An L2 Pronunciation Judgment Task

Adolfo Ausín and Megan Sutton
Michigan State University

1. Introduction

Elliott (2003) writes:

> "[w]ith the advent of more communicative approaches to language learning and instruction in the 1980s, explicit instruction in target language pronunciation became a thing of the past and was generally limited to the preliminary chapters of language textbooks or found in the back pages of the textbook appendixes." [1]

As a remedy to what we might call the pronunciation deficit of the communicative approach, many undergraduate Spanish programs across the US offer a course in phonetics and pronunciation in their third or fourth year. An indication of the interest in these types of courses is the number of phonetics / pronunciation textbooks that are available on the market, such as Guitart (2004), Hammond (2001), Quilis and Fernández (2003), Schwegler, Kempff and Barrutia (2007), Spicher, Sweeney and Pelayo Coutiño (2007), Stokes (2005), and Teschner (2000). Of course, not all these textbooks have the same emphasis, and there is great variation with respect to the degree of sophistication of the phonetic and phonological theory that one can find in these textbooks. Nevertheless, all of them have the practical purpose of improving the students' pronunciation.

At the same time that formal training in phonetics and pronunciation has found a way into the Spanish curriculum, there has been an increasing interest in finding out whether such an addition is adequate. One thing is to realize that the communicative approach to language teaching is not very effective at getting the students to pronounce correctly (after all, precise pronunciation is seldom an impediment for successful communication). It is a very different thing is to claim that formal teaching in phonetics can have an impact on improving the students' pronunciation. For this reason, there have been studies devoted to test whether formal instruction is an effective way of improving the students' pronunciation.

Although the students in a phonetics / pronunciation course might learn about the different modes of articulation or how the concepts of phonemes and allophones help us understand the differences between [s] and [z] in Spanish and English, the question is whether students actually improve in their practical knowledge of the language; that is, whether they put the phonetic theory to use. There are impressionistic data that seem to indicate that the students' pronunciation of Spanish does not improve after a phonetics class. If there is no improvement at all, it would be difficult to justify the need for this type of course, although not impossible, since one could argue that it is important that students are aware of the phonological properties of the language even if they cannot put them to practice.

Many studies that test the effectiveness of formal phonetics seem to indicate that formal training in phonetics has significant positive impact on the students' pronunciation. (See, for example, González-Bueno (1997), Lord (2005), Gaff and Loewen (2007), and Elliott (2003), which provides an overview of L2 studies in the acquisition of phonology.) A notable exception to this is Sutter (1976), who found that having formal training in pronunciation did not correlate with better pronunciation.

[1] Similar remarks can be found in Pennington and Richards (1986), Terrell (1989), Elliott (1995), Cenoz and Lecumberri (1999), and Lord (2008) among others.

2. Testing linguistic competence

In this paper we present an experiment designed to test whether a phonetics class actually makes an impact on the linguistic competence of the enrolled students. The novelty of our experiment resides in the way we test our subjects' knowledge of Spanish pronunciation. In most experiments that we are aware of, the subjects produce some speech samples which are then rated by native speakers of the target language, or are measured and the results compared with those of native speakers. For instance, Colantoni and Steele (2006) and Face (2006) had their subjects read a passage, recorded their passages, analyzed some relevant tokens, generated relevant spectrograms and then measured features such as stop length and voicing or the presence of a brief closure. On the other hand, Gaff and Loewen (2007) recorded their subjects performing a series of tasks (free response picture description, paragraph reading, sentences reading, word list reading) and then raters judged the subjects' utterances.

Although this way of testing pronunciation seems ideal, it frequently proves to be impractical. Since it requires a considerable amount of time to collect and analyze the data, frequently the number of subjects under investigation has to be kept small, which might end up making the (statistical) analysis of the data more problematic. Also, the reliance on subjective raters adds another level of complexity which again might make the data analysis more problematic and cumbersome. For these reasons we decided to look for an alternative way of determining the linguistic competence of the subjects, and we decided to assess perception, not production.

There have been many studies that have focused on the perception abilities of second language learners. For example, Zampini (1998) measured the VOT perceptual boundaries between /p/ and /b/. She modified the VOT of /p/ and /b/ in increments of 5 milliseconds and used this VOT continuum to determine at what point the subjects placed the difference between /p/ and /b/. (She found no correlation between perception and production of /p/ and /b/.) Flege and MacKay (2004) tested the ability of native Italian speakers to perceive English vowels correctly. They used two different testing strategies but in both the subject had to listen to a series of English minimal pairs (/bit/ vs. /bɪt/). Nibert (2006) presents what could be considered a Truth Value Judgment Task (TVJT). She is trying to see whether non-native speakers of Spanish can use the intonational clues that distinguish the two interpretations of sequences like "lilas y lirios amarillos" in the same way native speakers do.

(1) Example of a minimal group, using the text "lilas y lirios amarillos"

Tonal structure of the intonation contours		Possible meanings:
a. [[lilas] [y lirios amarillos]]	→	A. 'lilacs and yellow irises'
H- L- L%		
b. [[lilas y lirios] [amarillos]]	→	B. 'yellow lilacs and yellow irises'
H- L- L%		
c.[[lilas y lirios amarillos]]	→	A or B. No disambiguating medial H- phrase accent is present.
L- L%		

Nibert presented the above utterances and asked her subjects whether each "utterance refers to two types of flowers: lilacs (=lilas) of some unknown color, and yellow irises (=lirios)." Finally, Grenon (2006) used a perceptual discrimination test to test whether Japanese speakers could discriminate certain contrasts present in English but not in Japanese. Some of the nonsense words that she used were /sɪd-θɪd/ and /zug-ðug/.[2]

In all of these perception studies the subjects "perceive" correct grammatical sequences, that is, fragments that were pronounced correctly.[3] Now we would like to turn our attention to two studies that followed a different approach: Sakow and McNutt (1993) and Altenberg (2005).

[2] See Flege (2003) for more studies on L2 perception (and production) as well as for theoretical considerations that fall beyond the scope of this article.

[3] In the Zampini study the subjects were not listening to normal fragments in the sense that they had been artificially modified by the researcher, but the subjects were under the assumption that they were "normal" fragments. In the Grenon study, even though the words were made up, they were still pronounced correctly because they conformed to the phonological requirements of English phonology, and therefore, were grammatical.

Sakow and McNutt (1993) tested how well native speakers of Japanese and Korean could distinguish between correct and incorrect pronunciations of /r/ in English. They tested both external and internal perception. In external perception the utterances being judged were recorded by native speakers of English. In internal perception, subjects judged their own utterances that they had previously recorded. They found that subjects were better at external perception.

Altemberg's (2005) goal was to determine the knowledge that native speakers of Spanish learning English had regarding what constitutes a possible consonant cluster in English. One of the tasks that she used, a Metalinguistic Judgment Task, asked subjects to judge the acceptability of written non-words, such as "slad" and "fnod" as possible words in English. That is, she used an explicit grammaticality judgment task to determine the subjects' phonological knowledge.

The testing techniques that were used in Sakow and McNutt (1993) and Altenberg (2005) could be described as a grammaticality judgment task (GJT), since the subjects were asked to determine whether certain stimuli were acceptable or not in the target language. The use of grammatical judgments in SLA has been debated (see Sorace (1996)), however Chaudron (2003) notes that their use is standard in L2 syntactic and semantic research. As an example of the syntactic application of the GJT, consider Zyzik (2006). In her study, Zyzik focuses on the status of null objects in learners of Spanish. One of the tasks that her subjects performed was to determine whether the sentences that appear in (2) are acceptable or not.

(2) a. ¿Qué hiciste con el pastel? Lo puse en el refrigerador.
'What did you do with the cake? I put it in the fridge.'
b. *Ellos compraron la cerveza y pusieron en el refrigerador.
'They bought the beer and put in the fridge.'

Zyzik used her subject responses to these sentences to determine the (syntactic) knowledge that her subjects had about this type of constructions. Similarly, Pérez-Leroux, Majzlanova, and Sánchez-Naranjo (2003) use a grammaticality judgment task to determine their Spanish L2 learners' knowledge of the semantic factors involved in the choice between preterite and imperfect in Spanish. They asked their subjects to rate semantically well- and ill-formed sentences like the following:

(3) a. Hice la tarea pero no me dejaron terminarla. (semantically ill-formed)
b. Hacía la tarea pero no me dejaron terminarla. (semantically well-formed)
'I did/was doing the homework but they didn't let me complete it.'

Pérez-Leroux et al. use their subjects' ratings for sentences like the previous ones to determine their subjects' knowledge of the semantic features involved in the correct choice of preterite vs. imperfect.

As can be seen, the GJT has been used to assess the linguistic knowledge in the realms of syntax, semantics, and to a lesser degree phonetics / phonology. For our experiment, we decided to follow Altermberg's and Sakow and NcNutt's strategy and tested our subjects' knowledge of Spanish phonetics by using a pronunciation GJT. The general idea was to offer subjects good and bad recordings of Spanish words, and ask the students to rate them, to see whether they were able to tell the good recordings from the bad ones.

One immediate practical advantage of our approach was the simplification of the analysis of the subjects' responses. As mentioned earlier, in a pronunciation production task, after the subject produces the relevant samples, the researcher has to spend time and energy analyzing the data, in a fairly complicated process of analyzing what the subjects have produced. Furthermore, the complexity of the process increases the chances of error and decreases the possibility of studying a large subject population. In a pronunciation grammaticality judgment task, the analysis of the subject's response is much simpler, the possibilities of error are virtually zero, and the analysis of large subject population is completely plausible.

3. The experiment

3.1. Subjects

The test (a pre- and a post-test) was initially administered to all the students enrolled in two sections of a Spanish pronunciation and phonetics class at an American university. The class is described in the catalog as a "[p]honetic description of the sound system of Spanish, developed linguistically and applied to the improvement of pronunciation and spoken Spanish." Students take this class approximately during their third year of college-level instruction.

The class follows somewhat closely the structure of the textbook, Stokes (2005). The students are introduced to the spelling, phonetic and phonological systems of Spanish. In general a typical class period includes a review of daily homework (which might involve some phonetic transcription), lecture by the instructor[4] about some phonetic topic, and some oral practice exercises that cover the day's material. Even though the textbook comes with an audio program CD, it was rarely used. At the end of the semester there was an oral project where students had to make a recording of the poems in the book / CD.

A total of 53 students enrolled in the class took some portion of the test. Also, a group of 4 native speakers (graduate assistants in the Spanish department) also took the test. Initially we had hoped to run a control group with native speakers. However, since we were not able to obtain a group of native speakers large enough to establish a comparison, we decided to exclude native speakers from the present research. Thus, out of the 53 undergraduate students who took some part of the test, we eliminated the results from the 6 undergraduate native speakers. We also eliminated the results from the 8 students who missed the pre or the post-test.[5] In short, the results that we are reporting in this paper are the results obtained from the 39 nonnative speakers who took both the pre and the post-test.

3.2. Stimuli

The stimuli consisted of 34 words pronounced in one of three different ways. In one version, the pronunciation was intended to be native-like. In a second version, consonant sounds were pronounced in an English-like manner, and in the third version both consonants and vowels were English-like.[6] Thus, there was a total of 102 items. Some of the targeted sounds are: fricative [β,ð,ɣ], flap [ɾ], trill [r], un-aspirated [p, t, k], final [l], presence of [v] and [h], diphthongized vowels and [ə]. The words were recorded by a non-native speaker of Spanish who is a linguistically trained speaker of English with the ability to produce native-sounding Spanish words. All words were evaluated by native speakers: a group of native speakers went over the test and we used those results to verify the accuracy of the stimuli.

3.3. Method

The test was administered through the university course management system as part of the class requirements. The students were asked to go to a specific computer lab on campus where they would log on to the class web page.[7] There they would find a link to the test. From the students' perspective, this test was just an activity that was graded for completion only. At the end of the semester, the students were asked to give consent to use their responses in the present study.

The test was preceded by instructions regarding the test and a brief background questionnaire whose main purpose was to identify native speakers of Spanish. The instructions that the students were given are the following. (A Spanish version was also included.)

---

[4] The instructor of this class is a native speaker from Spain.

[5] Two of these excluded subjects actually took both the pre- and the post-test but had technical problems during the post-test (they could not hear the stimuli) and answered randomly.

[6] As pointed out by an anonymous reviewer, there is a logical fourth possibility: words pronounced with English-like vowels but Spanish-like consonants. We did try to produce this type of stimuli but we were not successful.

[7] The access to the test was restricted to one particular computer lab on campus through an IP filter. If the subjects tried to access it from anywhere else, access was denied.

(4) Experiment instructions

> In the following, you will see a series of written Spanish words and you will hear a voice that reads each word twice. After listening to the reading, you will have to evaluate the pronunciation of each word on a scale from 5 (perfect, native like) to 1 (very bad). Before that, you will have to answer a few questions about yourself.
>
> 5 Perfect / Native like -- Perfecto / Como un hablante nativo
> 4.
> 3.
> 2.
> 1. Very bad -- Muy mal

After reading the instructions and completing the background questionnaire, the actual test began. A word appeared on the screen and a recording was played twice. The subjects had to rate the recording and after doing that the new word / recording would appear. The order of presentation of the 102 items was randomized for each subject. The test was administered twice, once towards the beginning (the pre-test) and once towards the end of the semester (the post-test).

4. Results and analysis

For illustration purposes, we will describe one of the words contained in the stimuli and a selection of the subjects' response for that word. One of the 34 words included in the stimuli was "hubo". This word was recorded in three different ways. Version A was native-like: approximant "b" [β] and Spanish vowels. Version B had an incorrect occlusive "b" [b], but still with Spanish vowels. Finally, version C not only had occlusive "b" [b] but a diphthongized final "o" [o^{u}] also. The following table includes the responses from the first ten subjects and the averages for all the 39 subjects.

(5) Sample answers and averages for "hubo"

	Version A [ú.βo] Native-like consonants and vowels			Version B [ú.bo] English-like, stop "b"			Version C [ú.bo^{u}] English like, stop "b" and diphthongized final "-o"		
Subject	pre-test	post-test	Diff	pre-test	post-test	Diff	pre-test	post-test	Diff
1	5	5	0	3	2	1	1	1	0
2	4	5	-1	4	4	0	2	3	-1
3	5	5	0	2	4	-2	1	1	0
4	5	5	0	3	1	2	4	1	3
5	5	5	0	4	4	0	4	2	2
6	5	5	0	3	3	0	3	2	1
7	5	5	0	5	2	3	4	1	3
8	5	5	0	5	3	2	3	3	0
9	5	5	0	4	3	1	3	2	1
10	5	4	1	4	3	1	2	2	0
⋮									
Average for all 39 Subjects	**4.846**	**4.872**	**-0.026**	**3.692**	**2.769**	**0.923**	**2.410**	**1.718**	**0.692**

The data were averaged across the students' responses on the pre-test and post-test for each of the three versions (A, B, and C). Recall that group A contained both native-like consonants and native-like vowels, B had generally native-like vowels but English-like consonants, and C was composed of words with English-like consonants and vowels. The averages for the A, B, and C words in the pre- and post-test are shown below, along with the differences between the two tests and *p*-values for the differences as determined by a paired t-test.

(6) Averages for all words by group type

	Pre-test Average	Post-test Average	Difference	*p*-value
Group A	4.52	4.53	-0.01	0.811
Group B	2.58	2.12	0.46	0.000
Group C	1.67	1.39	0.22	0.000

The first item of interest is that the students appear to make a clear differentiation between groups A, B, and C even in the pre-test, and this distinction also appears in the post-test. Words in group A were assigned an average value of 4.52, very close to the ideal score of 5 for native-like productions. Words in group B received a 2.58, indicating a middle value, while words in group C were given an average value of 1.67, which is reasonably close to the value of 1 that words with a completely English-like value should have received.

In terms of change in the students' scores over the semester, it is clear that there is no improvement in the native-like words from group A. However, the average started at 4.52, close to the expected rating of 5 for native-like Spanish. Most of the A cases require no discussion since there is little or no change; one exception in the A group will be discussed in detail later. There is a change in the B and C words, which a paired t-test showed to be statistically significant. The results show that students gave the words higher ratings at the beginning of the semester than at the end, indicating that they realized at the end that the recordings were more English-like. Although we could use the comparison between the groups B and C to try to figure out the impact that the English-like vowels (diphthongized, long vowels, schwas) had in the ratings of each member of the group C, we will exclude from the discussion the results of group C. We cannot directly infer anything from the students' reactions to the words in group C, since the vowels and consonants were both non-native-like, and thus the students could have been reacting to either. Overall, the words in group B tended to show the largest differences between the pre-test and post-test, showing a change in students' reactions to English-like consonants.

In what follows we describe the word types that we tested, giving examples for each type of word.[8]

(7) Types of words that were tested

Word Type	**These words contain…**	**Some L2 learners…**	**Stimuli**	**Stimulus Versions**
Z	the grapheme "z"	pronounce the words with a [z]	"caza"and "paz"	A: [ká.sa] B: [ká.za]
BDG	intervocalic /bdg/	fail to produce the approximant allophone found in Spanish	e.g. "hubo", "oda", "abogado"	A: [ú.βo] B: [ú.bo]
Final L	word-final "l"	use the velarized "l" [ɬ] in word-final position	e.g. "sal", "cual"	A: [sal] B: [saɬ]
H	the grapheme "h"	pronounce the "h", forgetting that it is silent in Spanish	e.g. "mohoso", "bahía	A: [ba.í.a] B: [ba.hí.a]
V	the grapheme "v"	pronounce the words with a [v]	e.g. "ave", "uva"	A: [ú.βa] B: [ú.va]
PTK	/ptk/	pronounce them aspirated, as in English	e.g. "taco", "pato"	A: [tá.ko] B: [t^{h}á.ko]
R / RR	the flap [ɾ] or alveolar trill [r]	pronounce them as an alveolar approximant [ɹ], like in English	e.g. "ara", "arra"	A: [á.ɾa], [á.ra] respectively B: [á.ɹa] for both

[8] The Z-, H- and V-words are different from the rest in that the subjects' knowledge of the grapheme-phoneme relation is being tested. We thank an anonymous reviewer for bringing up this issue.

- Z words: These are words that have the grapheme “z”. This type of word is problematic for learners because they tend to pronounce them with a [z]. We tested two examples: “caza” y “paz”. For “caza”, the A version was [ká.sa] and the B version was [ká.za].
- BDG words: These are words that contain intervocalic /bdg/. These types of words are problematic for the students because they normally fail to produce the approximant allophone found in Spanish. Some of the examples that we tested were “hubo”, “oda” and “abogado”. See above for the different versions of “hubo”.
- Final L words: There are words with final “l”. Frequently students use the velarized “l” [ɬ] in word-final position. Some of the words that we tested are “sal” and “cual”. The A version for “sal” would be [sal] and the B version would be [saɬ].
- H words: These are words that contain the grapheme “h”. Some students tend to pronounce it forgetting that “h” in Spanish is silent. Some words that we tested are “mohoso” y “bahía”. The A version for the latter would be: [ba.í.a] and the B version [ba.hí.a].
- V words: These are words that contain the grapheme “v”. Some students pronounce these as a [v]. Some words that we tested are “ave” y “uva”. The A version for the latter would be [ú.βa] and the B version [ú.va].
- PTK words: These are words that contain /ptk/. Some students pronounce them aspirated as in English. Some of the words that we tested are “pato” y “taco”.
- R / RR words: These are words that contain the flap [ɾ] or the alveolar trill [r]. Some students pronounce them as an alveolar approximant [ɹ], like in English. Some of the words that we tested are “ara” and “arra”. The A versions were [á.ɾa] and [á.ra] respectively, and the B version was [á.ɹa] for both.

The following table summarizes the results for the word types that showed a statistically significant difference between the pre- and the post-test.[9] The results by word type were obtained in the same manner as for the general analysis of groups A, B, and C.

(8) Word types that showed a statistically significant difference ($p < 0.05$)

		Pre-test Average	Post-test Average	Difference	*p*-value
Z	Group A	4.14	4.44	-0.30	0.036
	Group B	2.24	1.55	0.69	0.000
	Group C	1.37	1.18	0.19	0.007
BDG	Group A	4.69	4.81	-0.12	0.015
	Group B	3.16	2.43	0.73	0.000
	Group C	2.20	1.67	0.53	0.000
-L	Group A	4.45	4.41	0.04	0.727 (ns)
	Group B	2.00	1.54	0.46	0.000
	Group C	1.53	1.28	0.24	0.006
H	Group A	4.45	4.56	-0.12	N/A[10]
	Group B	2.91	2.37	0.54	0.001
	Group C	1.83	1.36	0.47	0.000
V	Group A	3.78	3.39	0.40	0.016
	Group B	3.04	1.58	1.46	0.000
	Group C	1.97	1.30	0.68	0.000

[9] We have omitted from the discussion and will not report in this paper the data regarding words that could not fit in exactly one of the types that we just described. Thus, we disregard the data for words like “acaba” which in the B version has both an aspirated [k^h] and an occlusive [b], or “cascabel” which in the B version has an occlusive [b] and a final velarized [ɬ].

[10] The paired t test could not be run for this particular set of data because of the distribution of the data. A Wilcoxon Signed-Rank Test did not find the different statistically significant (p(2-tail)=0.3173).

Z WORDS: In Latin American Spanish, the grapheme "z" generally represents an unvoiced alveolar fricative, and the voiced alveolar fricative occurs only as an allophone of the /s/ before voiced consonants, so it is improper to pronounce words like "caza" as [ká.za]. This is unlike English, which uses the "z" to represent the voiced alveolar fricative. In light of this difference, it is understandable that students gave a rating of 4.14 to the productions of "caza" as [ká.sa]. This is a bit low when compared to the overall average of 4.52 for the A group as a whole, and might indicate a lack of awareness of the difference between the English Z and the Spanish Z. For group B, the students gave a 2.24 to productions of the Z with a voiced sibilant in the pre-test (for example, a word like "paz" was pronounced as [paz] in group B, in contrast to [pas], which would be more typical in general Latin American Spanish). In the post-test, this score fell by 0.69, to a 1.55, indicating a clear opinion that [paz] is not a native-like production. From this, we infer that some students benefited from a discussion of the differences between the Z in English and Spanish.

BDG WORDS: A similar pattern occurred for the voiced stops /b, d, g/. These phonemes differ from the English in that they are only realized as stops after a pause or nasal, or in the case of /d/, after an [l]. Intervocalically, they are realized as the approximants [β, ð, ɣ]. Thus, to maintain our pattern of incorrectly produced consonants in group B, students heard words like "abogado" pronounced with intervocalic stops, as in [a.bo.ɡá.do], rather than the appropriate intervocalic approximates in group A, as in [a.βo.ɣá.ðo]. Students originally gave the group B production with stops an average rating of 3.16, but lowered the score to a 2.43 in the posttest. This lends strong support to the idea that students benefited from the discussion of the difference between /bdg/ in English and Spanish.

FINAL L WORDS: The data for the final L words show generally the same pattern as the data overall: there was no sign of a significant change in the native-like words, while the non-native-like words were initially rated somewhat high and received lower scores in the posttest. However, there was a significant change in the B (and C group). Remember that there is an important difference between Spanish and English for final "l." While the word-medial and word initial "l" is the same in both languages, English contains the word-final "dark l." In English words like "wonderful," the final /l/ is velarized (and the tip of the tongue might not be touching the alveolar ridge), whereas Spanish words like "cual" contain the same sound as the word-medial "l". Final L words in the B group (pronounced with a dark "l": [saɬ], [kwaɬ]) were initially given a rating of 2.00 By the end of the semester, this rating dropped significantly to 1.54.

H WORDS: The data for H words also show the same pattern as the data overall: there was no sign of a significant change in the native-like words, while the non-native-like words were initially rated somewhat high and received lower scores in the posttest. The Spanish "h" differs from English in that in Spanish the grapheme "h" is not pronounced. The H words in the B group contained an English-like [h]. Thus, for the word "bahía," the stimuli included an [h], as in [ba.hí.a] rather than the Spanish-like [ba.í.a]. Subjects gave these words with [h] an average score of 2.91 at the beginning of the class, and a score of 2.37 at the end. Although a 2.37 is somewhat high considering that the students were hearing a sound that should have been entirely silent, this is an improvement.

V WORDS: The V was an interesting case for two reasons, the first of which was the large change in students' judgments of it from the beginning and end of the semester. For group B, students heard productions of words like "ave" as [á.ve], with the English voiced labiodental fricative, which is not used in Spanish. At the beginning of the semester, students rated this at a 3.04, a relatively high score, which plummeted to a 1.58 at the end of the semester. From this we infer that by the end of the semester the students differentiated between the voiced bilabial and voiced labiodental fricatives, recognizing that [á.ve] is not a native-like production.

The other notable feature of the data for the V was its remarkably low scores in group A on both tests. Words such as "ave" and "uva" in group A obtained a score of 3.78, much lower than the rest of the word types in group A, and the average fell to 3.39 in the post-test. This represents a significant change, which puzzled us because we had expected the students to rate words from group A closer to a five after the class due to their pronunciation with a bilabial approximant. It is important to note that subjects did not show an unusual response to the same bilabial approximant when the words shown contained the letter "b" rather than a "v." One possible explanation for this unexpected result is that the subjects saw a "v" and that made them "hear" a [v], which they know is not a good option in Spanish. The interference of the letter "v" in the pronunciation of /b/ has already been noted by Zampini (1994).

So far we have looked at word types that showed a significant difference between the pre and post-test. Now let's turn our attention to those word types that didn't.

(9) Word types that did not show a statistically significant difference

		Pre-test Average	Post-test Average	Difference	P-value
PTK	Group A	4.59	4.46	0.12	0.170 ($n_{s/r}$=29)
	Group B	2.39	2.33	0.06	0.342 ($n_{s/r}$=30)
	Group C	1.48	1.41	0.07	*0.039 ($n_{s/r}$=26)
RR	Group A	4.81	4.82	-0.01	0.85
	Group B	1.45	1.40	0.05	0.62
	Group C	1.05	1.13	-0.08	0.18
R	Group A	4.37	4.32	0.05	0.59
	Group B	1.72	1.55	0.17	0.20
	Group C	1.40	1.28	0.12	0.11

* = statistically significant

PTK WORDS: In the PTK words, there was no significant difference between the pre- and post-test in the A group (native-like) or the B group (aspirated consonants [p^h, t^h, k^h] but native-like vowels). There was a significant difference in the C group, but remember that in this group the vowels are also pronounced non-native-like.[11] In other words, for a word like "taco," there was no difference between the pre and the post-test for the A version ([tá.ko]) nor in the B version ([t^há.ko]). The only significant difference was for the C version ([t^há.ko^u]) with a final diphthongized vowel. Since the presence of the aspirated "t" [t^h] did not make a difference in the B version, the difference in the C version is most likely due to the diphthongized vowel, not the aspirated consonant.

R / RR WORDS: Finally, there was no significant difference between the pre and the post-test in any of the versions of the R / RR words. Thus, using as an example the word "arra", the A version [á.ra] was already identified as the right pronunciation by most subjects in the pre-test (4.81, fairly close to the highest possible value of 5), whereas the B version ([á.ɹa]) and the C version ([á.ɹə]) were already given a fairly low score in the pre-test: 1.45 and 1.05, respectively.

To summarize our results, we saw significant progress toward the expected ratings between the beginning and the end of the semester. Most of the word types showed an improvement, the only exception being PTK and R/RR words.

5. Discussion

Since there is a significant difference between the pre- and post-test in most of the types of words, formal phonetics instruction is supported. In most of the cases, students' ratings increased for the A type of words (those pronounced native-like) and decreased for B and C type of words (those pronounced "with an accent"). Thus, this study corroborates the idea that teaching formal phonetics is effective in improving the linguistic competence of the students.

[11] For the PTK words, the data followed a non-normal distribution and the t-test was not applicable, so we conducted a Wilcoxon signed-rank test on the data.

There were only two types of words that showed no significant improvement: PTK and R / RR words. The reason for the lack of improvement in the R / RR words is most likely due to the fact that the subjects' ratings of those words were already very good in the pre-test so there was not much room for improvement to begin with.

As for the lack of improvement in the PTK words, an explanation is less obvious. It is true that the general rating for these words is better than the average, but many types of words with similar or better ratings in the pre-test still showed improvement in the post-test. Several possibilities to explain the lack of improvement on PTK words come to mind: maybe the instruction on this particular topic might not have been good enough; alternatively, the difference between aspirated and unaspirated stops in English and Spanish is less salient than the other differences studied in this experiment.

Another important contribution of this study is further support for an underused approach to studying phonological proficiency. Compared to production based strategies, the Grammaticality Judgment Task strategy that we have used to determine the competence of our subjects is remarkably simpler, and still seems to be doing the work of tracking the students' progress. We believe that this opens the door for a reliable and simple way of testing phonological linguistic competence.

The approach that we have developed is obviously limited since it only measures the ability of the subjects to distinguish good and bad pronunciations in the target language. But on the other hand, it gives us the opportunity of obtaining quantitative data of one aspect of the linguistic competence of the learner. It might not be the most important aspect of the linguistic knowledge of the learner, but it is still part of it.[12]

We believe that the limitations of this technique are compensated by its ease of application. Once the experiment is set up, any number of subjects can participate without any extra time or energy on the part of the experimenter. The process of analyzing subject responses is normally very tedious, time consuming and prone to error. By simplifying this aspect of phonological assessment, we have a testing procedure that is simple enough to be applied to large subject populations.

Finally, we would like to comment on the overall performance of the subjects in the pre-test. In many cases, there was not a big difference between the pre-test and the post-test with the group A words because the initial values in the pre-test were already very high. Based on this, one could argue that students at this level already have sophisticated phonetic knowledge that allows them to differentiate between native and non-native pronunciation. On the other hand, there is impressionistic data that students at this level are far from being able to produce the pronunciations that they judge as native-like. This would be in agreement with Altenberg's (2005) findings: She found a clear contrast between the ability of her subjects to judge illicit consonant clusters and their ability to produce those clusters correctly. If it is true that intermediate students have a sophisticated knowledge of Spanish pronunciation that is much more refined than their performance would lead us to expect, there are two additional conclusions that could be drawn.

One is related to the way we teach formal phonetics. Since intermediate students fail to produce certain sounds in a native-like manner, it is assumed that they don't know these sounds, but this may turn out to be inaccurate. We should favor those approaches of phonetics that rely on the learners' knowledge and use that knowledge as a teaching tool. A recent and interesting approach along these lines is Lord (2008) who uses peer review of student created podcasts.

The other conclusion is related to the way we think about the SLA process. Frequently it is taken for granted that there is a fundamental difference between the acquisition of syntactic knowledge and phonological knowledge.[13] However, this might not be accurate. A cursory look at the literature seems to indicate that L2 learners achieve different degrees of success at syntactic and phonological levels. But it could be the case that this is due to the ways in which we are measuring their knowledge. Frequently L2 syntactic knowledge is assessed with judgment tests such as the grammaticality

[12] Following Altenberg (2005), we assume that measuring this type of knowledge is valuable even if it turns out that it is not directly correlated to other components of the phonological knowledge of the learner.

[13] For instance, Moyer (1999) says: "Anecdotally, it is common to point to nonnative speakers (NNSs) whose linguistic abilities are apparently native, with the exception of pronunciation. Seliger (1978) and Long (1990) have suggested separate critical periods for different levels of language."

judgment test or the truth value judgment test, which do not require the subject to produce any utterances. On the other hand, phonological knowledge is more frequently tested using production tests, and rarely tested using grammaticality judgment tasks as in the present study. The implicit phonological knowledge that students seem to have could be used to support the idea that there is UG access in phonology.[14] After all, the students are showing a clear ability, for instance, to distinguish between stop [b] and approximant [β], even in the pre-test.

As tempting as the conclusion from the previous paragraph may be, the evidence might not be completely compelling. The results are telling us that when the subjects have to decide whether [í.ba] or [í.βa] are native pronunciations of "iba", they overwhelmingly identify [í.βa] as a native pronunciation. This could be for two reasons. One is that they have access to UG, and even though there is no [b] / [β] contrast in English, access to UG enables them to figure out the difference between [b] and [β]. The other reason is that they simply assume that the most exotic variant, in this case [í.βa], is the right one; after all, it is the one that sounds less English-like.

Although these conclusions are interesting, there is no doubt that future research is needed to validate the findings in this study. For instance, to test the validity of the claims and arguments in the previous two paragraphs, a similar test should be run with speakers that have little or no knowledge of Spanish. There are also a few things in the test itself that could be changed. It would probably be a better idea to use made-up words in the stimuli to avoid any type of familiarity effects. Each item in the stimuli should exclusively target one particular sound. A control group of subjects not enrolled in a pronunciation course should also be used. In the present study we saw a clear difference between the pre-test and the post-test and we attributed this difference to the phonetics / pronunciation class they were enrolled in. However, we have not ruled out the possibility that the improvement could have been achieved by taking any other third year Spanish class. Having a control group would help us better understand the impact of a phonetics class.

To conclude, we wanted to test whether a phonetics class at the intermediate level helps students improve their knowledge of the pronunciation of Spanish. For that purpose, we designed an experiment based on a grammaticality judgment task. Our results indicate that formal phonetics instruction does have a positive effect. Also, our experiment reveals that our subjects have clear intuitions regarding what a native pronunciation should be like, even though they are far from using it in their production. If this is true, there could be an argument here for UG access. But before we continue that line of reasoning, more research needs to be conducted to confirm these results. Finally, we have developed a testing strategy that is remarkably simple, even if limited, which might open the door to further study of L2 phonology.

References

Altenberg, E. P. 2005. The judgment, perception, and production of consonant clusters in a second language. *IRAL-International Review of Applied Linguistics in Language Teaching* 43.1: 53-80.

Cenoz, J., and L. G. Lecumberri. 1999. The effect of training on the discrimination of English vowels. *IRAL-International Review of Applied Linguistics in Language Teaching* 37.4: 261-275.

Chaudron, C. 2003. Data collection in SLA research. In *The Handbook of Second Language Acquisition*, eds. C. J. Doughty & M. H. Long, 762–828. Malden, MA: Blackwell.

Colantoni, L. and J. Steele. 2006. Native-like attainment in the L2 acquisition of Spanish stop-liquid clusters. In *Selected Proceedings of the 7th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages*, eds. C. A. Klee and T. L. Face, 59-73. Somerville, MA: Cascadilla Proceedings Project.

Elliott, A. R. 1995. Foreign language phonology: Field independence, attitude, and the success of formal instruction in Spanish pronunciation. *The Modern Language Journal* 79.4: 530-542.

Elliott, A. R. 2003. Phonology: Staking out the territory at the turn of the century: Integrating phonological theory, research, and the effect of formal instruction on pronunciation in the acquisition of Spanish as a second language. In *Spanish Second Language Acquisition: State of the Science*, eds. B. A. Lafford and M. R. Salaberry, 19-46. Georgetown University Press.

Face, T. L. 2006. Intervocalic rhotic pronunciation by adult learners of Spanish as a second language. In *Selected Proceedings of the 7th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages*, eds. C. A. Klee and T. L. Face, 47-58. Somerville, MA: Cascadilla Proceedings Project.

[14] Grenon (2006) also supports the idea that there is UG access in L2 acquisition of phonology using an experiment based on perception.

Flege, J. E. 2003. Assessing constraints on second language segmental production and perception. In *Phonetics and Phonology in Language Comprehension and Production, Differences and Similarities*, eds. N. O. Schiller and A. Meyer, 319-355. Berlin: Mouton de Gruyter.

Flege, J. E., and I. R. A. MacKay. 2004. Perceiving vowels in a second language. *Studies in Second Language Acquisition* 26.1: 1-34.

Gaff, J. and S. Loewen. 2007. Teaching Spanish pronunciation: Does it work in the long run? Paper presented at the *Hispanic Linguistics Symposium* at the University of Texas at San Antonio.

González-Bueno, M. 1997. The effects of formal instruction on the acquisition of Spanish stop consonants. In *Contemporary Perspectives on the Acquisition of Spanish*, eds. W. R. Glass and A.T. Pérez-Leroux, Vol. 2, : 57–75. Somerville, Massachusetts, Cascadilla Press.

Grenon, I. 2006. Adults still have direct access to UG: Evidence from the perception of a non-native feature contrast. In *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006),* eds. M. G. O'Brien, C. Shea, and J. Archibald, 51-62. Somerville, MA: Cascadilla Proceedings Project.

Guitart, J. M. 2004. *Sonido y sentido: Teoría y práctica de la pronunciación del español*. Georgetown Studies in Spanish Linguistics. Washington, D.C.: Georgetown University Press.

Hammond, R. M. 2001. *The Sounds of Spanish: Analysis and Application (with Special Reference to American English)*. Somerville, MA: Cascadilla Press.

Long, M. H. 1990. Maturational constraints on language development. *Studies in Second Language Acquisition* 12.3: 251-85.

Lord, G. 2005. (How) Can we teach foreign language pronunciation? On the effects of a Spanish phonetics course. *Hispania* 88.3: 557-67.

Lord, G. 2008. Podcasting communities and second language pronunciation. *Foreign Language Annals* 41.2: 364-379.

Moyer, A. 1999. Ultimate attainment in L2 phonology. *Studies in Second Language Acquisition* 21.1: 81-108.

Nibert, Holly J. 2006. The acquisition of the phrase accent by beginning adult learners of Spanish as a second language. In *Selected Proceedings of the 2nd Conference on Laboratory Approaches to Spanish Phonetics and Phonology*, ed. M. Díaz-Campos, 131-148. Somerville, MA: Cascadilla Proceedings Project.

Pennington, M. C., and J. C. Richards. 1986. Pronunciation revisited. *TESOL Quarterly* 20.2: 207-225.

Pérez-Leroux, A. T., M. Majzlanova, and J. Sánchez-Naranjo. 2003. The light flashed until dawn: the L2 acquisition of the semantic mappings of Spanish tenses and problems of s-selection and compositionality. In *Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference*, eds. J. M. Liceras, H. Zobl and H. Goodluck, 227-237. Somerville, MA: Cascadilla Proceedings Project.

Piske, T., I. R. A. MacKay, and J. E. Flege. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29.2: 191-215.

Quilis, A., and J. A. Fernández. 2003. *Curso de Fonética y Fonología Españolas para Estudiantes Angloamericanos*. Madrid: Consejo Superior de Investigaciones Científicas, Instituto de Filología.

Sakow, M., and J.C. McNutt. 1993. Perception of /r/ by native speakers of Japanese and Korean: Internal and external perception. *IRAL-International Review of Applied Linguistics in Language Teaching* 31.1: 46-53.

Schwegler, A., J. Kempff, and R. Barrutia. 2007. *Fonética y Fonología Españolas*. Hoboken, NJ: John Wiley & Sons.

Seliger, H. W. 1978. Implications of a multiple critical period hypothesis for second language learning. *Second Language Acquisition Research: Issues and Implications*: 11-19.

Sorace, A. 1996: The use of acceptability judgments in SLA research. In *Handbook of Second Language Acquisition,* eds. W.C. Ritchie and T.K. Bhatia, 375-409. San Diego, CA: Academic Press.

Spicher, L. L., F. M. Sweeney, and R. Pelayo Coutiño. 2007. *La Voz: Fonética y Fonología Españolas*. Upper Saddle River, N.J.: Pearson, Prentice Hall.

Stokes, J. D. 2005. *¡Qué Bien Suena!: Mastering Spanish Phonetics and Phonology*. Boston, MA: Houghton Mifflin.

Sutter, R.W. 1976. Predictors of pronunciation accuracy in second language learning. *Language Learning* 26.2: 233-253.

Terrell, T. 1989. Teaching Spanish pronunciation in a communicative approach. In *American Spanish Pronunciation: Theoretical and Applied Perspectives*, eds. P. C. Bjarkman and R. M. Hammond, 196-214. Washington DC: Georgetown University Press

Teschner, R. V. 2000. *Camino Oral: Fonética, Fonología y Práctica de los Sonidos del Español*. Boston, MA: McGraw Hill.

Zampini, M. L. 1994. The role of native language transfer and task formality in the acquisition of Spanish spirantization. *Hispania* 77.3: 470-81.

Zampini, M. L. 1998. The relationship between production and perception of L2 Spanish stops. *Texas Papers in Foreign Language Education* 3.3: 85-100.

Zyzik, E. 2006. A look at missing objects in L2 Spanish. In *Selected Proceedings of the 9th Hispanic Linguistics Symposium*, eds. N. Sagarra and A. J. Toribio, 192-202. Somerville, MA: Cascadilla Proceedings Project.

Selected Proceedings of the 12th Hispanic Linguistics Symposium

edited by Claudia Borgonovo, Manuel Español-Echevarría, and Philippe Prévost

Cascadilla Proceedings Project Somerville, MA 2010

Copyright information

ISBN 978-1-57473-440-9 library binding

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, e-mail: sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Ausín, Adolfo and Megan Sutton. 2010. An L2 Pronunciation Judgment Task. In *Selected Proceedings of the 12th Hispanic Linguistics Symposium*, ed. Claudia Borgonovo, Manuel Español-Echevarría, and Philippe Prévost, 234-245. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2420.