This is the accepted version. The final published version is available here:

https://journals.sagepub.com/doi/10.1177/13621688221140106

Longitudinal study of Spanish vowel acquisition by Australian students

Alfredo Herrero de Haro

University of Wollongong, Australia

Manuel Delicado Cantero

Australian National University, Australia

Abstract

A growing body of research, especially on L2 English, has shown the positive effects of explicit pronunciation teaching. However, some beliefs which prevent explicit pronunciation teaching still remain, notably, the belief that regular speaking during class time is enough to improve pronunciation outcomes. This paper analyses the evolution of L2 Spanish vowels in four students at an Australian university. An analysis of 1387 vowels from the first, third and sixth semesters of a Spanish major with no particular focus on explicit pronunciation teaching shows minimal change in the quality of the students' vowels, indicating very little improvement in pronunciation across their six-semester language major. The results suggest that speaking during class is not enough to improve L2 Spanish pronunciation and support explicit pronunciation teaching in the language classroom.

Key words: Spanish as a foreign language, vowel acquisition, language learning, language teaching, pronunciation teaching

1. Introduction

After years reporting the neglected state of L2 pronunciation teaching and learning in the classroom (e.g. MacDonald, 2002; Derwing & Munro, 2005; Usó Viciedo, 2009; Celce-Murcia et al., 2010; Delicado Cantero et al., 2019), we now have extensive research showing that adequate explicit pronunciation teaching benefits students (e.g. Levis, 2005; Mellado, 2012; Kissling, 2013; Nagle & Huensch, 2020). However, most of this research has focused on L2 English (e.g. Thomson & Derwing, 2015).

In the case of L2 Spanish, the literature shows that certain obstacles are still present (Delicado Cantero et al., 2019). This is partly justified by the strong correspondence between Spanish spelling and pronunciation and the related belief that Spanish pronunciation will be easily picked up by default through regular, unassisted practice in a communicative approach (e.g. see discussions in Poch Olivé, 2004; Usó Viciedo, 2009; Orta Gracia, 2009; Delicado Cantero & Steed, 2015). This may translate into course outlines that aim to assess pronunciation but do not include specific pronunciation outcomes (e.g. Steed & Delicado Cantero, 2018).

We do in fact know that research has shown that some pronunciation gains may indeed take place naturally (without phonetic instruction). Thomson & Derwing (2015, p. 339) indicate that naturalistic gains may happen at the beginning of the learning process, in line with previous work by Flege (1988). However, these gains are minimal, as some of the previous papers reported, and, importantly, comprehensibility may nevertheless decline without explicit training. Sturm (2019) found that L2 French students in a US university showed improvement throughout several semesters without explicit pronunciation instruction but that improvement was inconsistent (e.g. sounds pronounced correctly on one occasion could be mispronounced at a later time). Derwing & Munro

(2014, p. 51), writing on English as L2, summarise the importance of explicit pronunciation teaching very clearly: "[a]n explicit focus on pronunciation in language classes, based on intelligibility priorities during that first year, may help learners to become sufficiently comprehensible that intervention for fossilised patterns several years later may not be necessary" (cf. also Couper, 2006; Munro & Derwing, 2008). Notice that foreign accentedness or nativism are not the focus anymore.

2. L2 Spanish vowel and our study

2.1. Research on L2 Spanish vowel

The present study focuses on vowels as these pose a problem to intelligibility and comprehensibility in L2 Spanish (e.g. Gil Fernández, 2007; Stevens 2011; Martínez Celdrán & Elvira-García, 2019). As shown in Herrero de Haro & Andión Herrero (2011, 2012a, 2012b) and Herrero de Haro (2016), transfer from some varieties of English to Spanish can cause intelligibility problems. For example, in Spanish, /o/ tends to mark masculine gender and /a/ feminine; vowels are also key in verbal morphology.

Previous literature has focused on the learning of Spanish vowels, mostly in US tertiary settings and with varying conditions and goals. Menke & Face (2010) study advanced learners of Spanish in the US using two tasks, a background questionnaire and an oral reading of a Spanish short story. The five Spanish vowels were analysed. Their results show that the more advanced students in the sample produced vowels that were similar to those of the native control group, with /a/ being the most problematic vowel. However, despite their improvements, the authors remark that L2 speakers continue to "retain a mark of English in their Spanish pronunciation" (p. 199), showing that foreign accentedness was an important factor.

Stevens (2011) explores the changes in Spanish vowels by L1 US English students enrolled in a third-semester Spanish class during study abroad. The author's goal was to explore whether those students "make more gains in acquiring a native-like pronunciation of Spanish in terms of reduced vowel length" (p. 80) compared to other students not on study abroad. Stevens (2011) shows that the study abroad students indeed improved their vowel duration compared to the control group. However, the author still states that the duration of the vowels in the study abroad students continued to be longer than that of the natives.

Cobb & Simonet (2015) focus on vowel reduction interference from English. They obtained data from volunteers at the University of Arizona, organised in three groups: native speakers of Spanish (control), native speakers of English who are proficient late L2 speakers of Spanish and a final group of English native speakers who are intermediate late L2 speakers of Spanish. This study explored vowels in stressed and unstressed positions with acoustic analysis. Their results show that there were no major differences for /i, a, o/, whereas there were larger differences with /e/ and /u/, especially /u/, which is more fronted in intermediate L2 Spanish speakers than in natives or advanced learners. Advanced learners produced more "native-like" /e/ and /u/ than the intermediate learners.

Solon et al. (2017) explore Spanish vowels in order to ascertain whether task difficulty aligns with improvement in L2 pronunciation. For this study, the authors obtained data from third-year Spanish university students in the US Midwest, with two native Spanish speakers as control. The students had to complete two tasks, a simpler one and a more complex one. The target phonetic context was stressed vowels between stop consonants or between a stop and a pause. Their results show that there was no statistical

significance for F1 with regards to task order (simple or complex first) or task complexity for any vowel. The only significant difference was in F2 differences: /e/ displays a significantly higher F2 during the simple task compared to the complex task. A comparison with the native informants shows that the /e/ of the students was moving towards the native goal in the complex task.

More recently, Long et al. (2018) studied the development of Spanish vowels during a short-term study abroad, as Stevens (2011). Long et al. (2018) obtained data from intermediate level students, focusing on changes in vocalic quality, diphthongisation, and duration over five weeks abroad. The control group continued to study in the US institution. The study used a reading task which was recorded twice: before the study abroad and in the host country. This study also shows minimal gains: no changes in quality, no changes in diphthongisation, and some changes in duration but in the opposite direction in that the study abroad students showed less target-like duration than the control group.

In sum, the previous literature on the development and improvement of Spanish L2 vowels points to little or simply no improvement in general, regardless of contexts analysed or educational settings. Some of the reported gains are very localised (e.g. the F2 of one vowel), are reported in advanced students mostly and, even then, are qualified as not being native-like enough or as still foreign accented. Furthermore, they are all US based.

2.2. The current study

The current paper's rationale sits on some persistent issues in L2 Spanish teaching and learning. First, we aim to test the traditional belief that pronunciation will be learned

by speaking without any further specific instruction (see Steed & Delicado Cantero, 2014 for discussion). Second, the literature reviewed above provides no longitudinal data in spite of the crucial value of such type of study (Ortega & Iberri-Shea, 2005; Long et al. 2018). Longitudinal studies are limited (Nagle 2017) and, to our knowledge, most of them focus on L2 English in English-speaking environments (cf. Derwing & Munro, 2014; Kennedy & Foote, 2015; Zielinski & Pryor, 2020). Finally, the work on L2 Spanish in Australia is still incipient, notwithstanding recent research (Delicado Cantero & Steed, 2015). It should be noted that our research does not seek to point out errors in the students' foreign accent or compare their productions with that of a native speaker. Rather, our aim is to study the acquisition process to measure the degree of improvement in vowel production that these students made without any explicit pronunciation teaching or participation in a study abroad programme to understand L2 acquisition better.

This study tracks a group of four students as they complete their six-semester Spanish major at an Australian university. Focusing on a select group of students over the six semesters of Spanish helps avoid potential unsystematic variation in our statistical model, keeping noise in the statistical models to a minimum so that the effects of time on performance are more likely to show (Field 2018, p. 20).

This paper aims to answer the following research questions:

1. What is the improvement in Spanish pronunciation vowels by a group of Australian university students based only on class attendance and target language use without explicit pronunciation training?

2. How do changes in vowel production through the semesters vary between the whole group and each individual student?

3. Methods

3.1. Participants

Our study concentrates on four students of a Spanish major (six semesters of 13 weeks each) at an Australian university. Students received four hours of instruction per week in first year and three hours in second and third year; one of those hours was a lecture with the whole subject cohort focussing mainly on grammar, and the remaining hours were in individual classes. Each class had a maximum of 24 students and all teachers followed a communicative approach incorporating the four language skills in every lesson; however, no explicit pronunciation teaching was conducted.

All participants grew up as monolingual speakers of Australian English, were between 18 and 22 years old at the time of recording, had no prior knowledge of Spanish before coming to university, did not attend any other Spanish lessons outside the ones for their major, did not report any major L2 exposure outside class (e.g. partner or close friend who was a Spanish speaker), and none of them had participated in any Spanish exchange or study abroad during or before their major; these characteristics were chosen as a way of controlling for the amount of exposure to Spanish outside class. Furthermore, the overall performance of these four students in Spanish was similar with the average marks across all their Spanish subjects being 85% for Fem 1, 88% for Fem 2, 74% for Male 1, and 79% for Male 2; their marks in Spanish after Sem 1, Sem 2 and Sem 3 were 93%, 86% and 85% for Fem 1, 94%, 89% and 87% for Fem 2, 75%, 69% and 78% for Male 1, and 92%, 85% and 75% for Male 2. The instructors were a male speaker from southern Spain, a female speaker from northern Spain, an Australian female who had learned Spanish in Spain, an American female who had learned Spanish in Spain, and an

Australian man who had learned Spanish in Peru. All these instructors taught the students at different stages during the majors.

More participants were recorded during their major but only four reliably met the specific criteria above. Some of the reasons for exclusion included speaking an additional language at home or having participated in an exchange abroad. Although comparing the evolution of the four chosen students with students who have taken part in an exchange abroad might seem like a promising route to understand language acquisition better and to test the early fossilisation hypothesis, there are several reasons not to include them in our analysis. It is much more difficult to track the progress of the students while abroad as they would have to be recorded by staff at the host institution. The students would all have to have been studying at the same institution abroad for consistency. Students in overseas programmes may experience very different levels of L2 exposure depending on their social networks abroad (e.g. whether they live with local students or with speakers of their L1) and measuring the level of interaction with the L2 can be problematic and unreliable. As a result, all these variables would add too much noise to the statistical model.

3.2. Data collection procedure

The students in our cohort were recorded after their first, third and sixth semester of Spanish in their Spanish major, which lasts six semesters. The recordings were part of class assessment. Ethics clearance and participants' permission were obtained. In Semesters 1 and 3, the recordings include performing a memorised dialogue in pairs with another student in front of their tutor and answering some questions individually about

topics covered in class. In Semester 6, the recording covers a general individual discussion with the tutor about different topics or class projects.

3.3. Data analysis procedure

3.3.1. Acoustic analysis

The vowels of the students were analysed acoustically to identify variation in L2 Spanish vowel pronunciation through the semesters. The F1 and F2 of each vowel were analysed on Praat (Boersma & Weenink, 2016). Waveform and spectrogram information was used to identify vowels, as is common in the literature (e.g. Hedia & Plag, 2017; Long et al. 2018). Following Torreira (2012), the onset of the vowels was identified by a sudden increase in amplitude in the waveform and by an abrupt energy increase in the F2 and F3; the latter manifests in the spectrogram by much darker formants in the F2 and F3 regions (e.g. Torreira, 2012). The offset of the vowel was marked at the point where the waveform showed an abrupt reduction of energy, which manifests in the waveform by a reduction of amplitude of the wave (e.g. Hedia & Plag, 2017; Long et al. 2018) and when the F2 shows a sudden decrease in energy (Henriksen & Harper, 2016).

Following other studies analysing vowel quality (e.g. Herrero de Haro, 2021), the F1 and F2 were measured at the stable segment in F1 and F2 energy closest to the peak of intensity (Figure 1). Vowel stability is characterised by minimal movement in the F1 and the F2. The section measured for each vowel was in the region of 20 ms. Analysing a section rather than a single point captures any minimal variation within vowels and helps reduce the impact of formant fluctuations or transitions (Warren, 2018). Furthermore, measuring the section closest to the peak of intensity, away from the onset

and offset of the vowels, minimises coarticulatory effects from neighbouring consonants (Herrero de Haro, 2019).

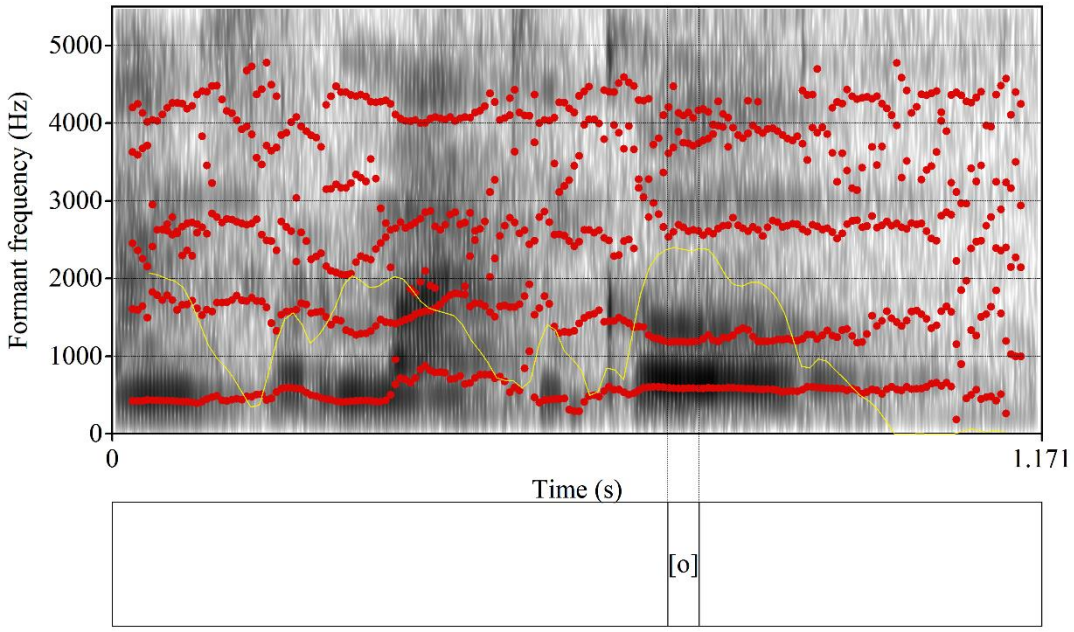

Figure 1. Stable section of word-final /o/ analysed in the word *informático* 'computer, adj.'. Fem 1, Semester 6.

All vowels were analysed word-finally except some cases of /u/ and /i/, since these are uncommon word-finally in Spanish (e.g. Alarcos Llorach, 1976) and not enough samples could be collected for these two vowels in a naturalistic way otherwise. These two vowels were also analysed in open syllables word-internally to increase the number of tokens. Vowels followed by other vowels were only included if they belonged to different words and if there was a visible pause between them in the spectrogram. This was done to minimise coarticulatory effects.

Following Torreira & Ernestus (2011), formant measurements were not normalised; any between-subject variation was accounted for in the statistical model by

coding speaker as a random variable. Normalisation results in data which are not linguistically meaningful and much more difficult to interpret (Thomas & Kendall, 2007).

The number of vowels analysed per speaker and semester are included in Table 1.

Table 1. Number of vowels analysed in the present study.

Speaker	Semester	/i/	/e/	/a/	/o/	/u/	Total
Fem 1	Sem 1	20	20	21	22	19	**102**
	Sem 3	12	25	23	13	8	**81**
	Sem 6	13	35	19	18	30	**115**
	Total	**45**	**80**	**63**	**53**	**57**	**298**
Fem 2	Sem 1	18	27	34	31	11	**121**
	Sem 3	6	29	31	16	6	**88**
	Sem 6	38	31	17	52	16	**154**
	Total	**62**	**87**	**82**	**99**	**33**	**363**
Male 1	Sem 1	12	13	27	21	12	**85**
	Sem 3	13	19	17	31	6	**86**
	Sem 6	22	36	37	35	17	**147**
	Total	**47**	**68**	**81**	**87**	**35**	**318**
Male 2	Sem 1	16	17	24	53	19	**129**
	Sem 3	15	27	37	38	12	**129**
	Sem 6	27	43	21	32	27	**150**
	Total	**58**	**87**	**82**	**123**	**58**	**408**
All speakers	Sem 1	66	77	106	127	61	**437**
	Sem 3	46	100	108	98	32	**384**
	Sem 6	100	145	94	137	90	**566**
	Total	**212**	**322**	**308**	**362**	**183**	**1387**

3.3.2. Statistical analysis

The F1 and F2 were analysed separately using the same statistical models since these two formants are, in fact, different continuous variables and their values are not comparable between them. Analysing F1 and F2 separately is not uncommon in studies analysing vowels (e.g. Clopper et al., 2005; Henriksen, 2017; Long et al., 2018; Herrero de Haro, 2021). A linear mixed model for F1 and another for F2 for each of the five Spanish vowels offers a much more robust statistical model; Long et al. (2018) follow a similar methodology.

The statistical analysis of F1 and F2 values (linear mixed model; see Eddington, 2015, p. 81) focuses on finding out whether any variation in the pronunciation of the

vowels between the semesters is statistically significant. While longitudinal studies analysing the improvement of exam performance may seem easy to interpret based on final marks, the interpretation of vowel variation is more complicated. We cannot conclude, for example, that a value of 1500 for the F2 of /a/ is better than a value of 1300, due to existing dialectal differences in the L1 itself. The four students in our samples were not always taught by the same staff member and were exposed to different accents and thus potentially different F1 and F2 values. We cannot conclude either that any differences are due to unsystematic variation (Field, 2018, p. 19), since the extra hours of class may be enough to expect a change in the pronunciation. More importantly, analysing students' pronunciation against a fixed native model would not measure improvement in intelligibility/comprehensibility, which is our ultimate goal.

All the assumptions for the linear mixed model (e.g. normal distribution of residuals) were tested following Eddington (2015, p. 133). As per Field (2018, p. 232), normal distribution of residuals applies to each level of the predictor, not to the overall data. Therefore, this was tested for the F1 and F2 of each vowel for each of the three semesters. The data for the F1 and F2 of all the vowels met all the assumptions and no transformation of data (e.g. Eddington, 2015, p. 38–40; Field, 2018, p. 268–276) was needed.

Different covariance structures were tested to find the best fitting model to account for the repeated measures (Eddington, 2015, p. 122). The more parsimonious statistical model with the fewest parameters and lowest -2RLL and BIC was selected (Eddington, 2015, p. 117). To control for repeated measures, each speaker was allowed a random intercept since this is necessary when there is more than one observation per subject (Barr et al., 2013, p. 262). *Semester* was coded as the independent categorical

variable, *Formant value* as the dependent continuous variable, and *Speaker* as a random variable. Furthermore, since the independent variable is a categorical value (different semesters), it is possible to consider each semester as a different condition/treatment. *Time* was coded as an independent categorical variable in line with other longitudinal linguistic studies (e.g. Coxhead & Boutorwick, 2018).

A random slope was fitted for each speaker and semester by speaker; this is needed if there is more than one observation per subject and treatment level as a way of controlling for each speakers' idiosyncrasies and to account for the repeated measures (Eddington, 2015, p. 124). Including random effects in our statistical model reduces the residuals and allows us to account for more of the variance (Eddington, 2015, p. 110). The resulting residuals were specified with a Variance Components covariance structure.

Changes in vowel production through the semesters was also analysed per speaker to understand individual progress better. We used ANOVA tests for the F1 and F2 of each vowel for each student.

As all the vowels to be analysed belonged to only one vowel phoneme and only to one speaker; linear mixed models could not be performed as there was no random variable. Repeated measures ANOVA was discarded for several reasons. First, a repeated measures ANOVA requires the same number of repetitions per token per condition (e.g. time) forcing us to disregard valuable tokens. Second, while this kind of test assumes a dependency between results, this is not appropriate for vowels. Thirdly, as the independent variable is not a continuous variable (e.g. number of years) but a categorical value (different semesters), it is possible to consider each semester as a different condition (e.g. Coxhead & Boutorwick, 2018).

All the assumptions of the ANOVA test (e.g. homogeneity of variance) were tested; as per Field (2018, p. 534), normality was tested within each group, not for the entire sample. All the F1 and F2 values for each of our participants met all the assumptions except F1 /o/ of Fem 1; this was analysed from the transformed data *F1 + 1000*. The effect size (partial eta^2) were interpreted as .01 - .06 weak, .06 - .14 medium, and > .14 large, as per Cohen (1988).

4. Results

4.1. Results: acoustic analysis

The F1 and F2 values of all the vowels analysed in the present study were plotted to carry out an exploratory analysis (Figure 2).

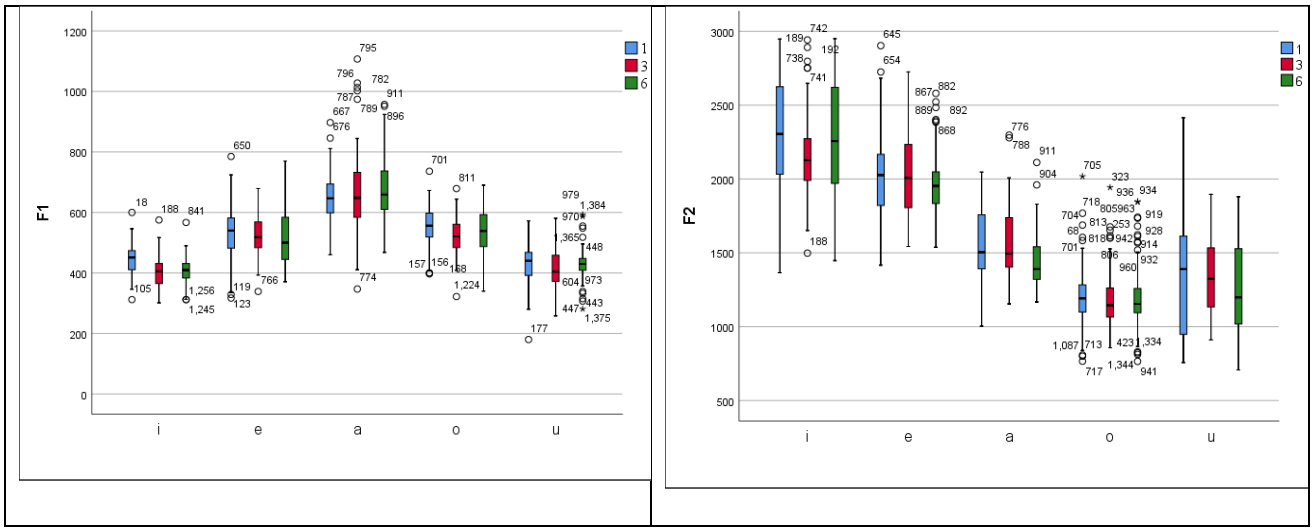

Figure 2. F1 (left) and F2 (right) values of /i/, /e/, /a/, /o/ and /u/ for each semester for the four speakers.

Figure 2 shows that the values for the F1 remain fairly constant across the semesters for the five vowels. Vowels /i/ and /u/ become more stable as the semesters advance as their F1 show a smaller degree of variation. This is not the case for /e/, /a/ and /o/. Regarding F2, the values are also quite similar between the semesters. The variation of the F2 for all

vowels except for /i/ is smaller in Semester 6 than in Semester 1; this could indicate increased stability. Sound instability is typical from L2 speech, due to the changes that take place during the acquisition of a foreign language (Moreno Fernández, 2002). An overlap in values shows no statistical significance (Eddington, 2015, p. 20), so there might be no statistical difference between the F1 and F2 of these vowels throughout the semesters. However, a statistical analysis is necessary to confirm this (Section 4.2). As a preliminary analysis, it is worth plotting the vowel values for each speaker across the semesters (Figures 3, 4, 5 and 6).

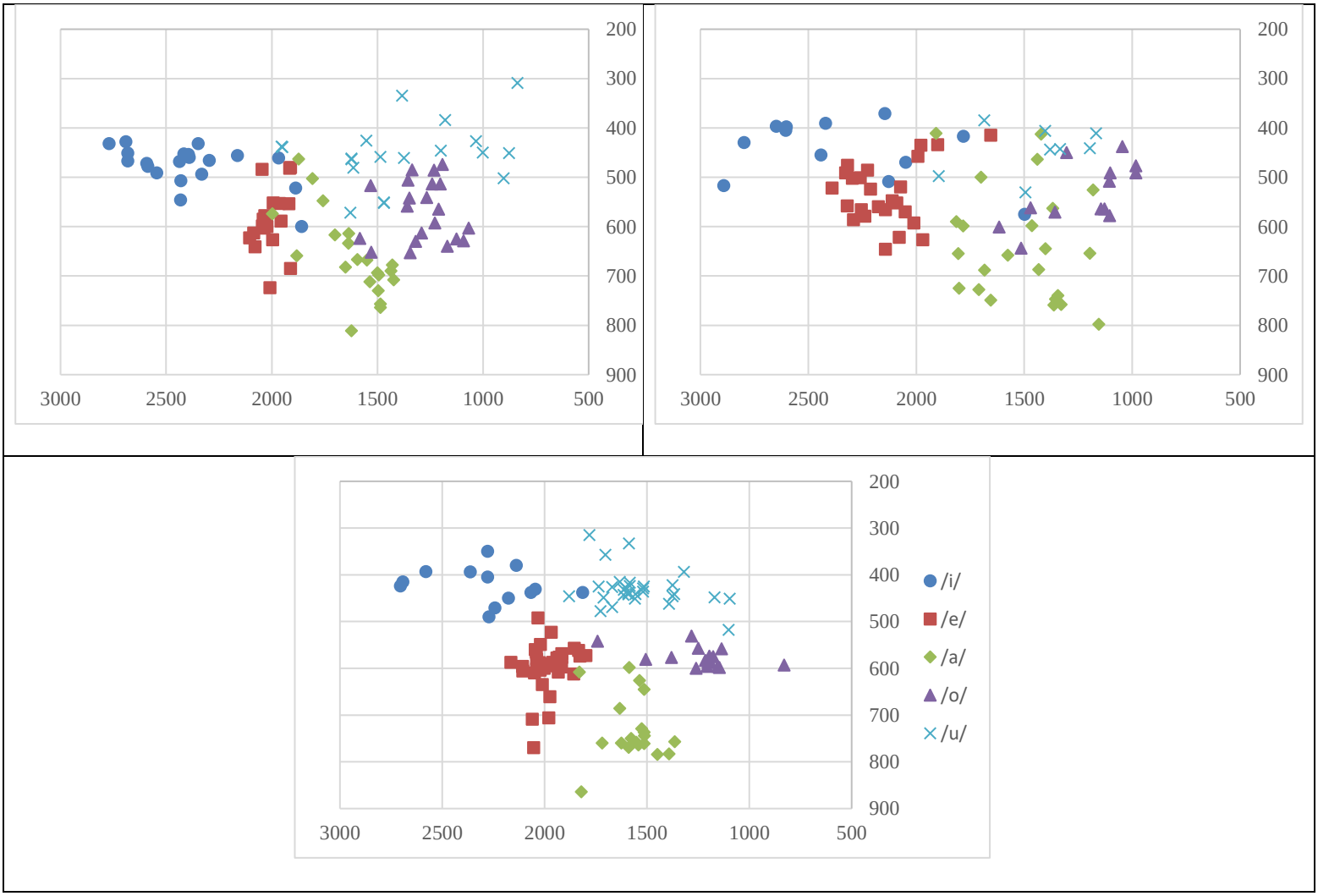

Figure 3. Formant values (Hz) for Fem 1. Semester 1 in top left, Semester 3 in top right and Semester 6 at the bottom.

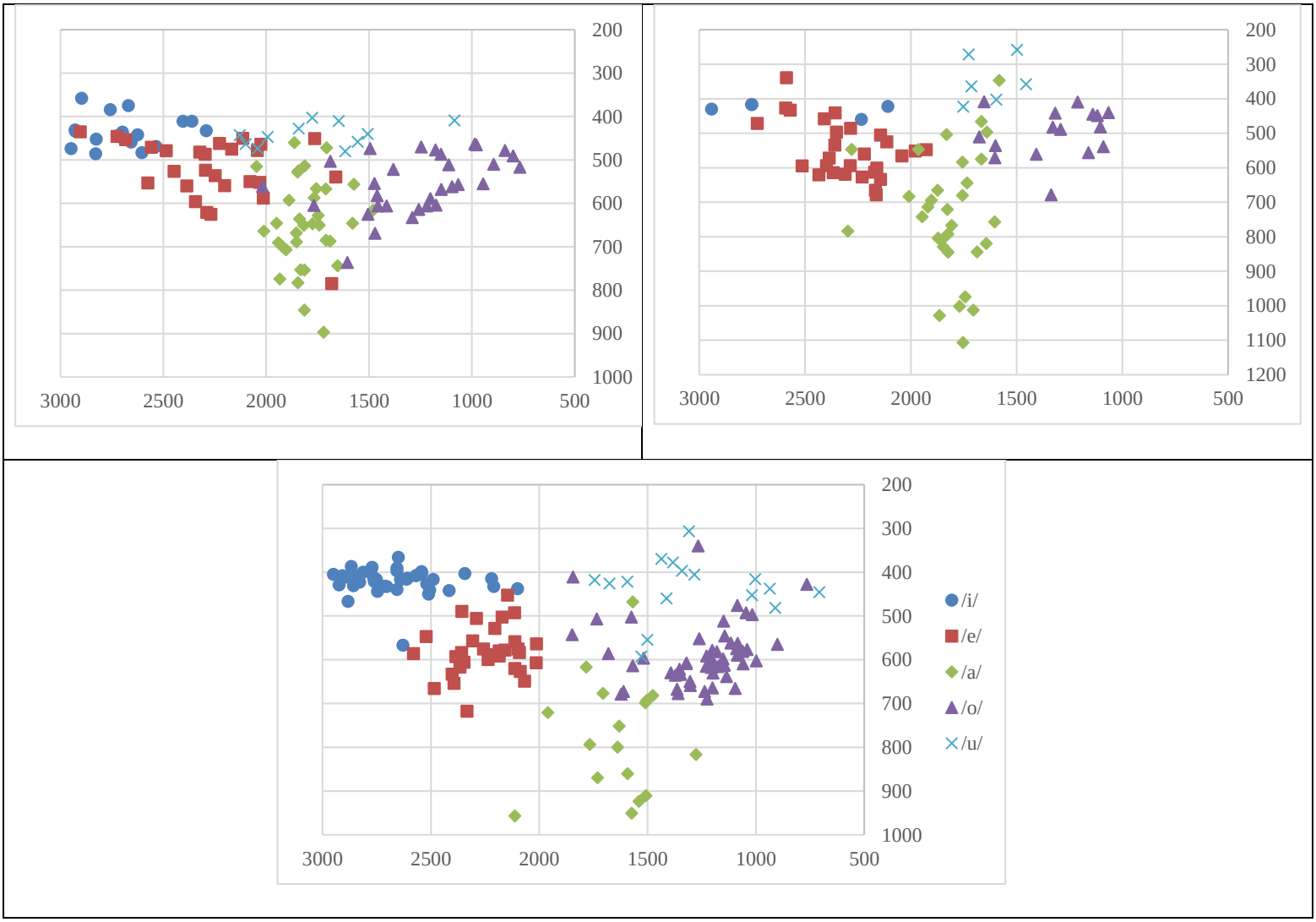

Figure 4. Formant values (Hz) for Fem 2. Semester 1 in top left, Semester 3 in top right and Semester 6 at the bottom.

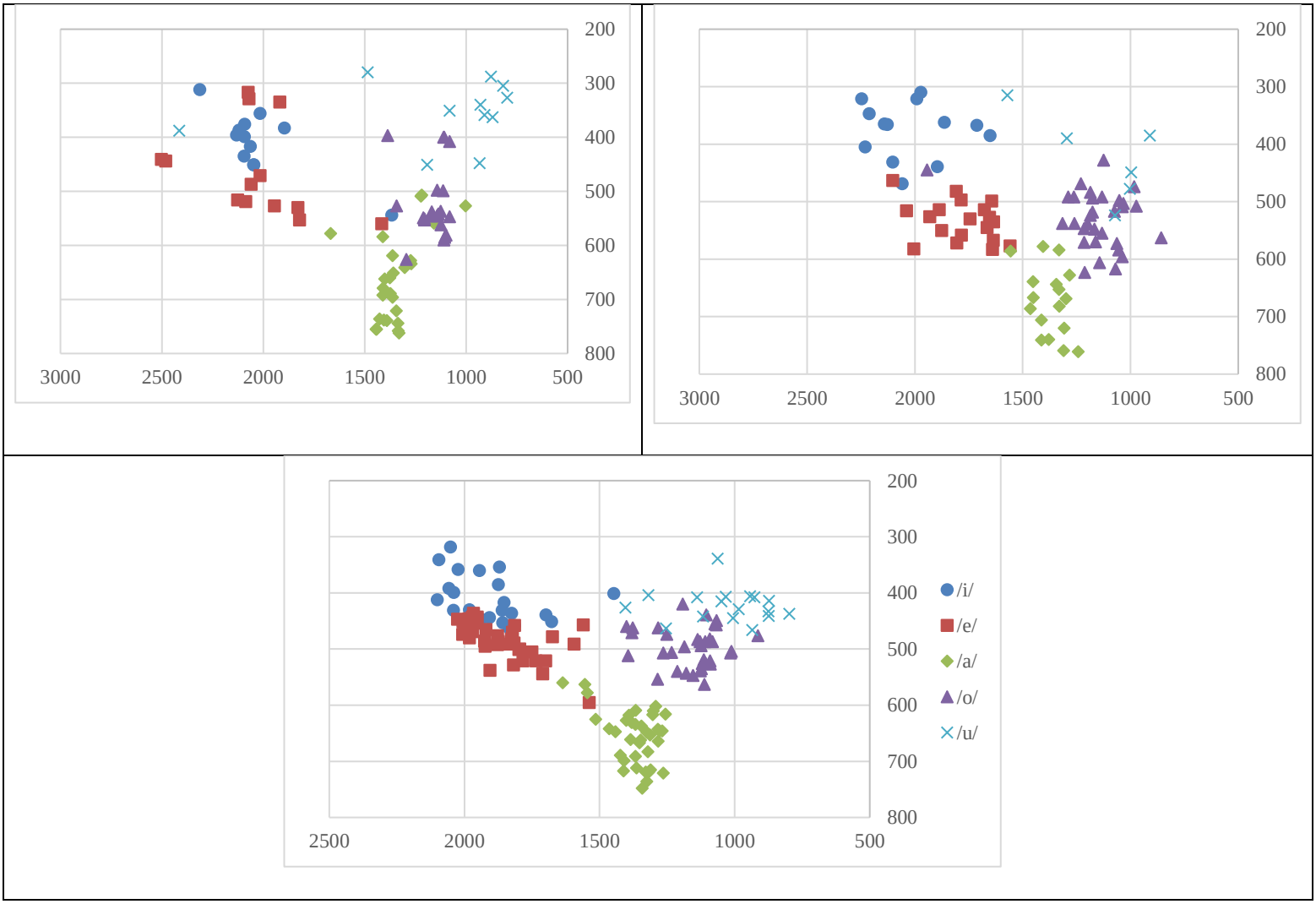

Figure 5. Formant values (Hz) for Male 1. Semester 1 in top left, Semester 3 in top right and Semester 6 at the bottom.

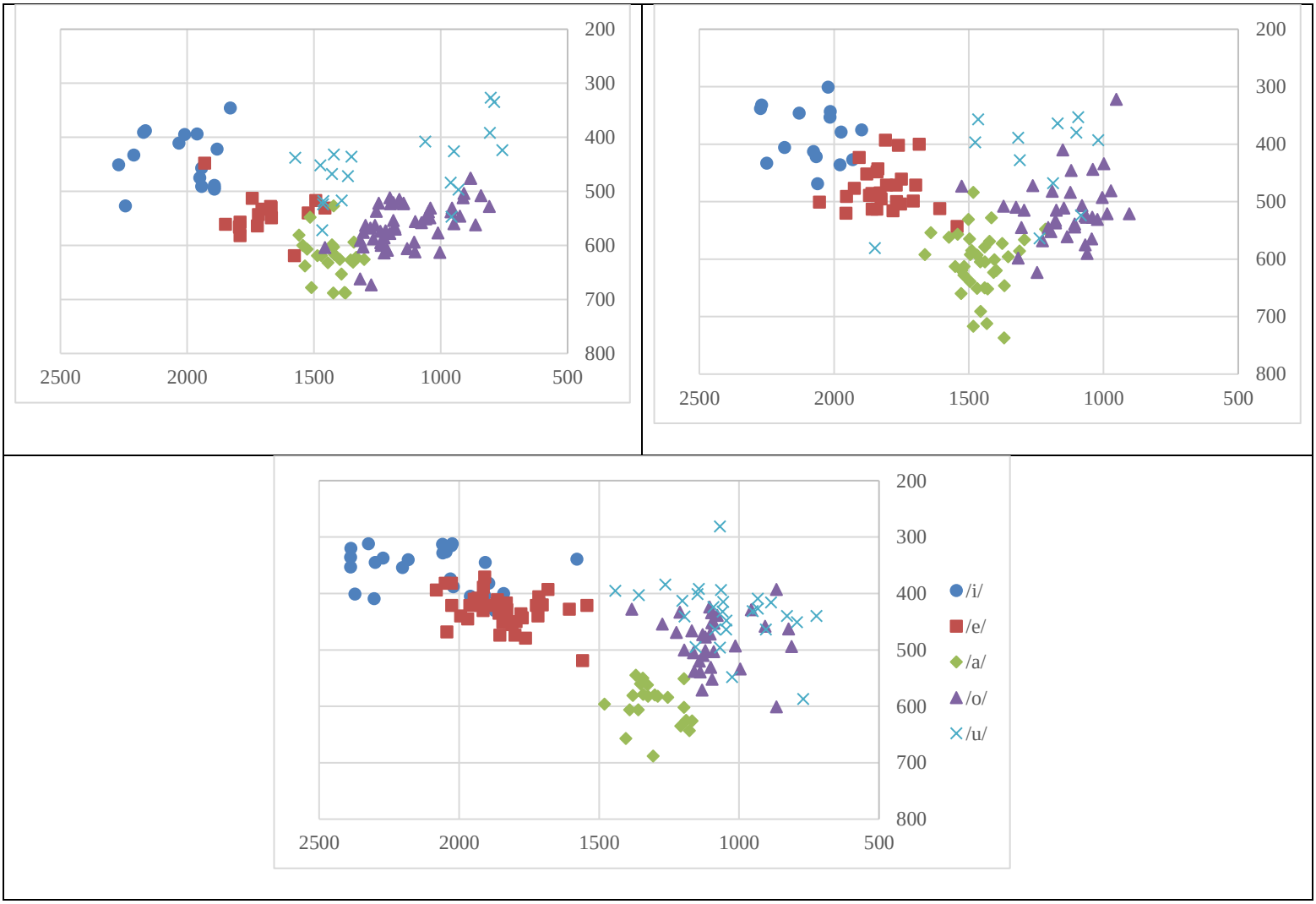

Figure 6. Formant values (Hz) for Male 2. Semester 1 in top left, Semester 3 in top right and Semester 6 at the bottom.

In Semester 1, Fem 1's /i/ and /u/ are much more spread out across the F2 than the other vowels and more so than in the other semesters. Fem 1 has a high degree of overlapping in the mid-central space, the space of an English schwa. In Semesters 3 and 6, the mid-central space is emptier, i.e. less vowel overlap. This indicates that vowels are differentiated better, which research identifies with increased intelligibility (e.g. Kim et al., 2011).

Fem 2 shows a high degree of merging in the F1 of high and mid vowels in Semester 1; this is much lower in Semester 6. Regarding the F2, vowels are more spread out in semester 6, indicating less merging of vowels. Vowel overlap in the mid-central space decreases throughout the semesters. While Fem 2's vowels become more stable as the semesters proceed (Figure 4), some vowel overlap is still present.

Male 1 has much fewer vowels in Semester 1; this unbalance in samples is typical of naturalistic samples (Eddington, 2015, p. 110). We find a high degree of merging between the high and mid, and mid and low vowels, which improves by semester 6. Semester 6 has a more V-like distribution of vowels than Semesters 1 and 3.

The situation is similar for Male 2. In semester 1, we see high vowel overlap and a busy mid-central vowel space, which has improved by semester 6, with less centralisation of Spanish vowels as is typical in English speakers, especially when vowels are unstressed (Herrero de Haro & Andión Herrero, 2011). Semester 6 also shows a much more V-like structure.

At this stage, it is necessary to carry out a statistical analysis to ascertain whether these changes in vowel quality are indeed statistically significant. This is done in Section 4.2.

4.2. Results: statistical analysis

The results for the F1 and F2 of each Spanish vowel for all our speakers are combined in Table 2. It is important to note that effect sizes are only given in the present paper for variables identified as statistically significant by the statistical model. Since none of the tests reported in Table 2 reach statistical significance, no effect sizes are included in this table.

Table 2. Results of the linear mixed models for F1 and F2.

Formant and vowel	Result of the linear mixed model
F1 /i/	F (2, 6.413) = 3.816, p = .081
F2 /i/	F (2, 206.152) = .937, p = .394
F1 /e/	F (2, 5.955) = .027, p = .974
F2 /e/	F (2, 5.542) = .076, p = .927
F1 /a/	F (2, 6.065) = 1.036, p = .410
F2 /a/	F (2, 5.956) = 1.827, p = .241
F1 /o/	F (2, 6.182) = 1.012, p = .417
F2 /o/	F (2, 357.309) = .633, p = .532
F1 /u/	F (2, 5.331) = .200, p = .825
F2 /u/	F (2, 6.781) = 1.245, p = .346

Importantly, the analysis shows no significant differences between F1 or F2 in the three semesters. Figure 2 already indicated minimal variation in the pronunciation of the vowels between each of the three semesters.

The results for the F1 and F2 of each Spanish vowel for each speaker are included in Table 3. To facilitate the interpretation of the data and to aid clarity, effect sizes in Table 3 are only considered for variables identified as statistically significant by the statistical model.

Table 3. ANOVA results for the F1 and F2 of /i/, /e/, /a/, /o/ and /u/ for each speaker through the semesters.

Speaker	Formant	ANOVA results	Effect size as per Cohen (1988)
Fem 1	F1 /i/	F (2, 42) = 5.712, p =.006, η^2 = .214	Large
	F2 /i/	F (2, 42) = .545, p =.584, η^2 = .025	
	F1 /e/	F (2, 77) = 9.528, p =.000, η^2 = .198	Large
	F2 /e/	F (2, 77) = 15.782, p =.000, η^2 = .291	Large
	F1 /a/	F (2, 60) = 5.595, p =.006, η^2 = .157	Large
	F2 /a/	F (2, 60) = .1777, p =.178, η^2 = .056	
	F1 /o/	F (2, 50) = 3.017, p =.058, η^2 = .108 (from *F1+100* transformed data)	
	F2 /o/	F (2, 50) = .738, p =.483, η^2 = .029	
	F1 /u/	F (2, 54) = 1.236, p =.299, η^2 = .044	
	F2 /u/	F (2, 54) = 2.360, p =.104, η^2 = .080	
Fem 2	F1 /i/	F (2, 59) = 1.160, p =.321, η^2 = .0.38	
	F2 /i/	F (2, 59) = .784, p =.461, η^2 = .026	
	F1 /e/	F (2, 84) = 4.698, p =.012, η^2 = .101	Large
	F2 /e/	F (2, 84) = .379, p =.686, η^2 = .009	
	F1 /a/	F (2, 79) = 5.560, p =.006, η^2 = .123	Large
	F2 /a/	F (2, 79) = 8.206, p =.001, η2 = .172	Large

	F1 /o/	F (2, 96) = 9.833, p =.000, η^2 = .170	Large
	F2 /o/	F (2, 96) = .354, p =.703, η^2 = .007	
	F1 /u/	F (2, 30) = 6.228, p =.005, η^2 = .293	Large
	F2 /u/	F (2, 30) = 8.911, p =.001, η^2 = .373	Large
Male 1	F1 /i/	F (2, 44) = 2.286, p =.114, η^2 = .094	
	F2 /i/	F (2, 44) = 2.416, p =.101, η^2 = .099	
	F1 /e/	F (2, 65) = 9.508, p =.000, η^2 = .226	Large
	F2 /e/	F (2, 65) = 7.946, p =.001, η^2 = .196	Large
	F1 /a/	F (2, 78) = .468, p =.628, η^2 = .012	
	F2 /a/	F (2, 78) = .712, p =.494, η^2 = .018	
	F1 /o/	F (2, 84) = 3.791, p =.027, η^2 = .083	Medium
	F2 /o/	F (2, 84) = .012, p =.988, η^2 = .000	
	F1 /u/	F (2, 32) = 8.460, p =.001, η^2 = .346	Large
	F2 /u/	F (2, 32) = .665, p =.521, η^2 = .040	
Male 2	F1 /i/	F (2, 55) = 15.019, p =.000, η^2 = .353	Large
	F2 /i/	F (2, 55) = .499, p =.610, η^2 = .018	
	F1 /e/	F (2, 84) = 75.445, p =.000, η^2 = .642	Large
	F2 /e/	F (2, 84) = 9.432, p =.000, η^2 = .183	Large
	F1 /a/	F (2, 79) = 1.723, p =.185, η^2 = .042	
	F2 /a/	F (2, 79) = 22.844, p =.000, η^2 = .366	Large
	F1 /o/	F (2, 120) = 27.412, p =.000, η^2 = .314	Large
	F2 /o/	F (2, 120) = 1.417, p =.247, η^2 = .023	
	F1 /u/	F (2, 55) = .707, p =.498, η^2 = .025	
	F2 /u/	F (2, 55) = 4.646, p =.014, η^2 = .145	Large

When the ANOVA test identified a statistically significant difference between the values of F1 or F2 across the semesters, a post-hoc test was carried out to locate the diverging semester. A Scheffé post-hoc test was carried out rather than a Tukey post-hoc since the latter assumes the same number of scores across each group (Eddington, 2015, p. 66). In the interest of brevity, the results from the post-hoc tests have been merged in one table per speaker (Tables 4, 5, 6 and 7).

Table 4. *p* value for the differences between F1 and F2 of /i/, /e/, /a/, /o/ and /u/ for Fem 1 across Semester 1, Semester 3 and Semester 6. * indicates statistically significant (*p value* < 0.05).

F1 /i/	**Sem 1**	**Sem 3**	**Sem 6**
Sem 1		.182	.008 *
Sem 3	.182		.475
Sem 6	.008 *	.475	

F1 /e/	**Sem 1**	**Sem 3**	**Sem 6**

F2 /e/	**Sem 1**	**Sem 3**	**Sem 6**

Sem 1		.010 *	.853
Sem 3	.010 *		.000 *
Sem 6	.853	.000 *	
F1 /a/	**Sem 1**	**Sem 3**	**Sem 6**
Sem 1		.707	.065
Sem 3	.707		.008 *
Sem 6	.065	.008 *	

Sem 1		.001 *	.706
Sem 3	.001 *		.000 *
Sem 6	.706	.000 *	

Regarding Fem 1's vowels, the only statistically significant difference for F1 /i/ is between Semesters 1 and 6. F1 /e/ and F2 /e/ change from Semester 1 to Semester 3 and from Semester 3 to Semester 6. F1 /a/ changes only from Semester 3 to Semester 6. This shows that Fem 1 shows a slightly higher change in vowel quality inside the second part of her major (Semester 3 to 6).

Table 5. *p* value for the differences between F1 and F2 of /i/, /e/, /a/, /o/ and /u/ for Fem 2 across Semester 1, Semester 3 and Semester 6. * indicates statistically significant (*p value* < 0.05).

F1 /e/	**Sem 1**	**Sem 3**	**Sem 6**
Sem 1		.354	.012 *
Sem 3	.354		.276
Sem 6	.012 *	.276	
F1 /a/	**Sem 1**	**Sem 3**	**Sem 6**
Sem 1		.053	.013 *
Sem 3	.053		.631
Sem 6	.013 *	.631	
F1 /o/	**Sem 1**	**Sem 3**	**Sem 6**

F2 /a/	**Sem 1**	**Sem 3**	**Sem 6**
Sem 1		.878	.004 *
Sem 3	.878		.001 *
Sem 6	.004 *	.001 *	

Sem 1		.045 *	.125
Sem 3	.045 *		.000 *
Sem 6	.125	.000 *	

F1 /u/	**Sem 1**	**Sem 3**	**Sem 6**
Sem 1		.011 *	.966
Sem 3	.011 *		.011 *
Sem 6	.966	.011 *	

F2 /u/	**Sem 1**	**Sem 3**	**Sem 6**
Sem 1		.669	.001 *
Sem 3	.669		.074
Sem 6	.001 *	.074	

For Fem 2, F1 /e/, F1 /a/ and F2 /u/ change from Semester 1 to Semester 6. F2 /a/, F1 /o/ and F1 /u/ change from Semester 1 to Semester 6 and from Semester 3 to Semester 6. This shows that Fem 2's vowels tend to change equally in the first and second part of the degree.

Table 6. *p* value for the differences between F1 and F2 of /i/, /e/, /a/, /o/ and /u/ for Male 1 across Semester 1, Semester 3 and Semester 6. * indicates statistically significant (*p value* < 0.05).

F1 /e/	**Sem 1**	**Sem 3**	**Sem 6**
Sem 1		.001 *	.368
Sem 3	.001 *		.004 *
Sem 6	.368	.004 *	

F2 /e/	**Sem 1**	**Sem 3**	**Sem 6**
Sem 1		.001 *	.010 *
Sem 3	.001 *		.403
Sem 6	.010 *	.403	

F1 /o/	**Sem 1**	**Sem 3**	**Sem 6**
Sem 1		.894	.184
Sem 3	.894		.037 *
Sem 6	.184	.037 *	

F1 /u/	**Sem 1**	**Sem 3**	**Sem 6**
Sem 1		.021 *	.002 *
Sem 3	.021 *		.999
Sem 6	.002 *	.999	

Regarding Male 1, F1 /e/, F2 /e/ and F1 /u/ change from Semester 1 to Semester 3 and from Semester 3 to Semester 6. F1 /o/ changes only from Semester 3 to Semester 6. For Male 1, his vowels improve slightly more in the first part of the degree than in the second one.

Table 7. *p* value for the differences between F1 and F2 of /i/, /e/, /a/, /o/ and /u/ for Male 2 across Semester 1, Semester 3 and Semester 6. * indicates statistically significant (*p value* < 0.05).

F1 /i/	**Sem 1**	**Sem 3**	**Sem 6**
Sem 1		.004 *	.000 *
Sem 3	.004 *		.346
Sem 6	.000 *	.346	

F1 /e/	**Sem 1**	**Sem 3**	**Sem 6**
Sem 1		.000 *	.000 *
Sem 3	.000 *		.000 *
Sem 6	.000 *	.000 *	

F2 /e/	**Sem 1**	**Sem 3**	**Sem 6**
Sem 1		.016 *	.000 *
Sem 3	.016 *		.400
Sem 6	.000 *	.400	

F2 /a/	**Sem 1**	**Sem 3**	**Sem 6**
Sem 1		.455	.000 *
Sem 3	.455		.000 *
Sem 6	.000 *	.000 *	

F1 /o/	**Sem 1**	**Sem 3**	**Sem 6**
Sem 1		.000 *	.000 *
Sem 3	.000 *		.035 *
Sem 6	.000 *	.035 *	

F2 /u/	**Sem 1**	**Sem 3**	**Sem 6**
Sem 1		.531	.159
Sem 3	.531		.021 *
Sem 6	.159	.021 *	

Regarding Male 2, F1 /i/, F2 /e/ and F2 /a/ change from Semester 1 to Semester 3 and from Semester 1 to Semester 6. F1 /e/ and F1 /o/ change from Semester 1 to Semester 3, from Semester 3 to Semester 6, and from Semester 1 to Semester 6. F2 /u/ changes from Semester 3 to Semester 6. For Male 2, vowels change equally in the first and second parts of the major.

The results are similar for all the participants. If we take into account changes in F1 or F2 between semesters, we have 11 statistically significant differences between Semesters 1 and 3 and 11 cases for Semesters 3 and 6. We can conclude that there is no global difference regarding degree of change between the first and the second parts of their degrees.

4.3. Vowel variability across the semesters for all speakers

Vowel production can also be analysed regarding variability, since L2 production is characterised by a lower degree of stability than L1 vowels (Moreno Fernández, 2002). L2 vowels are more unstable at the beginning of the learning process and vary less as the learner becomes more proficient (Menke & Face, 2010). The standard deviation for the F1 and F2 of each vowel for each semester for the four speakers was also analysed statistically (Table 8).

Table 8. Standard deviation values for F1 and F2 of /i/, /e/, /a/, /o/ and /u/ across the semesters for each participant.

Speaker	Vowel	Formant	Sem 1	Sem 3	Sem 6	Sem 3 – Sem 1	Sem 6 – Sem 3	Sem 6 – Sem 1
Fem 1	/i/	F1	41.58	62.10	37.98	20.52	-24.11	-3.59
		F2	260.99	421.12	257.73	160.13	-163.39	-3.26
	/e/	F1	61.63	61.94	51.04	0.31	-10.89	-10.59
		F2	59.43	166.41	83.73	106.99	-82.69	24.30
	/a/	F1	85.26	113.99	68.13	28.73	-45.86	-17.13
		F2	164.41	225.75	121.52	61.34	-104.23	-42.89
	/o/	F1	61.81	62.16	19.46	0.36	-42.70	-42.35
		F2	135.23	210.15	180.59	74.92	-29.56	45.36

	/u/	F1	65.98	48.56	39.47	-17.43	-9.09	-26.51
		F2	338.06	245.30	190.85	-92.76	-54.46	-147.22
Fem 2	/i/	F1	35.99	16.12	31.38	-19.87	15.26	-4.61
		F2	195.85	325.72	204.79	129.86	-120.93	8.93
	/e/	F1	76.46	81.56	56.27	5.11	-25.29	-20.19
		F2	299.68	193.89	149.46	-105.78	-44.43	-150.21
	/a/	F1	101.75	178.96	131.66	77.21	-47.30	29.92
		F2	122.45	164.43	195.00	41.98	30.57	72.55
	/o/	F1	67.17	71.45	71.60	4.28	0.14	4.43
		F2	294.92	213.03	227.11	-81.89	14.09	-67.81
	/u/	F1	26.50	67.62	68.16	41.13	0.54	41.67
		F2	313.86	126.63	301.18	-187.23	174.54	-12.69
Male 1	/i/	F1	58.18	48.57	41.77	-9.61	-6.81	-16.41
		F2	227.69	191.71	156.10	-35.98	-35.60	-71.59
	/e/	F1	86.20	34.66	33.42	-51.54	-1.25	-52.79
		F2	277.46	155.73	129.28	-121.73	-26.45	-148.18
	/a/	F1	81.60	59.58	47.80	-22.02	-11.78	-33.80
		F2	121.24	80.26	86.69	-40.98	6.43	-34.55
	/o/	F1	66.76	49.05	35.40	-17.71	-13.65	-31.35
		F2	83.91	177.66	113.25	93.75	-64.41	29.34
	/u/	F1	74.22	74.96	29.06	0.74	-45.90	-45.16
		F2	483.43	247.37	167.84	-236.07	-79.53	-315.60
Male 2	/i/	F1	51.27	48.43	39.21	-2.84	-9.22	-12.06
		F2	145.24	120.84	220.92	-24.40	100.08	75.68
	/e/	F1	35.87	38.82	28.52	2.95	-10.31	-7.35
		F2	128.39	108.29	124.82	-20.10	16.53	-3.57
	/a/	F1	39.65	54.23	37.66	14.59	-16.57	-1.99
		F2	75.65	88.90	86.83	13.25	-2.07	11.18
	/o/	F1	41.54	55.39	47.54	13.85	-7.85	6.00
		F2	148.82	129.92	127.85	-18.90	-2.07	-20.97
	/u/	F1	65.71	81.56	56.15	15.86	-25.41	-9.55
		F2	293.95	232.18	170.68	-61.76	-61.50	-123.27

As we can see in Table 8, in 31 out of 40 cases, formants have a smaller standard deviation in Semester 6 than in Semester 1, indicating a higher command of pronunciation. Nevertheless, a statistical analysis is necessary in order to assess the statistical significance of said changes.

The data were first analysed globally, including the four speakers as a random variable. Lower standard deviations of vowels and more distant vowels would indicate more intelligible speech (e.g. Kim et al., 2011). A linear mixed model of the standard deviation data was carried out for the F1 and F2 of each of the vowels to see if the variable *Semester* was significant. The data for the F1 and F2 of all vowels met all the assumptions

in all cases except for F2 /i/ and F2 /e/. Therefore, F2 /i/ and F2 /e/ were transformed following statistical procedures (Eddington, 2015, p. 38-40) using the function *LG10 (STDVF2)*. The same procedure was followed as for the analysis of F1 and F2 in Section 4.2 to find the most parsimonious statistical model (Eddington, 2015, p. 117). A random intercept was fitted for each speaker and a random slope was included for each speaker and for each semester by speaker (e.g. Barr et al. 2013, p. 262). The resulting residuals were specified with a Variance Components covariance structure.

Table 9. Results of the linear mixed model for the standard deviation of F1 and F2 of /i/, /e/, /a/, /o/ and /u/. A random intercept and slope were fitted for each speaker and for each semester by speaker. The specified covariance structure was Variance Components.

Formant and vowel	**Result of the linear mixed model**
F1 /i/	F (2, 6) = .884, p = .461
F2 /i/	F (2, 6) = .344, p = .722 (from *LG10 (STDVF2)* transformed data)
F1 /e/	F (2, 6) = 2.439, p = .168
F2 /e/	F (2, 6) = .618, p = .570 (from *LG10 (STDVF2)* transformed data)
F1 /a/	F (2, 6) = 2.252, p = .186
F2 /a/	F (2, 6) = .317, p = .740
F1 /o/	F (2, 6) = 1.789, p = .246
F2 /o/	F (2, 6) = .272, p = .711
F1 /u/	F (2, 6) = 1.240, p = .335
F2 /u/	F (2, 6) = 5.876, p = .023

According to the tests in the linear mixed models, the only statistically significant difference in the standard deviation is for the F2 of /u/. Since we are now dealing with standard deviations, all vowels, semesters and participants had the same number of data points, which means that a Bonferroni corrected Tukey post-hoc test could be run. This test shows that the only statistically significant contrast is the one between Semester 1 and Semester 6 (p = .043). The contrast between Semesters 1 and 3 is almost significant (p = 0.52). If instability of vowel pronunciation, measured by the size of the standard deviation, is used to assess whether students' vowels become more stable, this

improvement will have to be rejected. The standard deviation does not change across the semesters, except for the F2 of /u/.

4.4. Vowel variability across the semesters per speaker

In order to understand vowel stability better, we also analysed it for each of the participants though a series of ANOVA tests. All the assumptions of the ANOVA were tested and all the F1 and F2 values for each of our participants met all the assumptions except F1 /o/ of Fem 1; this was analysed from the transformed data *F1 + 1000*. As in Section 4.2, effect sizes are only considered for variables which are identified as statistically significant. It is worth noting that, since each vowel phoneme for each speaker only has one standard deviation value per semester, within-subject variation had to be analysed for all vowels at the same time in order to have enough data for an ANOVA test.

Table 10. ANOVA results for the variation of the standard deviation of F1 and F2 per speaker per semester.

Speaker	Formant	ANOVA results	Effect size as per Cohen (1988)
Fem 1	F1	F (2, 12) = 2.372, p =.136, η^2 = .283	
	F2	F (2, 12) = 1.142, p =.351, η^2 = .160 (from *LG10 (STDVF2)* transformed data)	
Fem 2	F1	F (2, 12) = .300, p =.746, η^2 = .048	
	F2	F (2, 12) = .323, p =.730, η^2 = .051 (from *LG10 (STDVF2)* transformed data)	
Male 1	F1	F (2, 12) = 11.985, p =.001, η^2 = .666	Large
	F2	F (2, 12) = 1.034, p =.385, η^2 = .147 (from *LG10 (STDVF2)* transformed data)	
Male 2	F1	F (2, 12) = 1.464, p =.270, η^2 = .196	
	F2	F (2, 12) = .103, p =.903, η^2 = .017 (from *LG10 (STDVF2)* transformed data)	

As Table 10 shows, the only statistically significant difference between the standard deviation of F1 or F2 across the semesters is for the F1 of Male 1.

Table 11. p value for the difference between standard deviations for Male 1 across the semesters. * indicates statistically significant (p *value* < 0.05).

F1 Male 1	Sem 1	Sem 3	Sem 6
Sem 1		.045 *	.001 *
Sem 3	.045 *		.119
Sem 6	.001 *	.119	

Table 11 shows that the standard deviation of F1 for Male 1 is statistically significant between Semesters 1 and 3, and Semesters 3 and 6. Since the data from Table 8 show that the standard deviation for the F1 of all the vowels for Male 1 is smaller for Semester 3 than for Semester 1, and for Semester 6 than Semester 3 (with the exception of the F1 of /u/ having a slightly larger value in Semester 3 than in Semester 1), we can conclude that Male 1's vowels get more stable as the student advances through the semesters.

4.5. Difference of variability between vowels across all semesters

Another statistical analysis was run to compare changes of stability between the five Spanish phonemes using a linear mixed model. F2 /i/ and F2 /e/ were the only standard deviation values which violated any of the assumptions, and these were transformed using the function LG10 (STDVF2), as per Eddington (2015, p. 38–40). A random intercept was fitted for each speaker and a random slope was fitted for each speaker and for each vowel by speaker, as per Barr et al. (2013, p. 262).

Table 12. Results of the linear mixed models for the standard deviation of the F1 and F2 of /i/, /e/, /a/, /o/ and /u/ for all speakers across the three semesters.

Formant and vowel	Result of the linear mixed model
F1	F (4, 12) = 2.221 p = .128
F2	F (4, 12) = 5.183 p = .012*

The results of a Tukey post-hoc test are included in Table 13.

Table 13. Results of the Tukey post-hoc test for the linear mixed models of the standard deviation of F2 of /i/, /e/, /a/, /o/ and /u/ for all speakers across the three semesters

ST DV F2	**/i/**	**/e/**	**/a/**	**/o/**	**/u/**
/i/		.427	0.74	1.000	1.000
/e/	.427		1.000	1.000	.119
/a/	.074	1.000		1.000	.021 *
/o/	1.000	1.000	1.000		.415
/u/	1.000	.119	0.21 *	.415	

Regarding the standard deviation for F1 and F2, all vowels show the same variability between them in the F1. Regarding the F2, the Tukey post-hoc test identifies that the only statistically significant difference of standard deviation is the contrast of /a/-/u/, with /u/ displaying a larger variability across the semesters.

5. Discussion and conclusions

The acoustic and statistical analyses provide evidence that the production of L2 Spanish vowels of these four Australian students does not improve in a way which is statistically significant when there is no explicit pronunciation instruction, as little improvement is observed.

A comparison with previous studies of L2 Spanish vowel acquisition can be carried out in order to offer a general view of the learning process; however, differences in study designs and in methods makes comparing results between studies challenging (Long et al. 2018). For example, while Stevens (2011) focuses on vowel duration and analyses data using ANOVA, our study focuses on F1 and F2 values and analyses data

with a series of linear mixed models. Unlike the present study, Cobb & Simonet (2015) analyse the speech of L1 Spanish speakers and of intermediate and advanced speakers of Spanish with L1 English; despite some methodological differences between both studies, such as log-transforming F1 and F2 measurements and using by-subject ANOVAs in Cobb & Simonet (2015), their study and the present paper both report better improvement in the pronunciation of /u/. Although Long et al. (2018) analyses F1 and F2 of L2 Spanish vowels using linear mixed models, as in the present study, they measure vowels at midpoint and normalise F1 and F2 measurements, while we measured vowels in a section of a vowel of approximately 20 ms and did not normalise data for the reasons explained in the methodology section. Long et al. (2018) find no improvements in L2 Spanish vowels regarding the F1 or F2 for students on a four-week study abroad program, although they admit that the short duration of the course could be a limitation of the study; the current paper reaches similar results for at-home students with no explicit pronunciation training, as explained below.

Regarding our first research question –the improvement in Spanish vowels pronunciation by our participants–, the data show that the students' vowels became slightly more stable throughout their major but those changes were not statistically significant; Long et al. (2018) also report no improvement in F1 or F2 in vowels for students on a four-week study abroad program. As some studies point out (e.g. Flege 1988; Thomson & Derwing 2015, p. 339; Sturm 2019, p. 40), *some* improvement can take place without explicit pronunciation teaching. We have also seen that certain gains were indeed reported in Menke & Face (2010), Stevens (2011) or Cobb & Simonet (2015), but these changes were qualified as minimal due to their focus on foreign accentedness and native-like goals. In our case, the statistical analysis shows no significant difference

between the F1 and F2 values of vowels for all combined speakers or for each independent speaker between the three semesters. Our analysis suggests that the vowels of these speakers experience the same amount of change in the first part of their major and in the second part; we cannot claim that students make more progress in the first half of their major or in the second half. Flege (1988) and Thomson & Derwing (2015, p. 339) believe that some minimal gains may happen at the beginning of the learning process without phonetic instruction. Menke & Face (2010) conclude that more advanced students of Spanish produce more native-like vowels, although they have more issues with /a/, and Cobb & Simonet (2015) only find improvements in more advanced learners in /e/ and /u/. Although previous studies have seen some improvements in more advanced students, our informants' vowels do not show actual, consistent improvement throughout their major, even at more advanced levels of proficiency. We interpret this as additional evidence in favour of explicit pronunciation teaching (Lord, 2005; Sturm, 2019, p. 42; among many others) and we argue that it points to a possible degree of early fossilisation in the analysed students (Munro & Derwing, 2008; Derwing & Munro, 2014, p. 51).

Regarding research question 2 –vocalic changes across the semesters as a group and between each individual student–, the data do not show any improvements in the pronunciation of the vowels except for the F2 of /u/ when we analyse all speakers together. When we analyse each speaker separately, only the vowels of Male 1 become more stable throughout the major, although there are some small differences between the speakers. Fem 1 shows a very slightly higher change in vowel quality inside the second part of her major (Semester 3 to 6), although Male 1's vowels improve slightly more in the first part of the degree than in the second. For Fem 2 and Male 2, vowels tend to change equally in the first and second part of the major. Whereas some of the previous

literature pinpointed some specific issues with particular vowels (e.g. /a/ in Menke & Face 2010), our analysis shows that the limited changes are attested in the pronunciation of vowels throughout the major and are consistent for all vowels.

Finally, we acknowledge the limitation posed by the fact that our data sample may be too small for conclusions to be extrapolated to the whole population of students of Spanish who have Australian English as their first language. However, we consider that our article contributes data to an important line of research on longitudinal studies in L2 pronunciation teaching and learning and provides information collected from students in Australia, a country underrepresented in the field of L2 Spanish.

6. References

Alarcos Llorach, E. (1976). *Fonología Española* (4th ed.). Gredos.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random Effects Structure in Mixed Effects Models: Keep It Maximal. *Journal of Memory and Language, 68,* 255-278.

Boersma, P., & Weenink, D. (2016). *Praat: doing phonetics by computer*. In (Version version 6.0.19) <http://www.praat.org>

Celce-Murcia, M., Brinton, D. M., Goodwin, J. M., & Griner, B. (2010). *Teaching pronunciation. A course book and reference guide.* (2 ed.). Cambridge University Press.

Clopper, C. G., Pisoni, D. B., & De Jong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *The Journal of the Acoustical Society of America*, *118*(3), 1661-1676.

Cobb, K., & Simonet, M. (2015). Adult second language learning of Spanish vowels. *Hispania*, *98*(1), 47-60.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioural Sciences* (2 ed.). Lawrence Erlbaum.

Couper, G. (2006). The Short and Long–term Effects of Pronunciation Instruction. *Prospect*, *21*(1), 46 - 66.

Coxhead, A., & Boutorwick, T. J. (2018). Longitudinal Vocabulary Development in an EMI International School Context: Learners and Texts in EAL, Maths, and Science. *TESOL Quarterly*, *52*(3), 588 - 610.

Delicado Cantero, M., & Steed, W. (2015). La enseñanza de la pronunciación del español en Australia: actitudes y creencias de los profesores. *Journal of Spanish Language Teaching*, *2*(1), 18 - 35. https://doi.org/10.1080/23247797.2015.1012895

Delicado Cantero, M., Steed, W., & Herrero de Haro, A. (2019). Spanish pronunciation and teacher training: challenges and suggestions. In R. Gao (Ed.), *Key Issues in the Teaching of Spanish Pronunciation* (pp. 304 - 323). Routledge.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, *39*, 379 - 397.

Derwing, T. M., & Munro, M. J. (2014). Myth 1. Once you have been speaking a second language for years, it's too late to change your pronunciation. In L. Grant & D. M. Brinton (Eds.), *Pronunciation Myths. Applying Second Language Research to Classroorm Teaching* (pp. 34 - 55). University of Michigan.

Eddington, D. (2015). *Statistics for linguists: A step-by-step guide for novices*. Cambridge Scholars Publishing.

Field, A. (2018). *Discovering statistics using IBM SPSS Statistics: North American Edition* (5 ed.). Sage.

Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America*, *84*(1), 70 - 79.

Gil Fernández, J. (2007). *Fonética para profesores de español: de la teoría a la práctica*. Arco/Libros.

Hedia, S. B., & Plag, I. (2017). Gemination and degemination in English prefixation: Phonetic evidence for morphological organization. *Journal of Phonetics*, *62*, 34-49.

Henriksen, N. (2017). Patterns of vowel laxing and harmony in Iberian Spanish: Data from production and perception. *Journal of Phonetics*, *63*, 106-126.

Henriksen, N., & Harper, S. K. (2016). Investigating lenition patterns in south-central Peninsular Spanish /sp st sk/ clusters. *Journal of the International Phonetic Association*, *46*(3), 1 - 24.

Herrero de Haro, A. (2016). The acquisition of Spanish pronunciation by Welsh learners: Transfer from a regional variety of English into Spanish. *Circulo de Linguistica Aplicada a la Comunicacion*, *65*, 169 - 198.

Herrero de Haro, A. (2019). Consonant deletion and Eastern Andalusian Spanish vowels: The effect of word-final /s/, /r/, and /θ/ deletion on /i/. *Australian Journal of Linguistics*, *39*(1), 107-131. https://doi.org/10.1080/07268602.2019.1542935

Herrero de Haro, A. (2021). The vowel system of Eastern Andalusian Spanish speakers with articulation disorders: The effect of consonant deletion on preceding vowels in Western Almería. Lingua (Elsevier), 249 Article 102958. *Lingua*, *249*, Article 102958. https://doi.org/https://doi.org/10.1016/j.lingua.2020.102958

Herrero de Haro, A., & Andión Herrero, M. A. (2011). La enseñanza de la fonética española a hablantes de escocia e irlanda del norte. *Circulo de Linguistica Aplicada a la Comunicacion*, *47*, 28-64.

Herrero de Haro, A., & Andión Herrero, M. A. (2012a). La enseñanza de la pronunciación del castellano a aprendices irlandeses. Contrastes dialectales de interés. *Porta Linguarum*, *18*(18), 191-212.

Herrero de Haro, A., & Andión Herrero, M. A. (2012b). La fonética de aprendices de español del sur y del norte de Inglaterra: Consideraciones acerca de su enseñanza. *Revista Española de Lingüística Aplicada/Spanish Journal of Applied Linguistics*, *25*(25), 45-68.

Kennedy, S., & Foote, J. A. (2015). Second Language Speakers at University: Longitudinal Development and Rater Behaviour. *TESOL Quarterly*, *49*(1), 199 - 209.

Kim, H., Hasegawa-Johnson, M., & Perlman, A. (2011). Vowel contrast and speech intelligibility in dysarthria. *Folia Phoniatrica et Logopaedica*, *63*, 187 - 194. https://doi.org/10.1159/000318881

Kissling, E. M. (2013). Teaching Pronunciation: Is Explicit Phonetics Instruction Beneficial for FL Learners? *The Modern Language Journal*, *97*(3), 720 - 744.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, *39*(3).

Long, A. Y., Solon, M., & Bongiovanni, S. (2018). Context of learning and second language development of Spanish vowels. *Studies in Hispanic and Lusophone linguistics*, *11*(1), 59-87.

Lord, G. (2005). (How) Can we teach foreign language pronunciation? On the fffects of a Spanish phonetics course. *Hispania*, *88*(3), 557-567.

MacDonald, S. (2002). Pronunciation - views and practices of reluctant teachers. *Prospect*, *17*(3), 3 - 18.

Martínez Celdrán, E., & Elvira-García, W. (2019). Spanish vowels description and some guidelines to teach them. In R. Gao (Ed.), *Key Issues in the Teaching of Spanish Pronunciation. From Description to Pedagogy* (pp. 17 - 39). Routledge.

Mellado Prado, A. (2012). La pronunciación, un objetivo crucial para la enseñanza/ aprendizaje de la competencia comunicativa. In J. Gil Fernández (Ed.), *Aproximación a la enseñanza de la pronunciación en el aula de español* (pp. 11 - 41). Edinumen.

Menke, M. R., & Face, T. L. (2010). Second Language Spanish Vowel Production: An Acoustic Analysis. *Studies in Hispanic and Lusophone linguistics*, *3*, 181 - 214.

Moreno Fernández, F. (2002). *Producción, expresión e interacción oral*. Arco/Libros.

Munro, M. J., & Derwing, T. M. (2008). Segmental Acquisition in Adult ESL Learners: A Longitudinal Study of Vowel Production *Language Learning*, *58*(3), 479 – 502.

Nagle, C. L. (2017). A Longitudinal Study of Voice Onset Time Development in L2 Spanish Stops. *Applied Linguistics*, *40*(1), 86 - 107.

Nagle, C. L., & Huensch, A. (2020). Expanding the scope of L2 intelligibility research. Intelligibility, comprehensibility, and accentedness in L2 Spanish. *Journal of Second Language Pronunciation*, *6*(3), 329 - 351.

Orta Gracia, A. (2009). Creencias de los profesores acerca de la pronunciación y sus repercusiones en el aula. *Phonica*, *5*, 48 - 73.

Ortega, L., & Iberri-Shea, G. (2005). Longitudinal research in second language acquisition: Recent trends and future directions. *Annual Review of Applied Linguistics*, *25*, 26 - 45.

Poch Olivé, D. (2004). La pronunciación en la enseñanza del Español como Lengua Extranjera. *RedELE*, *1*.

Solon, M., Long, A. Y., & Gurzynski-Weiss, L. (2017). Task complexity, language-related episodes, and production of L2 Spanish vowels. *Studies in Second Language Acquisition*, *39*(2), 347-380.

Steed, W., & Delicado Cantero, M. (2014). First things first: Exploring Spanish Students' Attitudes toward Learning Pronunciation in Australia. *The Language Learning Journal*, *46*(2), 103 - 113. https://doi.org/10.1080/09571736.2014.963644

Steed, W., & Delicado Cantero, M. (2018). *Exploring pronunciation teaching and assessment in Australian L2 subject outlines* Third Pronunciation Symposium, University of Wollongong, Australia.

Stevens, J. J. (2011). Vowel duration in second language Spanish vowels: study abroad versus at-home learners. *Arizona Working Papers in SLA & Teaching*, *18*, 77 - 104.

Sturm, J. L. (2019). Current approaches to pronunciation instruction: A longitudinal case study in French. *Foreign Language Annals*, *52*(1), 32 - 44.

Thomas, E. R., & Kendall, T. (2007). *NORM: The vowel normalization and plotting suite. Online Resource.*

Thomson, R. I., & Derwing, T. M. (2015). The Effectiveness of L2 Pronunciation Instruction: A Narrative Review. *Applied Linguistics*, *36*(3), 326 - 344.

Torreira, F. (2012). Investigating the nature of aspirated stops in Western Andalusian Spanish. *Journal of the International Phonetic Association*, *42*(1), 49-63.

Torreira, F., & Ernestus, M. (2011). Realization of voiceless stops and vowels in conversational French and Spanish. *Laboratory Phonology*, *2*(2), 331-353.

Usó Viciedo, L. (2009). Creencias de los profesores en formación sobre la enseñanza de la pronunciación". , 8. . *MarcoELE. Revista de didáctica español como lengua extranjera*, *8*, 1 - 32.

Warren, P. (2018). Quality and quantity in New Zealand English vowel contrasts. *Journal of the International Phonetic Association*, *48*(3), 305-330.

Zielinski, B., & Pryor, E. (2020). Comprehensibility and everyday English use. An exploration of individual trajectories over time. *Journal of Second Language Pronunciation 6*(3), 352 – 379