never taught online before the pandemic semesters. All had access to educational technology training workshops offered by the language lab and the university. All but one of the instructors had taught the course on-ground before the spring 2020 pandemic semester and all instructors had taught the course online at least twice before (i.e., in the second half of the spring 2020, fall 2020, and spring 2021 semesters).

Data Collection

The participant data harnessed in this study included STAMP test scores for students enrolled in 30 sections of a third-semester Spanish course in the spring semesters of 2019 and 2021. All data were provided by the external assessment company Avant Assessment. Participant data were given an identifying code and all names and identifying markers were eliminated in order to protect participant confidentiality.

STAMP Proctoring

The STAndards-based Measurement of Proficiency (STAMP) test (Santos, 2019), a computer-adaptive test that measures student proficiency in a world language, was proctored on-ground in the language lab prior to the pandemic and remotely post-2020. For remote proctoring, Avant Assessment implemented robust measures in spring 2021 to ensure the reliability of the online test results. These included a trained online human proctor, artificial intelligence monitoring software, a lockdown browser, screen recording, test-taker webcam feed, and flagging/reporting of suspected instances of academic misconduct.
**STAMP Reliability**

The STAMP 4S Spanish test is accredited by the American Council on Education (ACE). Information about the internal reliability of the STAMP 4S Spanish for the Listening and Reading sections of the test is provided in Table 1 (Santos, 2022). Inter-rater reliability statistics for the Writing and Speaking sections of the test, wherein test takers respond to three prompts, are also provided in Table 1 (Santos, forthcoming).

<table>
<thead>
<tr>
<th>STAMP Section</th>
<th>Reliability (Cronbach’s alpha)</th>
<th>Avg. CTT Item Difficulty (Range)</th>
<th>Avg. CTT Item Discrimination (Range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>0.90</td>
<td>0.61 (0.26–0.98)</td>
<td>0.42 (0.20–0.58)</td>
</tr>
<tr>
<td>Listening</td>
<td>0.89</td>
<td>0.58 (0.28–0.95)</td>
<td>0.41 (0.23–0.55)</td>
</tr>
<tr>
<td>STAMP Section</td>
<td>Exact Agreement / Exact + Adjacent Correlation</td>
<td>Standardized Mean Difference (SMD)</td>
<td></td>
</tr>
<tr>
<td>Writing</td>
<td>0.84 / 0.99</td>
<td>0.96</td>
<td>-0.01</td>
</tr>
<tr>
<td>Speaking</td>
<td>0.80 / 0.99</td>
<td>0.96</td>
<td>0.01</td>
</tr>
</tbody>
</table>

**Data Analysis**

The data analysis for this study involved non-parametric analyses performed in SPSS on the ordinal data output provided by the STAMP proficiency overall and sub-skill scores for students enrolled in third-semester Spanish courses in the spring 2019 and spring 2021 semesters. In particular, we used non-parametric independent-samples Mann-Whitney U tests in order to ascertain differences in overall median STAMP proficiency scores and sub-test skill scores between the data obtained from students in two semesters, pre- and post-2020 (before the pandemic semester in spring, 2020 and afterwards). Initially, because of their timeline proximity, we chose to compare students' STAMP scores from the fall semester of 2019 to the spring semester of 2021, with fall semester 2019 being the last semester before the COVID-19 pandemic semester (spring 2020), when courses were offered fully on-ground. However, we later reasoned that because of differences in student demographics across fall and spring semesters, we would compare only spring semesters.

To perform the data analysis, multiple nonparametric Mann-Whitney U tests were performed in R (R Core Team, 2018) on ordinal student STAMP scores (response variable). As nonparametric tests do not rely on normal distributions, where certain outcomes are more likely, the use of the median rank values generated distributions from the actual data in order to determine significant differences. Further, using a nonparametric approach when dealing with non-normal data allowed us to detect with a higher level of confidence significant differences between the groups.

With the main objective of detecting differences between two independent populations pre (on-ground) and post (online) the nonparametric Mann-Whitney U test had a higher statistical power. The assumption of independence was met as the students from the pre (on-ground) population took the STAMP exam in the Spring of 2019 and the post (online) took the exam in the Spring of 2021.

**Results**

**Overall Proficiency Outcomes: Total STAMP Scores**

In order to answer Research Question 1 (was there a significant difference in overall Spanish language proficiency scores between on-ground and online students in respective spring 2019 and 2021 semesters?), we conducted a non-parametric independent samples Mann-Whitney U test to compare the median ranked
overall STAMP scores for Spanish proficiency for on-ground pre-pandemic and online post-2020 sections of third-semester Spanish.

Descriptive statistics comparing scores reveal a marked increase in proficiency results both overall as well as for three of the four skills (listening, reading and speaking) in the online courses offered during the pandemic, as shown in Table 2. Under the null hypothesis that there is no difference in distribution between STAMP scores in the Spring 2019 compared to the distribution of STAMP scores in the Spring 2021, the Mann-Whitney U test revealed a significant p-value (< .001) (see Table 3). Figure 2, below, shows online students outperforming their on-ground peers, thus we have evidence to conclude that students from the spring 2021 semester (on-line) obtained significantly higher levels of proficiency than students from the spring 2019 semester (on-ground).

Figure 2

*Overall STAMP Scores for Spring 2019 to Spring 2021*

As shown in Figure 2, we can observe a major shift in the distribution of scores. In comparing spring 2019 to spring 2021 semester STAMP scores, we found a significant shift to the right in the distribution for proficiency scores on the STAMP test in the spring 2021 semester. The statistically higher overall STAMP score results were obtained from students after having offered Spanish 1 and 2 fully online for one and a half semesters prior.

**Sub-Skill Student Proficiency Outcomes: Reading, Writing, Speaking, and Listening**

In order to answer Research Question 2 (was there a significant difference in proficiency skill (reading, writing, listening, and speaking) scores between on-ground and online students in respective spring 2019 and 2021 semesters?), we conducted a non-parametric, independent samples Mann-Whitney U test to compare the ranked differences between Spring 2019 (pre) and Spring 2021 (post) for each STAMP sub-skill (reading, writing, speaking, and listening) proficiency in a third-semester Spanish class.

From the descriptive statistics shown in Table 2, we can observe that the mean rank scores within each sub-skill category is visibly higher for the students from the Spring 2021 semester. On average, there is also more variability within the ranked scores from each sub-skill category in the Spring 2021 semester. Of note is the smaller sample size from the speaking sub-skill category due to the fact that students took this task last, and many of them likely ran out of time to complete the entire portion of the speaking task.
Additionally, sometimes students received a grade of NR – not ratable – for the speaking section because the sound recording was not clear enough to rate, either due to a faulty microphone or a soft voice.

**Table 2**

*Descriptive Statistics for Sub-Skill STAMP Results for Spring 2019 (Pre-) to Spring 2021 (Post-)*

<table>
<thead>
<tr>
<th>Sub-Skill</th>
<th>Pre N</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAMP</td>
<td>269</td>
<td>13.33</td>
<td>13</td>
<td>3.38</td>
</tr>
<tr>
<td></td>
<td>299</td>
<td>17.13</td>
<td>17</td>
<td>4.69</td>
</tr>
<tr>
<td></td>
<td>269</td>
<td>4.20</td>
<td>4</td>
<td>1.35</td>
</tr>
<tr>
<td>Reading</td>
<td>299</td>
<td>5.79</td>
<td>6</td>
<td>1.42</td>
</tr>
<tr>
<td></td>
<td>299</td>
<td>3.68</td>
<td>4</td>
<td>0.79</td>
</tr>
<tr>
<td>Writing</td>
<td>293</td>
<td>3.71</td>
<td>4</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>269</td>
<td>2.80</td>
<td>3</td>
<td>0.98</td>
</tr>
<tr>
<td>Listening</td>
<td>294</td>
<td>4.80</td>
<td>5</td>
<td>1.61</td>
</tr>
<tr>
<td></td>
<td>241</td>
<td>2.97</td>
<td>3</td>
<td>0.93</td>
</tr>
<tr>
<td>Speaking</td>
<td>272</td>
<td>3.28</td>
<td>3</td>
<td>1.29</td>
</tr>
</tbody>
</table>

As shown in *Figure 3*, 50.17% of students in the post-reading skill category received a score of "6," while the rest of the students' scores were uniformly distributed between a score of "2" and "9." While the students' scores from the pre-reading skill category were highly concentrated between a "3" and a "5." Both listening and writing scores were highly negatively skewed in the 2019 group, revealing that these sub-skill categories were the most challenging for students. Students' post-listening and writing scores appear to be approximately normally distributed between the scores. In terms of students’ post-speaking skill scores compared to their pre- scores, we observed a slight shift to higher scores, indicating a better performance.

*Figure 3*

*Reading, Writing, Listening, and Speaking STAMP Results for Spring 2019 to Spring 2021*
Since we have non-parametric ordinal data, we used the Mann-Whitney U test instead of the standard parametric $t$-test. As shown in the inferential statistics in Table 3, a Mann-Whitney U test revealed a statistically significant difference in the sub-skill STAMP Spanish proficiency scores between students in the spring 2019 and 2021 semesters across all areas except for writing, which revealed no significant difference in the two distributions of pre- and post-scores. Reading and listening had a large effect size ($r > 0.5$) and the speaking and writing had a small effect size ($r < 0.3$), thus the overall effect size was medium ($0.3 < r < 0.5$).

**Table 3**

**Mann-Whitney Test Results for Sub-Skill STAMP Results for Spring 2019 to Spring 2021**

<table>
<thead>
<tr>
<th>STAMP Test</th>
<th>Test Statistic</th>
<th>$P$-value</th>
<th>Effect Size ($r$)</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>$W = 19969$</td>
<td>$&lt; .001$</td>
<td>0.4361</td>
<td>Reject $H_0$</td>
</tr>
<tr>
<td>Reading</td>
<td>$W = 17184$</td>
<td>$&lt; .001$</td>
<td>0.5059</td>
<td>Reject $H_0$</td>
</tr>
<tr>
<td>Writing</td>
<td>$W = 38482$</td>
<td>0.6031</td>
<td>0.0219</td>
<td>Fail to Reject $H_0$</td>
</tr>
<tr>
<td>Listening</td>
<td>$W = 10937$</td>
<td>$&lt; .001$</td>
<td>0.6365</td>
<td>Reject $H_0$</td>
</tr>
<tr>
<td>Speaking</td>
<td>$W = 27872$</td>
<td>0.002</td>
<td>0.1281</td>
<td>Reject $H_0$</td>
</tr>
</tbody>
</table>

To summarize, when comparing spring 2019 to spring 2021 semesters, we found that when students took the STAMP test during their final exam period after in the spring of 2021, having taken their third-semester Spanish course in a fully online modality, participants performed significantly higher overall on the STAMP test, as well as across the sub-skills of reading, listening and speaking. We also found no significant difference in students' median STAMP writing performance across semesters and course modalities. Despite a potentially detrimental impact of the COVID-19 pandemic on student performance, with the abrupt move to online instruction, students' median STAMP scores in spring of 2021 were significantly higher in three out of the four skills as well as overall. We will now delve deeper into these results, discussing their potential implications.

**Discussion**

This study has revealed important increases in student Spanish proficiency with the move from a solely on-ground to a fully online course modality as a result of the 2020 pandemic. The results of the pre- vs. post-pandemic comparison of student proficiency outcomes as measured by the STAMP test showed that students’ overall language either remained consistent or improved significantly in the online course modality, which is in line with other research (Aldrich & Moneypenny, 2019; Blake, et al., 2008; Moneypenny & Aldrich, 2016). In addition, their reading, listening, and speaking proficiency scores
significantly improved, and their writing proficiency remained constant. Given the challenges of the pandemic crisis context and the close pedagogical equivalence between on-ground and online courses, it may seem surprising that students' overall proficiency scores and sub-proficiencies for three out of the four skills significantly improved.

Certain features of the online modality may have played a greater role in the development of student proficiencies, thus outweighing the negative aspects, such as a lack of technology resources (e.g., inadequate internet connections that precluded camera use, other less-than-optimal contexts for learning) experienced early on in the pandemic. One possible advantage of the online classes was the fact that the Zoom modality may have fostered opportunities for other forms of rich interaction, a key ingredient in second language acquisition (SLA) (Long, 1996; Swain, 2005). Guillén et al. (2020), for instance, found that despite the loss of body language and other embodied communicative techniques that teachers use in on-ground classes, online classes in Zoom during the COVID-19 crisis held the potential to foster other types of diverse interactions, including multimodal collaboration (e.g., sharing of authentic media, breakout rooms, polls, etc.) which may have helped bolster language proficiency over traditional on-ground classroom collaboration in our context (Kosmas, 2021; Nathan, 2022).

Another fundamental pedagogical difference between the on-ground and online courses was the move from a physical, 50-minute language lab experience to a fully online lab experience wherein students would perform communicative types of activities as provided in the Supersite. The 50-minute physical lab experience, at its best, may have proven inferior for proficiency development to the types of communicative tasks offered in the Supersite, such as synchronous partner chats (opportunities for spontaneous interaction), interpretive tasks with access to authentic or semi-authentic videos, simulated conversation practice with an avatar, and more (Tao & Gao, 2022). An on-ground lab with older computers and a firewall that often may have made startup slow, the physical barriers of the space itself, and other logistics could have been a reason for the uptake in proficiency outcomes with the fully online course modality.

It is also likely that, over time, instructors and students improved their ability to teach and learn in an online modality, as the university adopted policies and allocated resources to aid in the transition post-spring 2020. In fact, many instructors who had initially suffered a lack of preparedness and/or experience with online teaching in the crisis context (Gacs et al., 2020), were likely able to secure adequate professional development and practical experience to support online language course delivery by the spring of 2021.

Related logistical matters may be another plausible explanation for the increase in student Spanish language proficiency and skill-based proficiencies among students in online courses in comparison to their peers in on-ground courses (Gacs et al., 2020). Students who took the STAMP test in the spring of 2021, and who scored significantly higher than their peers who had taken it prior to the COVID-19 pandemic, had taken at least one language course online, and likely took as many as three online language courses either in high school or at the university, which may have affected their score. Alternatively, students who took STAMP in the spring semester of 2019 had no opportunity to take any language course online, either at the university or in high school. In other words, and as previously mentioned, the students who took Spanish in the spring of 2019 and scored significantly lower overall on the STAMP had not been afforded the option to take any online general Spanish courses (although they likely had taken technology-enhanced courses in other disciplines), and thus their scores reflect their Spanish language proficiency after having solely taken the fully on-ground Spanish courses offered by the university and in feeder high schools prior to the pandemic.

Students and instructors likely had additional time to prepare for their online classes due to the fact that they did not necessarily have to commute, get dressed, get themselves mentally prepared to move out into the world, etc. (Moser et al., 2021). This additional time could have been invested in their teaching and learning experience. This becomes more significant when one considers the student population in the context of the given study (e.g., a large percentage of first-generation university learners; many living with families, commuting great distances – an hour or more – to attend on-ground classes, and maintaining jobs outside of their role as students to make ends meet). Despite the many challenges of emergency remote learning – especially at the beginning of the pandemic – the added benefit of not having to commute to
language classes, a more flexible schedule to take care of other responsibilities, and more time to devote to their language learning may help to explain the increase in student language proficiency.

A generational divide due to the rapid introduction of digital technologies in the latter decades of the 20th century ought also to be considered (Prensky, 2001). If so-called digital natives preferred the online language learning experience, their motivational response may have fostered the enhancement of their language learning results. This potential preference could help to explain why overall student proficiency and three out of the four sub-skill proficiencies increased in the online course modality in spite of all the setbacks and challenges of the pandemic. Furthermore, it is possible that students in the online courses had grown accustomed to learning language remotely after doing so for two-and-a-half semesters. Thus, this population of students was likely more comfortable producing language in an online setting and hence better prepared to take the online STAMP test.

Digital native preferences for the online course modality have been observed in other contexts (Goertler et al., 2012; Whitford, 2021). Although beyond the scope of the present study, it is important to note that while the university in question experienced an overall decline in enrollment between the spring 2019 and spring 2021 semesters, the third-semester Spanish courses under observation actually increased.6 This increase represented a general trend experienced at this university not only in third-semester Spanish courses, but among the courses in the lower-division language programs across five of the languages offered (Tu et al., 2022). Thus, a plausible rationale for an increase in lower-division course enrollment during the semesters when the university experienced an overall decrease in enrollment is that students wanted to take the required language courses in an online modality while they were being offered virtually.7 Adding to this likelihood is the data garnered from the fall 2021 semester, during which time course sections were offered in both online and on-ground modalities, which show that not only do students tend to select online sections of their language courses first, but also that the online sections tend to be fuller than their on-ground counterparts (Gleason et al., 2022, Gleason & Bartlett, 2022).

These observations about why students' overall language proficiency and sub-level proficiencies in three out of the four skills significantly improved lead to additional questions. For example, many would argue that it is our responsibility as universities and educators to adapt and cater to the changing needs of our learners, especially in light of the fact that these changes have been ongoing for decades having been brought about by collective changes in society at large (Prensky, 2001). A large body of educational research exists on the importance of instructional differentiation for world language and other educational contexts (Blaz, 2016; Mastropieri & Scruggs, 2018; Subban, 2006) as well as for culturally and linguistically-responsive instruction (Hammond, 2014). It could be argued that teenage and 20-to-30-year-old digital native students belong to a different culture than their 40+ counterparts, and thus, their learning needs are different from these latter "digital immigrants" (e.g., most university administrators, instructors, adult learners, etc.).

Additionally, there is something to be said about the notion that "[t]echnology will not replace teachers, but teachers who do not use technology will be replaced" (Clifford, 1983, as cited in Patnoudes, 2014). The COVID-19 pandemic certainly underscored the importance and extended the significance of this passage, since most instructors who may have been reluctant or resistant to using technology in the spring of 2020 no longer had a choice whether to use it or not. In other cases, those instructors who were particularly reluctant or unable to adapt their teaching practice to an online context may have chosen to retire from the profession entirely (Dickler, 2021), and those who instead persisted may have been rewarded for their ability to meet the challenges posed by a new teaching and learning modality (Tao & Gao, 2022).

Concerning the validity of the claims made regarding the results of this study, an argument-based approach to validity (Chapelle & Voss, 2021; Gleason, 2013), which entails including multiple sources of evidence, can help us support or refute a potential claim regarding the superiority of online instruction. We wish to emphasize that while our findings reveal that the students in our context demonstrated significantly higher skills-based proficiencies in reading, listening, and speaking, as well as overall, we cannot make the claim that the online course modality is superior to an on-ground one. Nonetheless, such a claim could be implied...
and thus, we would like to assert that the validity of such a claim would need to be bolstered by additional pieces of evidence, including student course grades, university enrollment and retention rates, student satisfaction surveys, interviews with multiple stakeholders, and more.

Many questions remain unanswered, such as whether or not there is anything worth preserving from an on-ground classroom. In other words, if students garner increased language proficiency and prefer the online course modality, why not do away with on-ground learning altogether? While a full discussion of this is beyond the scope of the present paper, we wonder if prioritizing language proficiency as a field (e.g., ACTFL, 2012) at all jeopardizes other potentially beneficial aspects of taking on-ground language courses that could be of equal if not greater importance. Such underrepresented or otherwise ignored aspects might include: (a) the joy, novelty, and anxiety of connecting and conversing with one’s peers in person; (b) the embodied, multi-sensorial, and subtly nuanced experiences that are part and parcel of what it means to be human; (c) the in-person undergraduate world language learning experience of learning to communicate in a new language, and (d) the evolving meaning of the construct of SLA interaction in the context of online instruction (Chapelle, 2009).

What is more, we might also examine the potential disadvantages of an online-educational experience as promoting a constantly-connected, computer-mediated lifestyle. Consequences of such a choice might include an increased lack of privacy that the online modality engenders (Zuboff, 2019), a difficulty in establishing work-life balance brought about by a perceived need to be online and available 24 hours a day, seven days a week (Duxbury & Smart, 2011), a potential tendency to over-work (Moss, 2021), and a manic energy that one sometimes adopts when constantly checking one’s digital devices and platforms (e.g., cell phone, tablet, laptop, and social media outlets) for notifications and updates (Schmitt et al., 2021). It will be important to consider such questions of ethical importance regarding the future of online teaching and learning so as to become aware of the potential risks of over-eager technological adoption in various educational contexts.

**Conclusion**

The conversation regarding the successes and challenges of online pandemic language instruction will undoubtedly continue for years to come. Within the context of third-semester Spanish courses in a pandemic context, we have revealed some perhaps unexpected results, which will certainly have implications for well-planned online and on-ground courses moving forward. One shortcoming of our study is that because we only focused on changes in language proficiency, we were unable to ascertain why students improved in proficiency overall, in three out of four skills, as well as remained constant in writing proficiency. Future research will undoubtedly be needed, both more broadly (e.g., by comparing results across languages) as well as more granular (e.g., by specifically examining aspects of the online course modality that bolster proficiency). Many of the most important questions left in the wake of the COVID-19 pandemic remain unanswered, as we consider the importance of language proficiency outcomes in lower-division language courses.

**Acknowledgements**

The authors would like to acknowledge three people who were instrumental to this project: Dr. Elu Tu, the former lab director at the university where these data were collected, who helped proctor the STAMP and collect data for this study; Dr. Nick Gossett from Avant Assessment, who answered numerous questions and helped obtain our data sets; and Dr. Victor Santos, also from Avant Assessment, who graciously answered all of our questions regarding validity and reliability.

**Notes**

1. While an ACTFL language proficiency level of Intermediate-Low is the general education requirement
for students of Indo-European languages, students of non-Indo-European languages must obtain a level of Novice-High. Additionally, students arriving at this university with 30 or more initial transfer credits were eligible for a language requirement waiver.

2. The STAMP test had been available for students to take remotely by the Spring 2020 semester; however, the higher cost of proctoring the exam online and the difficulty of administering it remotely to hundreds of students off campus prompted the Department to wait until Spring 2021 to reintroduce STAMP test as the final exam as it had prior to the pandemic. Remote STAMP proctoring thus resumed in the spring of 2021.

3. Participation counted for 20% of the course grade, Online Homework for 15%, Lab Work for 10%, three Integrated Performance Assessments (IPAs), each worth 10% (30% total), and the STAMP test, which was administered as a final exam, and was worth 10%.

4. The Supersite conveniently allows instructors to see an estimated time that students will take to complete each of the assigned activities. Instructors were granted the flexibility and freedom to decide which activities to assign, but consistently reminded of the learning outcomes of the course, which were based on a communicative curriculum.

5. Specifically, incoming students at this institution are generally advised not to take a language in their first fall semester, but rather to wait until the spring semester after taking their placement exam during their first-semester seminar.

6. The university declined in enrollment by 10% (9,485 students to 8,648 students respectively) between the spring 2019 and spring 2021 semesters, while the third-semester Spanish course increased in enrollment by 9% (269 and 299 respectively).

7. Recall that all lower-division language courses offered prior to the COVID-19 pandemic at this university were on-ground. Further, the departmental policy, which has since changed, was not to accept transfer credit for any lower-division language courses that students may have taken at other institutions.

References


Tu, I., Gleason, J., Cardone, R., & Bartlett, A. (2022, June 2). *To teach online or not to teach online: A comparative analysis of proficiency outcomes pre-pandemic and post-2020 across five languages* [Conference presentation]. Computer Assisted Language Instruction Consortium (CALICO). Seattle, WA, United States.


About the Authors

Jesse Gleason, Ph.D. is Associate Professor of Spanish and Applied Linguistics at Southern Connecticut State University in New Haven, CT, where she presently coordinates the Lower-Division Spanish program. Her teaching and research interests include world language education, pedagogy and curriculum, technology-enhanced language learning, and assessment. Jesse Gleason is the corresponding author.

E-mail: gleasonj8@southernct.edu

ORCiD: https://orcid.org/0000-0001-7980-4500

Resha Cardone, Ph.D. is Associate Dean of the College of Liberal and Creative Arts at San Francisco State University. Before becoming an administrator, she was Chairperson of the Department of World Languages and Literatures at Southern Connecticut State University and Professor of Spanish and Latin American literary and cultural studies.

E-mail: cardoner1@sfsu.edu

ORCiD: https://orcid.org/0009-0002-6382-5408

Andy Bartlett, Ph.D. is Assistant Professor of Statistics and Mathematics at Southern Connecticut State University in New Haven, CT, where he is presently the Data Science Coordinator. His teaching interests include Bayesian Statistics, Non-Parametric Statistics, and Data Science models, while his research interests include Extreme Value Theory and Time Series.

E-mail: bartletta4@southernct.edu