THE DEVELOPMENT OF ENGLISH VOWELS BY NATIVE SPANISH SPEAKERS

by

Andrew Jeske

B.A. University of Wisconsin-Madison, 2007

MA, University of Pittsburgh, 2012

Submitted to the Graduate Faculty of

The Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2016

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Andrew Jeske

It was defended on

March 31, 2016

and approved by

Claude Mauk, PhD, Director of Less Commonly Taught Languages Center

Matthew Kanwit, PhD, Assistant Professor

Natasha Tokowicz, PhD, Associate Professor

Dissertation Advisor: Alan Juffs, PhD, Professor

THE DEVELOPMENT OF ENGLISH VOWELS BY NATIVE SPANISH SPEAKERS

Andrew Jeske, PhD

University of Pittsburgh, 2016

Previous studies have shown that exposure to a second language (L2) changes one's perception and production of L2 sounds to become more native-like (e.g., Flege, 1995; Flege & MacKay, 2004). This change has been documented most commonly among immigrants after they move to an environment where the L2 is the main means of communication (e.g., Jia & Aaronson, 2003). However, many people get introduced to a foreign language in school and it has been found that the amount of L2 exposure provided to students in a foreign language classroom is not equivalent to the amount of exposure experienced by immigrants and, therefore, will not produce the same kinds of benefits (e.g., White & Genesee, 1996). This dissertation aimed to examine the effects different amounts of L2 exposure in a classroom environment can have on the perception and production of English front vowels (/i ɪ e ɛ æ/). The participants for this study were a group 2nd, 4th, and 6th graders from two schools (one bilingual, one non-bilingual) who shared the same L1 (Spanish) and age of first exposure to L2 English, but who differed in the amount of L2 exposure they received each week (5 hours vs. 14 hours). The participants' perception was examined through a categorical discrimination task and their production of English front vowels was elicited via a picture-naming task. Predictions surrounding the relative discrimination difficulty of certain vowel pairs were made through the Perceptual Assimilation Model for Language Learners (PAM-L2) (Best & Tyler, 2007) and their production was evaluated through the Speech Learning Model (SLM) (Flege, 1995). Results from the perception task found a significant effect for school - the additional L2 exposure provided to the bilingual

school students beneficially impacted their overall performance on the categorical discrimination task. However, this advantage was not clearly exhibited in their production because participants from each school were able to produce each English front vowel in a significantly distinct way. Further research will have to be conducted to see if the differences in production between the two schools affected the intelligibility of the target words.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I would like to offer my sincere thanks to the members of my committee: Claude Mauk, Natasha Tokowicz, Marta Ortega-Llebaria and Matthew Kanwit, for the time they have spent with me throughout this process offering advice and guidance. I would also like to thank my advisor, Alan Juffs, whose support has been crucial to the completion of this dissertation.

1.0 LITERATURE REVIEW

1.1 INTRODUCTION

Much of the second language acquisition (SLA) research indicates that people who are exposed to a second language (L2) early in life are more likely than older learners to reach native-like performance in many, if not all, linguistic domains (e.g., Granena and Long, 2013). Second language speech acquisition often supports this position because early learners often display native (Flege and MacKay, 2004; Flege, MacKay and Meador, 1999) or near-native (Højen and Flege, 2006; Fabiano-Smith & Goldstein, 2010) perception of certain L2 vowels and consonants and speak with less of an accent (Oyama, 1976; Baker, Trofimovich, Flege, Mack & Halter, 2008; Flege, Yeni-Komshian and Liu, 1999). In an attempt to explain the advantage held by early learners, researchers have focused on three variables of L2 learning – (1) the critical period, (2) native language transfer, and (3) environmental factors like quantity and quality of input.

The aim of this dissertation is to examine the effects of previous language experience and amount of L2 exposure on the perception and production of the English front vowels /i ɪ e ɛ æ/. This was accomplished by examining L1 Spanish students enrolled in either a Spanish-English bilingual school or a non-bilingual school in Colmenar Viejo, Spain. Results were determined through an analysis of the participants' performance on (1) a Categorical Discrimination task and (2) a picture-naming task and findings were discussed through the lens of the Speech Learning

Model (SLM) (Flege, 1995) and Perceptual Assimilation Model for Language Learners (PAM-L2) (Best, 1995; Best & Tyler, 2007).

1.2 AGE EFFECTS

The first of three theories that have lent support to the *earlier is better* axiom, is based around the biological changes that occur in the brain as one matures (Lenneberg, 1967). A critical period, as it pertains to second language acquisition, is the time during which a person must acquire or begin acquiring a second language. Some scholars believe that maturational developments preclude achieving native-like competence in an L2 if a person were to begin acquiring this language after the termination of this critical period. Lenneberg (1967) claimed that a loss of plasticity, or a lateralization, of the brain inhibits L2 speech acquisition past the point of puberty. Scovel (1969), citing Penfield (1965) for support, highlighted Penfield's remark that the speech center of the brain cannot be relocated to the non-dominant cerebral hemisphere after severe injury to the left hemisphere past the age of 12. It is this immobility, or *cerebral dominance*, that impedes the accurate learning of L2 speech starting around puberty (Scovel, 1969).

However, as it relates to L2 speech acquisition, people exposed very early in life to a second language often continue to speak with a non-native accent (Flege, Munro & MacKay, 1995; Guion, Flege, & Loftin, 2000) and are generally less accurate in perception tasks than monolinguals (Flege, MacKay & Meador, 1999). And even after years of experience, perceiving and/or producing specific L2 contrasts remains a problem for many people. Examples of such difficult contrasts are: (1) Japanese /r/-/l/ distinction (Goto, 1971; Yamada, 1995) and (2) Spanish /i/-/ɪ/ (Escudero & Boersma, 2004; Flege, Bohn & Jang, 1997).

But even though native-like accuracy may be impossible for the majority of the population, there are numerous instances of late L2 learners being indistinguishable from native speakers of the target language (e.g., Ioup et al., 1994; Moyer, 1999) or score within the same range as native speakers when their pronunciation was judged on the amount of foreign accent (Muñoz & Singleton, 2007). In a study by Bongaerts (1999), native speakers of French judged the pronunciation of L2 French produced by nine adult L1 Dutch speakers. In spite of that fact that the L1 Dutch adults all began learning French at or after the age of 12 and received only approximately 2-3 hours of French instruction per week until they were 18, three of them acquired a native-French accent. It is exactly this variability in results that has made many researchers call into question the finality of the Critical Period Hypothesis' (CPH) claims.

In his review of CPH research, Scovel (2000) claimed that some version of the Critical Period Hypothesis is supported by the majority of applied linguists and psycholinguists. This can be seen through the emergence of a less rigid definition of what constitutes a critical period that reflects the fact that there is not one age at which all learning stops or is impossible. Instead, some scholars have begun to acknowledge the notion of *sensitive periods*: windows during which language learners are more likely to acquire native-like levels of performance (Oyama, 1976; Long, 1990; Granena & Long, 2012).

Moreover, many scholars have also concluded that different aspects of language may have unique sensitive periods. Seliger (1978) and Long (1990) suggest that the sensitive period for phonetics/phonology happens earlier than the sensitive period of syntax. This is supported through studies that have found that earlier first exposure is needed to speak an L2 without a non-native accent (Oyama, 1976) than it is to perform at native-like levels on syntactic tasks (Patkowski, 1980).

In a first-of-its-kind study, Granena and Long (2012) examined the same group of adults across phonological, lexical-collocations, and morphosyntactic domains. The participants in this study were highly proficient L2 speakers who differed primarily in the age at which they were first exposed to the second language. The researchers compared these L2 speakers' scores against those of native speakers and identified a distinct age of onset (AOA) window for each domain, at which, once passed, native-like performance would be very rare. Within a single population of L2 learners, Granena and Long showed that there is not one sensitive period but rather multiple sensitive periods that depend on the aspect of language being examined: phonological (~6 years), lexical-collocations (~12 years), morphosyntactic (late teens).

But even L2 exposure during the earliest of these sensitive/critical periods does not guarantee native-like competence in all linguistic domains (e.g., Flege & MacKay, 2004), which is why researchers such as Flege (1995) and (1999) began to dissect AOA and concluded that it could actually be broken down into separate variables that could have significant impacts on language learning on their own. Two of the variables that emerged from this decoupling were the quantity and quality of language exposure that came along with age.

1.3 QUALITY AND QUANTITY OF INPUT

Age interacts with many variables, including one's exposure to the L2. One's exposure to an L2 can be measured in various ways: (1) age of first exposure (2) length of exposure and (3) quality of exposure – and at times, it can be difficult to separate these variables. The majority of the studies investigating the role of quantity of exposure on L2 speech acquisition have looked at it in terms of age of arrival (AOA) - the age at which a participant arrives in a country where their first language (L1) is not the language of the wider community (Munro, Flege, & MacKay, 1996; Flege, MacKay & Meador, 1999; Cebrian, 2006; Abrahamsson and Hyltenstam, 2009). These studies focus primarily on adult immigrants after they have already received large amounts of L2 exposure in a natural environment and, in many cases, had presumably already reached their level of ultimate attainment. Results here indicate that early arrivers perceive and produce L2 sounds significantly better than late arrivers (Munro et al., 1996; Flege, Yeni-Komshian, & Liu, 1999; Højen & Flege; 2006), and when coupled with infrequent usage of the L1, early exposure with high quality input becomes a good predictor of native or near-native perception and production (Flege, MacKay & Meador, 1999; Flege & MacKay, 2004).

In another study, Jia and Aaronson (2003) followed a group of L1 Chinese speakers through their first three years of living in the US (thereby removing the variable of length of residence (LOR)) and assessed how AOA impacted L2 acquisition. They divided their subjects into two groups based on their AOA: Group 1 arrived at age nine or below and Group 2 arrived at age 10 or after. The researchers saw that the younger children significantly outperformed the older children on grammaticality and translation tasks after three years in the US. The authors suggest that the younger children surpassed their older counterparts because, even though they had been exposed to the language for the same amount of time, the younger participants were exposed to richer L2 experiences through peer and social interactions. Similar studies have mentioned that younger learners are able to interact with native speakers of the target language more because they are often enrolled in school whereas adult learners enter the workplace and have more exposure to L2-accented speech (e.g., Dekeyser & Larson-Hall, 2005; Flege, 2009; Flege, Birdsong, Bialystock, Mack, Sung, & Tsukada, 2006; Jia & Aaronson, 2003).

As studies like the one above show, younger learners' overall attainment is often higher than older L2 learners' (Baker et al., 2008; Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004; Krashen, Long, & Scarcella, 1979), but when first starting out in an L2 environment, younger learners do not fare as well. In one study examining age effects on immigrant language learners' ability to perceive and produce English vowels (Jia, Strange, Wu, Collado, & Guan, 2006), the authors found an age effect that changed over time: initially, the adults performed significantly better than the children, then after two years of residence in the US the adults and children performed equally well, but after 3-5 years of living in the US, the children performed significantly better than the adults. These studies by Jia and Aaronson (2003) and Jia et al. (2006), along with many others (e.g., Krashen, Long, & Scarcella, 1979; Snow & Hoefnagel-Höhle, 1979) highlight an important aspect of age and L2 learning: in the beginning, adult learners learn an L2 faster but, with time, younger learners often surpass their more mature counterparts in nearly all linguistic domains (Granena & Long, 2012).

This rate advantage that adults initially have in a naturalistic setting (e.g., Ekstrand, 1976; Snow and Hoefnagel-Hohle, 1977, 1978) has been replicated in the foreign language classroom (Genesee, 1987). One theory behind the adults and adolescents learning faster after the moment of first exposure is due to their ability to take advantage of strategies that younger learners don't or can't utilize because of the state of their cognitive development (Collier, 1989; Genesee, 1978). Not all variables that are significant in a natural environment are also found to be significant in foreign language classrooms though.

Age of arrival has been a relatively reliable predictor of L2 attainment in perception and production (Flege, Bohn & Jang, 1997; Højen & Flege, 2006) but the age at which someone starts formal foreign language classes is not. In fact, when moving from a natural to formal language learning environment, multiple studies have found that learning a foreign language at a younger age does not necessarily correspond with having a higher level of attainment (DeKeyser, 2000; Mayo, 2003; Johnson & Newport, 1989; Muñoz, 2006).

Lecumberri and Gallardo (2003) examined the potential effects of AOA on Basque children who began learning English in their schools at either 4, 8 or 11 years old. Each group received 2-3 hours of English per week at school. After approximately 600 hours of English (roughly 5 years later) the participants took part in a phoneme discrimination task and a story telling task, which L1 English speakers later used to give the participants intelligibility scores. Unlike some cases with early exposure to an L2 in a natural environment, the early learners of L2 English (4 and 8) performed significantly worse on both the perception and production tasks. Similar studies evaluating the effects of age in the foreign language classroom in which students receive fewer than five hours of exposure per week often find no consistent benefit to early L2 education (Larson-Hall, 2008) or they have found an advantage to those starting later (e.g., Mayo, 2003). White and Genesee (1996) suggested that this is because the amount of exposure one receives in a school is insignificant compared to the exposure people receive in a natural setting. If children learn in a more implicit manner (DeKeyser, 2000), such a low level of L2 input may not be enough for them to build L2 systems.

In summary, in natural L2 settings, AOA is a good predictor of high ultimate attainment of L2 speech acquisition. The age of first exposure in the classroom, however, does not have the same effect – neither in the short-term nor the long-term. Even though many studies find that older learners perform better than younger ones in syntactic and morphological tasks (Flege, Yeni-Komishian, Liu, 1999), only those with a large exposure to the L2 sound system early in life perceive and produce L2 sounds significantly better (e.g., Flege et al., 2006). Due to the nature of many foreign language programs, significant exposure to an L2 sound system is not possible. Therefore, a question arises: *how much L2 exposure is needed to significantly benefit L2 speech acquisition?* The aim of the present study is to evaluate performance on a perception and production task by groups of students that share the same age of first exposure, but who differ in the amount of L2 English they are exposed to on a weekly basis (approximately 4 hours vs. 14 hours).

1.4 L1 TRANSFER/INTERFERENCE

Infants are able to distinguish any two sounds used by speakers of a natural language (e.g., /i/ and /ɪ/ in the minimal pair /hit/-/hɪt/) (Kuhl, 1994). But by six months they have already begun to create L1 phonetic prototypes that attract non-native phones (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992) and, over the second six months, they lose the ability to discriminate between certain sounds that are not contrastive in their L1 (Werker and Tees, 1984; Conbboy & Kuhl, 2010). As this universal awareness weakens, their ability to discriminate language-specific contrasts strengthens (Rivera-Gaxiola, Silva-Perevra, & Kuhl 2005) due to increased experience with their ambient language and a more established L1 sound system (Werker and Tees, 1984).

However, L1 phonetic category formation is a years-long process and is the result of increased exposure to the ambient language (Hazan & Barrett, 2000; Flege & Eefting, 1986). Even though L1 attunement happens within the first year, adjustments are made to this developing sound system throughout adolescence resulting in less rigid phonetic categories and less accurate L1 discrimination until adulthood (Hazan and Barrett, 2000). One reason many researchers believe early language learners are more receptive to interlanguage phonetic/phonemic differences is because the L1 sound system is still forming (Baker et al., 2008; Flege, 1995). Bond and Adamescu (1979) used this rationale to support their finding of how a 4-year-old was able to distinguish L1 and novel L2 plosives at a rate better than chance whereas adolescents and adults were not.

There are several theories that predict how language learners acquire the sounds of an L2 including the Speech Learning Model (SLM) (Flege, 1995) and the Perceptual Assimilation Model for Language Learners (Best, 1995; Best & Tyler, 2007). Each theory makes its own predictions about how people learn the L2 sounds, but they both base their predictions around the interactions between the L1 and L2 sound systems.

In the Speech Learning Model, Flege (1995) predicts that certain L2 sounds will be easier to acquire than others based on the magnitude of perceived differences between that L2 sound and its closest L1 counterparts: sounds that are unlike any of the L1 sounds (*new sounds*) will be perceived easily, whereas those L2 sounds that are relatively *similar* will be the most difficult. In many cases when a sound from the L1 and L2 are perceived to be *nearly identical*, perception and production will not change. Instead, a diaphone is established – a single phone that is used in both languages. The SLM also makes predictions about the ease with which language learners will be able to produce new sounds. Flege (1995) claims that perception necessarily precedes production claiming that if a learner cannot distinguish two sounds perceptually, there is an expectation that the same learner will produce the sounds similarly (Flege et al., 1997).

When comparing the phonetic categories of highly proficient English and French bilinguals, Flege (1987) noticed that the /u/ category produced by both groups was not native-like in either language. Instead, the French and English bilinguals produced an intermediate vowel between the monolinguals' French /u/ and the monolinguals' English /u/. Flege suggested that the similarities between the two categories blocked accurate L2 category formation conflating the two categories into one.

Another model that predicts language learners' success of perceiving L2 sounds based on cross-linguistic interaction is the Perceptual Assimilation Model for Language Learners (PAM-L2) (Best, 1995; Best and Tyler, 2007). PAM-L2 hypothesizes that the ease of discriminability of two L2 sounds is based on how those two sounds assimilate to the L1 sound system (Figure 1). If two L2 sounds are assimilated equally well onto a single L1 category (*single-category contrast*), the language learner will have a difficult time hearing a difference. If two L2 sounds are mapped onto a single category, but at different rates, perceiving a difference between the two sounds will be relatively easy (*category-goodness*). When two L2 sounds are perceived as being most similar to two different L1 categories, it will be easy to perceive a difference in these two sounds (*two-category*). And lastly, when L2 sounds do not assimilate well onto any L1 category, they are said to be *uncategorizable* and the difficulty of discriminating these sounds depends on their level of similarity to all surrounding L1 sounds.

Even though PAM-L2 doesn't make direct predictions about the difficulty experienced in producing L2 sounds, there is the recognition that perception and production are related.

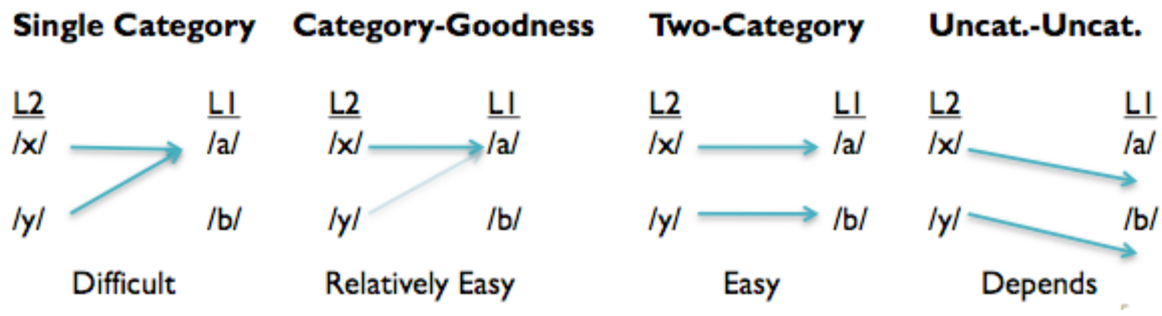


Figure 1. *Four L2-L1 Vowel Assimilation Patterns (PAM-L2)* Illustrates the four different kinds of L2-L1 vowel assimilation patterns that are used to predict how easily (or difficult) it will be for second language learners to perceive any given L2 contrast.

Another tenet shared by SLM and PAM-L2 is the belief that perception is not stagnant, even after having learned an L1 (Best & Tyler, 2007; Flege, 1995). Perception patterns can develop to more closely follow the patterns of native speakers of the target language (Best & Strange, 1992; Gottfried, 1984; Neufeld, 1988; Fabra & Romero, 2012; Scovel, 1969). Perceptual training in the classroom has been shown to improve language learners' ability to distinguish L2 contrasts more accurately (Iverson & Evans, 2009; Logan, Lively, & Pisoni, 1991) and help them establish representations of new sounds (Lee & Lyster, 2015; Strange & Dittman, 1984). High variability perception training with feedback can be of special help to perceive the most difficult sounds accurately (Cenoz & Lecumberri, 1999) and has been shown to have beneficial effects on perception even months after the training has ended (Wang & Munro, 2004).

Specialized perceptual training isn't always required for learners to establish more native-like L2 categories. In a study examining the perception and production abilities of adult American immigrants from varied language backgrounds (e.g., German, Korean, Mandarin, Spanish), Flege et al. (1997) found that adults with more L2 experience were able to perceive and produce English vowels better than adults with less experience.

1.5 THE RELATIONSHIP BETWEEN PERCEPTION AND PRODUCTION

1.5.1 Perception before production

In the field of L2 speech acquisition, the relationship between perception and production has frequently been examined, but no complete consensus exists over how these two aspects of speech interact. At its most basic, the question becomes *must a learner be able to accurately perceive an L2 sound before he/she can produce it or can production precede accurate perception?* Even though there is no definitive proclamation that can be made regarding the relationship between L2 perception and production, the culmination of 70+ years of L2 speech acquisition research has led the majority of contemporary researchers to support the claim that, in most cases, perception precedes production (Best, 1995; Borden, Gerber & Milsark, 1983; Escudero, 2006; Flege, 1995; Leather, 1999; McAllister, 1997).

Marckwardt (1946), after conducting a phoneme identification task with L1-Spanish/L2-English speakers, noticed parallels between his students' perception and production errors and suggested that if a language learner can't "hear accurately" the sounds in an L2, they will not be able to produce them (p. 106). Since then, numerous other studies have continued to examine this relationship using more quantitative measures.

One of these studies (Neufeld, 1988) assessed the perception and production capabilities of L1 English speakers who learned French after the age of 16. Even though a number of late learners performed equally as well as native French speakers on perception tasks requiring them to identify minute phonological anomalies in spoken words and phrases, their semi-spontaneous speech was judged by L1 francophones as being foreign. It was this *phonological asymmetry* – perceiving sounds that they could not produce – that led the author to suggest that accurate L2 perception precedes its production.

Rochet (1995)[1] came to a similar conclusion upon examining the perception and production patterns of L1 English and L1 Portuguese speakers on the French high vowels /i/, /y/ and /u/. Neither Portuguese nor English has a high, front rounded vowel, but both do have phones acoustically similar to the French /i/ and /u/. Despite the similarity of phones inhabiting the upper portions of the vowel space in Portuguese and English, the L1 Portuguese speakers perceived /y/ to be most similar to Portuguese /i/ and the L1 English speakers perceived /y/ to be more similar to English /u/. When participants from these two language groups produced French words containing /y/, their mispronunciations mirrored their perception biases – L1 French speakers judged the English participants' mispronunciations of /y/ as /u/ and the Portuguese's as /i/. He concluded that issues in production were rooted in inaccurate perception.

Taking this assumption that inaccurate perception leads to inaccurate production, a number of studies emerged examining the effects of perceptual training on production (Bradlow, Pisoni, Akahane-Yamada & Tohkura, 1997; Bradlow, Akahane-Yamada, Pisoni & Tohkura, 1999; Rvachew, 1994; Rvachew, Nowak & Cloutier, 2004). Many of these studies have found that by addressing areas of perceptual confusion through high-variability perceptual training and feedback, the benefits extend beyond perception itself and have a positive impact on production as well. Lambacher, Martens, Kakehi, Marasinghe and Molholt (2005), for example, provided

[1] Inspired by a similar study by Valdman (1984)
[2] Malta, Cyprus, Spain, Italy, Austria, Norway, Macedonia, Liechtenstein and Luxembourg all require foreign language

L1 Japanese speakers with six weeks of perception training of five English vowels /æ ɑ ʌ ɔ ɝ/. The participants' performance in a forced identification perception task improved significantly from the pretest to the post-test (16%), which was also significantly higher improvement compared to the control group (5%). Importantly, only the training group's productions became significantly more identifiable from the pretest to the post-test.

The benefits of perception training are not limited to the acquisition of individual segments, like vowels (Thomson, 2011) and consonants (Bradlow et al., 1997, 1999; Hardison, 2003), but extend to tone as well (Wang, Jongman, Sereno, 2003; Wang, Spence, Jongman, Sereno, 1999). Wang and colleagues provided perceptual training to a group of L1 English speakers learning Mandarin and found that the participants could identify Mandarin tones significantly more accurately after the training and they were able to generalize their training to new stimuli and new speakers. Additionally, L1 Mandarin speakers were able to identify the tone being produced by the language learner significantly more frequently after training.

It is important to note, however, that perceptual training can affect the perception and production to varying degrees. In the example above, the L1 English participants' identification of Mandarin tones improved by 21%, but their production only improved by 18%. A similar pattern was found by Bradlow et al. (1997) after native Japanese (NJ) speakers went through training to help them identify the difference between English /ɹ/ and /l/. She and her colleagues found a substantial improvement in perception between the two phones but only modest gains in production. Although these results still support claims by SLM that perception precedes production, it also highlights a shortcoming: *when does production catch up to perception?*

Following the introduction of SLM (Flege, 1995), Flege and his colleagues amended aspects of model to address other possible factors that can affect L2 phonetic category formation like AOA, amount of L1 use (Flege & MacKay, 2004), and amount of exposure (Bohn & Flege, 1997). They explain that early on in the learning of an L2, learners' speech acquisition follows that of their native language – where perception necessarily precedes production. However, Bohn & Flege (1997) also claimed that with increased exposure to an L2, learners' production is more likely to be beneficially affected than their perception: perception and production no longer progress at the same speed.

1.5.2 Production before perception

Despite the evidence supporting the theory that accurate perception necessarily precedes production, findings from a few studies suggest that this is not universally so. Darcy & Krüger (2012), for example, found that L1 Turkish children who began learning L2 German in kindergarten could accurately produce both segments of the vowel contrast /iː/-/ɪ/ but were not able to perceive the difference in an AXB discrimination task as well as their L1 German counterparts. A study by Sheldon and Strange (1982) found that accurate L2 production of /ɹ/ and /l/ actually preceded accurate perception of L1 Japanese participants. Similar findings with another group of native Japanese speakers suggested that accurate pronunciation of these /l/ and /ɹ/ could be accomplished through an awareness of the difference in articulator placement and not the difference in acoustic cues (Goto, 1971). This point was also recognized by Saito (2013) who acknowledged that accurate production of L2 sounds can precede accurate perception when the learners rely on articulatory memory and not acoustic cues.

There have also been a number of studies in which language learners are perceived to have native-like pronunciation (Neufeld, 1979; Neufeld, 1988), but these often rely on imitating native speakers of the target language. Different production tasks require varying levels of cognitive processing and imitation is ‘optimal for performance’ because it requires minimal processing (Strange, 2006) and says little about the language learner’s understanding of the L2’s phonological system (Neufeld, 1988). Studies have shown that Broca’s aphasics often mimic others’ productions better than when they produce the same sounds spontaneously (Trost & Canter, 1974). Galunov & Chistovich (1966) stated that imitation tasks partially bypass categorization because perception is only required at the sub-phonetic level, meaning the perception of the sound is linked to its production before any phonemic segments are realized.

Although not specifically a theory used to predict or explain L2 speech perception and production, the Motor Theory can be used to explain accurate responses and decreased choice reaction times in shadowing tasks conducted in the L1 (Fowler, Brown, Sabadini & Weihing, 2003; Porter & Castellanos 1980) and L2 (Muchinsky, 1983). The perception of gestures allows participants in shadowing tasks to partially bypass categorization; in the case of L2 learning, imitation decreases the effect of L1 interference.

In Shockley, Sabadini, and Fowler (2004), participants produced better imitations of a target word when asked to repeat it during a shadowing task than when the target word was read. The native English speakers in that study imitated the stimuli with longer than average VOT even though extra long VOT is not phonemic in their L1. In a separate imitation study (Neufeld, 1979), adult L1 English participants received pronunciation instruction and listened to an 18-hour video with a Chinese or Japanese speaker. At the end of their exposure, they imitated a number of phrases presented to them in the L2, which were later presented to native Chinese or Japanese speakers. Forty-five percent of the L2 Chinese speakers and 55% of the L2 Japanese speakers were identified as being native speakers of those languages.

These studies add support to the claim that adults have not lost the ability to perceive or produce phonetic distinctions that are not phonemic in the L1 – they may just have trouble categorizing them. The SLM posits accurate production for those phones that are perceived accurately. However, there exists the possibility that the seemingly inaccurate predictions made by SLM are not shortcomings of the model itself, but the result of a production task that cannot realistically measure more abstract concepts like L2 category formation.

1.6 MOTIVATION

The studies outlined above generally support the notion that early exposure to a sufficient amount of quality input beneficially affects L2 learners in a way that often ultimately separates them from late learners. These studies also show that even though immigrant (e.g., Flege & MacKay, 2004; Jia & Aaronson, 2003) and adult foreign language learners (e.g., Larson-Hall, 2008; Lee & Lyster, 2015) have been studied extensively, much less has been done to understand the acquisition of L2 perception and production in the first years of foreign language instruction in elementary education. To the best of my knowledge, no studies have been published examining the early effects of differing amounts of exposure to an L2 in a classroom environment – specifically on the perception and production of L2 vowels.

Understanding how the perception and production of L2 sounds may change over the course of formal foreign language programs is of crucial importance. Childhood foreign language courses are gaining popularity in the US and Europe and many people in these regions are starting foreign language education early on. According to the European Commission ("Foreign language learning statistics," 2016), a majority of students are studying English in primary school and, as of 2014, foreign language instruction of some kind in primary school was mandatory in nine European Union countries.[2] Examining secondary schools, a 2014 investigation revealed the number of students studying English as a foreign language to be 94.1%, French 23.0%, Spanish 19.1% and German 18.9%, and the number of secondary students learning more than one foreign language was 51.2%. Even though the US is far behind Europe in terms of access to foreign language education, in a speech to the Foreign Language Summit held at the University of Maryland, the Secretary of Education, Arne Duncan, reported that in 2010 a quarter of all US elementary schools offered a foreign language (U.S. Department of Education, 2010).

It is clear that foreign language education is the first point of contact for many people. Therefore, determining how the amount of L2 exposure affects speech acquisition in formal foreign language instruction is vital to serving this large, yet largely overlooked, population. As mentioned earlier, the benefits of a young AOA for immigrants is not paralleled for those foreign language students who begin formal instruction early – but with the growing importance being placed on foreign language education and the emerging popularity of bilingual programs, it is important to know the different effects these two kinds of foreign language education have on the students' performance in different aspects of L2 acquisition. What this dissertation intends to examine is the perceptual and production ability of a group of elementary school students who share the same native language (Spanish) and the same age of first exposure (first grade) to English, but who differ in the hours of L2 English instruction provided to them each week.

[2] Malta, Cyprus, Spain, Italy, Austria, Norway, Macedonia, Liechtenstein and Luxembourg all require foreign language instruction in primary school. Of these nine, English is the obligatory language of all but Luxembourg, which requires its students to learn German.

Foreign language education in Spain has become a priority for the Spanish government in recent years. In 2004, the Spanish Ministry of Education started a bilingual school initiative that has expanded to include over 500 elementary and secondary schools in the autonomous community of Madrid alone. The program is held in high regard and has the goal of providing their students with the tools needed to reach linguistic competence that will contribute to their eventual professional placement in an intercultural, globalized environment (Ministerio de Educación, Cultura, y Deporte, 2016). To that extent, the bilingual school teachers, in addition to earning their teaching degree with a specialization in English as a L2, must also pass the Test of Linguistic Qualification (*Examen de Habilidad Lingüística*) to show that they have the necessary language skills to excel in a bilingual environment (Consejería de Educación, Juventud y Deporte, 2016).

As an additional way of providing their students opportunities to speak English, each school in the bilingual initiative also has *language auxiliaries* – native English speakers who act as teaching assistants. These L1 English speakers come from countries all over the world and must be in a classroom with the students from 12-16 hours per week. This interaction exposes the students to native English speakers and allows them to use what they are learning in class with an expert of the language. This interaction also helps equip them for the mandatory external assessment of their language competency.

To prepare their students to interact within the global community, as well as to validate the quality of each school's bilingual program, language examiners from Trinity College London assess the linguistic abilities of each second, fourth, and sixth grader. The examiners evaluate the oral proficiency of these students based on their ability to communicate and their control of specific, pre-determined aspects of English, like language functions (*greetings, asking questions,* *expressing certainty and requesting opinions etc.*), grammar, lexis, and phonology (Trinity College London, 2009). Touching on this last point, the examiners look for, among other things, correct pronunciation. As many L2 English teachers and L2 English learners will tell you, certain English sounds are more difficult to perceive and produce than others.

1.7 AMERICAN ENGLISH AND SPANISH VOWELS

When compared with the sound systems of the world, American English, referred to after this simply as *English*, has a reasonably dense vowel system with 12 monophthong vowels (Maddieson, 1984; Odden, 2005). The present investigation focuses on how L1 Spanish students perceive and produce the five English front vowels /i, ɪ, e, ɛ, æ/, all of which are unrounded. English vowels are often divided into five categories based on height: high, mid-high, mid, mid-low, and low. Native American English (L1 English) speakers distinguish vowels principally by attending to spectral cues, though duration has been shown to be a secondary cue in the discrimination of tense-lax pairs (Hillenbrand, Clark, & Houde, 2000).

The primary acoustic feature of vowels is the location of three formants, or bands of energy, that are present in the acoustic signal (Ladefoged & Johnson, 2011). L1 English speakers discriminate between different vowels based on the arrangement of these three formants. Although durational differences are present between tense-lax vowel pairings (Bohn & Flege, 1990), vowel length is not a distinctive phonological feature of American English vowels.

Spanish, in contrast, contains only five vowels, two of which are front (/i/ and /e/), one central (/a/) and two back (/o/ and /u/). The large difference between the vowel inventory of Spanish and English often leads to complex mappings when L1 Spanish participants compare

Spanish and English phones during assimilation tasks (Bradlow, 1995). Unlike English, Spanish does not contrast tense and lax vowels, though there has been evidence of L1 Spanish speakers laxing /e/ to /ɛ/ in closed syllables that do not end in /s/ or /z/ (Dalbor, 1980).

As mentioned above, the American English vowel space is divided into five categories based on height (high, mid-high, mid, mid-low, and low) and three based on the lateral movement of the tongue (front, central, back). Even though the Spanish vowel system also categorizes sounds into front, central, and back groups, it has only three vertical sections – high, mid, and low. The small differences between English tense-lax vowels of neighboring height categories (ex. high vowel /i/ versus mid-high vowel /ɪ/) often result in difficulty for many L1-Spanish learners of English (Flege, Bohn & Jang, 1997; Gulinello, 2010). An inability to discriminate between the tense-lax pairs based on spectral differences can result in the overuse of vowel duration as the defining feature that separates /i/ from /ɪ/ and /ɛ/ from /æ/ (Barrios, Jiang, & Idsardi, 2016; Bohn & Flege, 1990; Cebrian, 2006; Cebrian, 2007; Flege & Bohn, 1989).

1.7.1 Perceptual and acoustic similarities of L1 Spanish-L2 English vowels

To obtain information concerning cross-language perceptual similarities between Spanish and English vowels, we will look at how a group of native Spanish-speaking adults from Flege (1991) categorized American English vowels (/i ɪ ɛ æ/) onto their L1 sound system. These L1 Spanish speaking adults were initially divided into two groups based on their differing levels experience with English, but results indicated that there was no significant difference in the way the members of these two groups performed. Results from both groups are conflated here, as in Flege (1991).

The native Spanish speakers from Flege (1991) categorized four English front vowels /i, ɪ, ɛ, æ/ in terms of the vowels found in the L1 /i e a/. In Figure 2, we see the frequency with which these L2 English vowels were assimilated onto specific L1 categories.

An acoustical analysis measuring the F1 and F2 values of the English and Spanish phoneme /i/ (Mendez, 1982) found no significant difference between the productions of /i/ in either language. However, in an acoustic analysis of her own, Bradlow (1995) found that English vowels were generally produced with a higher F2, indicating that they were more fronted that the Spanish counterparts. Moreover, English /e/ is often a diphthong [eᴵ] unlike the Spanish monophthong [e]. But if the English learners perceive the English and Spanish phones to be more or less identical, the L1-Spanish/L2-English speakers will maintain their pronunciation regardless of the language they are speaking (Flege, 1995; Cebrian, 2007).

In Figure 2, we see that the Spanish speakers assimilated 84% of the English /i/ presentations to Spanish /i/ suggesting that Spanish and English /i/ are not just acoustically similar, but perceptually similar. Conversely, both English /ɪ/ and /ɛ/ are assimilated to three native Spanish categories and do not fit well into any Spanish category. According to SLM and PAM-L2 this may mean that there is not an equivalent sound in Spanish, which means that there is strong potential for L2 English learners to establish mental representations for these sounds. Support for this is most clearly seen for English /ɪ/, which is more evenly assimilated onto Spanish /i/, /e/, and the *none* category. The inconsistent, variable mapping of these two vowels may mean the L2 English learners perceive these as *new* categories.

Finally, English /æ/ was well assimilated on Spanish /a/ at 82%, although *none* was selected at a rate significantly higher than zero. This suggests a relatively strong perceptual similarity between English /æ/ and Spanish /a/ even though, acoustically, there are differences.

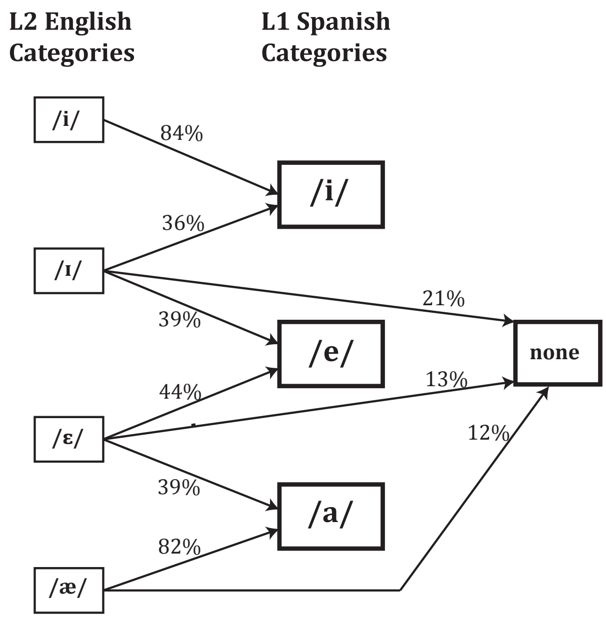


Figure 2. *L2 English-L1 Spanish Vowel Assimilation.* The illustration displays mapping patterns of L2 English vowels onto the Spanish sound system by native Spanish speakers. This figure was created with results from Flege (1991).

1.8 METHODOLOGICAL APPROACHES

One popular method used to examine L2 perception is done through categorical discrimination tasks (e.g., Cebrian, 2006). Researchers use these tasks, in part, to see if language learners have established new L2 phonetic categories (Cebrian, 2008; Fabra & Romero, 2012; Flege & MacKay, 2004;). During an AX Categorical Discrimination task, two L2 sounds or words are presented to the participant and they must determine whether the two target sounds are instances of the same sound (within-category vowel pairing) or if they represent two different sounds (between-category vowel pairing). The establishment of new L2 phonetic categories increases language learners' sensitivity to between-category differences while, at the same time, decreasing sensitivity to within-category differences that can occur between speakers (e.g., pitch, tone). It is expected that language learners will have greater difficulty discriminating the tokens of an L2 contrast if they do not have the necessary L2 categories formed (Flege, 1995).

To test the participants' production of L2 English front vowels, a picture-naming task is used. Orthography has been shown to affect pronunciation in the L2 (Bassetti, 2006) but using pictures to elicit pronunciation removes the potential influence. There is a one-to-one grapheme-to-phoneme mapping in Spanish unlike the many-to-one mapping found in English. When shared graphemes (e.g., English and Spanish 'i') map to different pronunciations in the L1 (*pito* [pito]) and L2 (*bit* [bɪt]) this can cause pronunciation difficulties.

1.9 PRESENT STUDY

The purpose of the present dissertation is to examine the effect that varying amounts of L2 exposure may play on the perception and production of English front vowels by L1 Spanish-speaking children. In the following section (Section 2) I will lay out the details of the perception experiment, introduce the participants and the stimuli. Then, I will present the specific research questions associated with L2 perception and provide insight into the predicted results. After that, I will present the results from the perception study and will finish Section 2 with a discussion of those results. Section 3 will follow the same format as Section 2, but will focus on the production task. Finally there will be a general discussion and conclusion section (Section 4) that explains the implications of the results of both tasks and what it says about the current interpretation of the interaction of perception and production.

2.0 EXPERIMENT I – PERCEPTION

2.1 METHODS I (CHILDREN)

2.1.1 Participants

One hundred twenty four school-age children from Colmenar Viejo, Madrid (Spain) participated. They were students attending either a bilingual or non-bilingual school and enrolled in either second, fourth, or sixth grade (Table 1). To be eligible to participate, the students from the bilingual school needed to meet the following criteria: be a speaker of L1-Spanish/L2-English, speak Spanish at least 85% of the time at home, not receive additional language instruction outside of class, and they must have enrolled in the bilingual school by the time the bilingual program began in first grade. The requirements for the students from the non-bilingual schools were similar, with two noticeable exceptions – they must have strictly attended a non-bilingual school starting in first grade and, due to the high demand for private English classes in Spain, they were allowed to have up to one year of English class outside school for up to two hours per week.

Table 1. *Distribution of Native Spanish-Speaking Participants.* The distribution of NS students who participated in the study by school type and grade level.

Grade level	Bilingual School	Non-Bilingual School	Total
2nd	23	17	40
4th	19	25	44
6th	22	18	40
Total	64	60	124

The students at this particular bilingual school receive instruction in English in three classes: art, science, and English language, which constitute roughly 40% of their schedule, or 14 hours per week. The classroom teachers are L1-Spanish/L2-English speakers who passed an English proficiency test organized by the Spanish government, giving them certification to teach in a bilingual school. Additionally, the classroom teachers are paired with native English speakers who act as teaching assistants, often leading small group activities and periodically teaching lessons. The teaching assistants working at the time the study was conducted were from the United States (California, Washington State and Missouri).

The students from the non-bilingual schools receive instruction in English only during English language class, which meets approximately once per day resulting in 3-5 hours of exposure to English each week. The teachers at this school do not need to pass the same government-organized proficiency test mentioned above and do not have any native English-speaking assistants.

2.1.2 Stimuli

Two monolingual English speakers from the Pittsburgh area provided the stimuli for the AX Categorical Discrimination task. Both participants were male, 28 years old and grew up within 30 minutes of Pittsburgh proper. They produced words containing the five English front vowels in a /b_t/ context contained in the following carrier phrase *"_________. Say _________ to your mother"*[3] (Table 2). The words were recorded in a quiet room directly onto the principle investigator's (PI) computer using a USB microphone. Two productions of each word from each speaker were chosen to be part of the final task.

The target environment during the word selection process was bVt for two reasons: Levy (2009) found that responses to an assimilation task were more consistent when the vowels appeared in bilabial contexts than alveolar; and second, bVC has already been utilized in multiple L2 vowel perception studies (Mayr and Escudero, 2010; Flege & MacKay, 2004; Cebrian, 2008) allowing for potential comparisons to be made more easily. The average F1 and F2 of the native English speakers can be found in Table 2. The average F1 and F2 measurements of English vowels from Bradlow (1995) have been included in the same table and act as a point of comparison.

Table 2. *English Stimulus List and Acoustic Measurements Across Speakers.* Five English vowels were placed into a single CVC context (/b_t/) for the AX Categorical Discrimination task.

Vowel	Context	Gloss	Present Study		Bradlow (1995)	
			Mean F1 Hertz (SD)	Mean F2 Hertz (SD)	Mean F1 Hertz (SD)	Mean F2 Hertz (SD)
/i/	/bit/	*beat*	293.5 (13)	2444 (174)	268 (20)	2392 (239)
/ɪ/	/bɪt/	*bit*	445.5 (25)	1974 (119)	463 (34)	1995 (199)
/e/	/beᶦt/	*bait*	421 (9)	2168 (172)	430 (45)	2200 (168)
/ɛ/	/bɛt/	*bet*	643 (31)	1757 (143)	635 (53)	1796 (149)
/æ/	/bæt/	*bat*	829 (42)	1600.5 (134)	777 (81)	1738 (177)

The stimuli taken from the NE speakers were parsed using Praat and presented to the participants via the experiment builder SuperLab 5.0 using Tritton AX 180 headphones. Each

---

[3] Due to individual differences in aspiration of the final /t/ in the second production of the word embedded in the phrase, this production could not be used. Therefore, for consistency in production, the first utterance of the word in the carrier phrase was presented to the participants.

utterance from the first NE speaker was contrasted against every utterance from the second NE speaker. The sequence was then reversed to prevent possible order effects. This resulted in 200 stimulus word pairings. The contrasts were preceded by 800ms of silence with 500ms of silence between the individual tokens of the pairing.

2.2 PROCEDURE (CHILDREN)

The study was conducted using a MacBook Pro laptop computer in a private room provided by the schools in two eight- to ten-minute sessions. Before beginning the task, the PI read the instructions aloud in English and the participants read along on the computer screen. When needed, clarification questions were addressed in Spanish.

Immediately after the instructions, the participants completed the first of two practice sessions. In order to confirm that the students understood the task, they listened to six minimal pair contrasts in Spanish and were instructed to determine if what they heard were instances of the same word (ex. *pito-pito*) or different words (ex. *pito-pato*). All participants responded accurately to five or six of the Spanish word pairings. Next, the participants continued to the second practice session containing a sample of the L2 English words from the main portion of the task. Subjects were required to respond to five out of six vowel pairings correctly before continuing on to the main portion of the task. If a subject was unable to achieve this on their first try, they repeated the practice session one more time. After passing the two practice sessions, the real task began.

At this time the auditory stimuli (200 word pairs) were randomly presented to the participants. After hearing an English word-pairing (ex. *bait-bet*), they were instructed to click on one of the two circles displayed on the screen that said *the same word* or *different words*. Once their response was recorded, the next contrast began. This continued over two sessions of 100 word pairs each until they finished the 200 word pairings at the end of the second session. To prevent fatigue, the participants were given breaks after every fifteen responses.

2.3 METHODS II (ADULTS)

2.3.1 Introduction

In a previous study (Jeske, 2012), twelve native Spanish speakers (5 male, 7 female) from the Department of Hispanic Language and Literature at the University of Pittsburgh participated in an AX Categorical Discrimination task very similar to the one the Spanish children participated in. This study is detailed below and their overall score from the perception task is compared against those of the bilingual and non-bilingual school students in the results section. The purpose of this comparison is to determine if the adults perform significantly differently from the children, thereby supporting or contradicting the notion that L2 perception becomes more native-like with experience and remains accessible to even late learners (Flege, 1995).

2.3.2 Participants

The L1 Spanish-speaking adults who participated in the study from 2012, came from one of nine Spanish-speaking countries[4], with varying ages (m=31.25 years, range: 26-41), ages of first exposure to English (m=12.42 years, range: 4-20 years), lengths of residence (m=3.7 years, range: 7 months-120 months), and years studying English (m=14.3 years, range: 2-20 years).

Each participant had passed an English proficiency exam administered by the university before they were given clearance to work as a graduate teaching assistant. On a language questionnaire filled out before the first task, everyone indicated they had normal hearing, eyesight, and that they did not attended an English or Spanish-English bilingual school. Some participants had experience with French as a foreign language but indicated that their proficiency in English was higher. All activities were completed in the Phonetics Lab on campus, after which each participant was paid $15.

Table 3. *Demographics of Adult L1 Spanish Participants.* Demographic information about the adult participants used in the perception study (Jeske, 2012).

Partic. #	Gender	Age	Country	Age of First Exposure	Years learning English	Months in the US
1	F	28	Spain	12	16	42
2	M	41	Bolivia	39	2	7
3	F	30	Nicaragua	14	9	60
4	F	26	Bolivia	11	6	7
5	M	30	Mexico	20	10	120
6	F	35	Bolivia	8	16	42
8	F	31	Ecuador	4	12	30
9	F	30	Colombia	10	20	84
10	F	28	Uruguay	7	15	17
11	M	36	Peru	8	25	57
12	M	31	Mexico	10	21	36
AVG	**F, 5M**	**1.45**	**--**	**13.0 years**	**13.8 years**	**45.64** (3.8 years)

[4] Nationality of L1 Spanish participants was not of importance for eligibility due to the relative stability of vowel production across dialects (Hualde, 2005; Mendez, 1982; Morrison & Escudero, 2007)

2.3.3 Stimuli

Four L1 English speakers (2 male, 2 female) from the Linguistics graduate program at the University of Pittsburgh supplied the stimuli for the adults' perception study. They recorded the full list of words directly onto a PowerBook using a USB microphone in the Robert Henderson Language Media Center at the University of Pittsburgh. Each speaker read a randomized list of words containing ten English vowels; each word was produced three times. A male from Flint, Michigan, and a female from Buffalo, New York provided the majority of the stimuli to the L1 Spanish speakers because of the clarity of their speech. In the chance that the PI perceived one of their vowels as atypical, it was replaced with a production of the same word from the other native English speaker of the same gender.[5]

The perceptual stimuli presented to the L1 Spanish adults consisted of consonant-vowel-consonant (CVC) English words. Because surrounding consonants have been shown to affect the perception of vowels (Bohn & Steinlen, 2003), every attempt was made to keep the onset and coda constant while still producing a word that appeared on the British National Corpus' (BNC) list of the 2,000 most commonly used words. Real, frequent words were chosen to maximize the chance that participants would have a corresponding entry in their lexicon for each word. The vowels were contained within the bVt context except for *bait*, which was replaced with *date* to meet the word frequency requirement.

[5] Of the words produced by a female, 88% were recorded by the woman from Buffalo, New York (12% by a woman from Nashville, TN). Of the words produced by a male, 85% were recorded by the man from Flint, MI (the other 15% were produced by a male from Columbus, OH.)

Table 4. *AX Categorical Discrimination Task Stimuli.* All the stimuli used in the original perception and production tasks with the L1 Spanish speakers. The current study focuses on English /i/ /ɪ/ /e/ /ɛ/ and /æ/

	English		Spanish	
Vowel	Context 1	Gloss	Context 1	Gloss
i	bit	*beat*	pito	*Pito*
ɪ	bɪt	*bit*		
e	deɪt	*date*	peto	*Peto*
ɛ	bɛt	*bet*		
æ	bæt	*bat*		
ɑ	dɑt	*dot*	pato	*Pato*
ɔ	bɔt	*bought*		
o	boʊt	*boat*	poto	*Poto*
ʊ	bʊk	*book*		
u	but	*boot*	puto	*Puto*

2.4 PROCEDURE (ADULTS)

The testing for this task was completed in the phonetics lab at the University of Pittsburgh. The auditory stimuli that were presented through headphones followed the bVC pattern found in Table 5. Participants heard two words and were told to pay special attention to the vowels. They were asked to press the green button on the response pad if the words contained the same vowel sound and the red button if the words contained different vowel sounds. Each contrast was preceded by 800 ms of silence with 500 ms of silence between the individual tokens of each pairing. After five practice trials the participants were given the opportunity to ask questions or repeat the practice trial before continuing on; no one asked any questions or repeated the practice trial. No feedback was given during the activity.

A total of 200 discrimination contrasts were randomly presented to the participants. Each of the ten English vowels was presented in a contrast with the other nine vowels and itself (10 vowels x 10 vowels = 100 vowels). To prevent any order effects, each contrast was presented in the opposite sequence as well (ex. male *bat* – female *bought*; female *bought* – male *bat*) [100 vowel contrasts x 2 sequences = 200 vowel contrasts].

2.5 RESEARCH QUESTIONS AND HYPOTHESES

Q1. Will School, Grade, or an interaction between school and grade have an effect on the participants' overall perception in the AX Categorical Discrimination task?

In the present study, the amount and quality of exposure to L2 English is the primary difference between the schools being analyzed. The students from the bilingual school have 14 hours of English per week and have an L1 English speaker in the classroom for up to nine of those hours. Conversely, the students from the non-bilingual schools have English class between three and five hours a week without a native English speaker. Due to the bilingual school's students' increased contact with L2 English, I expect these participants to perform significantly better than the non-bilingual students on the AX Categorical Discrimination task.

However, if the amount of exposure to English from both schools is enough to change L2 perception, the perceptual accuracy scores from all participants may increase with age as the participants gain more experience with the second language (Werker and Tees, 1984). Because the participants from the bilingual school are receiving more exposure per week, as they move from second, to fourth, to sixth grade they may become more accurate overall when compared to the non-bilingual school participants. On the other hand, if one or both of the schools does not offer enough high-quality input, perceptual scores may decrease with age due to the increased influence of their strengthening L1 sound system and its inhibitory effects on their ability to decipher non-native phones.

When the participants are first introduced to English with any regularity in the schools, they are in first grade (5-6 years old). It becomes increasingly difficult, if not impossible, to perceive and produce L2 sounds (perform phonological tasks) in a native-like fashion for those first exposed to the L2 after age six (Granena and Long, 2013; Meisel, 2009). Therefore, in the present study, if the students' first exposure to English at age five is insufficient, the second graders (6-7 year olds) may have more accurate overall perception scores than the fourth and sixth graders because their L1 sound system is less rigid.

Q2: Does additional weekly exposure to L2 English improve participants' ability to accurately perceive individual between-category and within-category vowel pairings?

The same arguments made regarding how additional exposure could impact overall perception scores (Q1) remain relevant in the discussion of the participants' performance on individual vowels. If the bilingual school provides significantly more exposure to L2 English for their students than the non-bilingual school, we should see an improvement in performance on the discrimination task.

Q3. Does an greater length of formal instruction improve participants' accuracy in perceiving individual between- and within-category vowel pairings?

Perception difficulties in an L2 often vary according to learners' level of familiarity and amount of use; with increased exposure and use, perception/production generally improve (Flege et al., 1995; Ingram & Park, 1997; Tsukada, Birdsong, Bialystok, Mack, Sung & Flege 2005).

PAM-L2 claims that L2 phonetic and phonological attunements occur during the early stages of language learning in an immersion setting (Best & Tyler, 2007). If phonological reattunement can occur without immersion, as suggested in Bundgaard-Nielson et al. (2011) with adult NJ speakers, or if 14 hours of English per week constitutes immersion, it is possible that some of the bilingual elementary school students could have already gone through this reorganization of L2 phones. Studies examining perceptual change in an L2 (e.g. Aoyama et al., 2004; Flege & Liu, 2001) have noted that the biggest changes to L2 vowel perception for adults immersed in an L2 environment have occurred within the first 6-12 months and then level off. If this timeline can be generalized to the students of the bilingual program, this could suggest that reattunement to the L2 occurs during the first year of the bilingual program – in first grade. If there is no significant improvement across grades within a school, this may explain why.

Q4. Will certain school-grade groups perceive specific vowel pairings significantly differently from the other school-grade groups?

As students advance through the non-bilingual school, they will most likely not see significant improvement in their perception scores because the amount of L2 exposure they have is minimal (Larson-Hall, 2008; White and Genesee, 1996). However, if 14 hours of English per week is comparable to the exposure child immigrants receive in a target language environment, we may see the bilingual school participants' perception accuracy improve with age.

Q5. Does the order that the English stimuli are presented to the participants affect their accuracy of that vowel pairing?

Polka and Werker (1994) found that L1 English infants exposed to two pairs of German vowels, /u/-/y/ and /ʊ/-/ʏ/, perceived a difference more easily when they were presented as /y-u/ and /ʏ-ʊ/ instead of in the opposite order. To determine if this asymmetry was due to the infants' familiarity with /u/ and /ʊ/ through their L1 (English), Polka and Bohn (1996) examined the discrimination ability of infants from two language groups: English and German. They presented both groups with a contrast from each language and found that the infants exhibited the same perceptual asymmetry in both German and English, regardless of their L1 - the perceptual asymmetry occurred in the infants' L1 and L2.

Polka and Bohn (2011) expanded on their previous study with infants and found that adults showed signs of perceptual asymmetries too. These repetitive observations surrounding perceptual asymmetries lead to the creation of the Natural Referent Vowel (NRV). The NRV claims that there is a perceptual bias toward vowels that are in the periphery of the F1/F2 vowel space and that peripheral vowels act as an anchor (Polka & Bohn, 2003). When presented with a vowel contrast, people will have an easier time perceiving a difference when the peripheral vowel is presented after the vowel that is located more centrally in the vowel space. For example, the difference between /u/ and /ʊ/ will be easier to hear when the vowels are presented /ʊ-u/ instead of /u- ʊ/.

Because both adults and infants have exhibited perceptual asymmetries while performing L2 and L1 vowel discrimination tasks, and because this asymmetry has been seen with L1 Spanish/L2 English adults, it is probable that the children from Madrid will also show signs of perceptual asymmetry, regardless of age and school type. Using the guideline described in the NRV, we can predict the presentation orders of L2 English front vowels that will result in better discriminability: /ɪ-i/, /ɪ-e/, and /ɛ-æ/.

Q6. Do the adult L1 Spanish speakers exhibit evidence of perceptual learning when their overall performance on the AX Categorical Discrimination task is compared to those of the participants from the bilingual and non-bilingual school?

The adults who participated in the first experiment did not attend a bilingual school growing up, but they had been studying English for many years (*M*=16.25 years) and had been living in an L2 environment for numerous years (*M*= 3.6 years), as well. An analysis comparing overall perception accuracy between the L1 Spanish-speaking adults and children can help us determine if adults can overcome the possible disadvantage of never having attended a bilingual school.

Both the PAM-L2 (Best & Tyler, 2007) and SLM (Flege, 1995) have stated that perceptual learning does not disappear in adulthood and can still be accessed under the right environments: increased exposure to and use of the L2 (e.g. Flege & MacKay, 2004). If the adults have better overall perception than the children who did not attend a bilingual school either, this will add support to the claim that with continued exposure and practice adults can significantly improve their L2 perception accuracy from when they were young. However, if the non-bilingual school children and adults have comparable scores resulting in significant underperformance when compared to bilingual school children, this could add support for the implementation for extensive and early bilingual education.

2.6 RESULTS

Q1. Will School, Grade, or an interaction between school and grade have an effect of the participants' overall perception in the AX Categorical Discrimination task?

To test for possible effects of school, grade and an interaction between school and grade on overall perception scores, a 2x3 ANOVA was performed (2 school x 3 grades), the results of which can be found in Table 1. For overall perception scores by school, grade, and individual grades within each school, see Table 5.

Table 5. *Overall Perception Accuracy on the Categorical Discrimination task.* Results from a 2x3 ANOVA show a significant effect for school (*p*=.024) but no significant effect for grade (*p*=.857) or school*grade (*p*=.568).

	Type III Sum of Sq.	df	Mean Square	F	Sig.	Partial Eta Sq.	Noncent. Param.	Obs. Power
Intercept	73.609	1	73.609	10271.881	.000	.989	10271.881	1.0
School	.037	1	.037	5.230	.024*	.042	5.230	.621
Grade	.002	2	.001	.155	.857	.003	.310	.073
School*Grade	.008	2	.004	.568	.568	.010	1.136	.142
Error	.846	118	.007					

Effect for school on overall perception accuracy score

The students from the bilingual school reliably outperformed the students from the non-bilingual school in the AX Categorical Discrimination task. Out of 200 possible points, the bilingual school participants earned an average of 159.13 points (79.56%) and the non-bilingual school participants earned average of 151.93 points (75.97%). This significant, main effect of school type indicates that, overall, the students from the bilingual school were able to perceive the English vowel word pairings significantly more accurately than the students from the non-bilingual school [$F(1, 118) = 5.23$, $p=.024$]. However, the partial eta squared test revealed only a small effect size of $\eta^2=.042$.

Table 6. *Discrimination Scores for Bilingual and Non-Bilingual Schools.* Mean accuracy scores and percentages for participants at all levels from both the bilingual and non-bilingual school.

	Grade	Overall Mean (out of 200)	% Accuracy	School Mean (% correct)
Bilingual	2nd	158.3	79.15%	159.13 (79.56%)
	4th	158.05	79.03%	
	6th	160.91	80.46%	
Non-Bilingual	2nd	155.18	77.59%	151.93 (75.97%)
	4th	151.36	75.68%	
	6th	149.67	74.84%	
Combined	2nd	156.98	78.49%	All Partic.
	4th	154.25	77.13%	155.65 (77.83%)
	6th	155.85	77.93%	

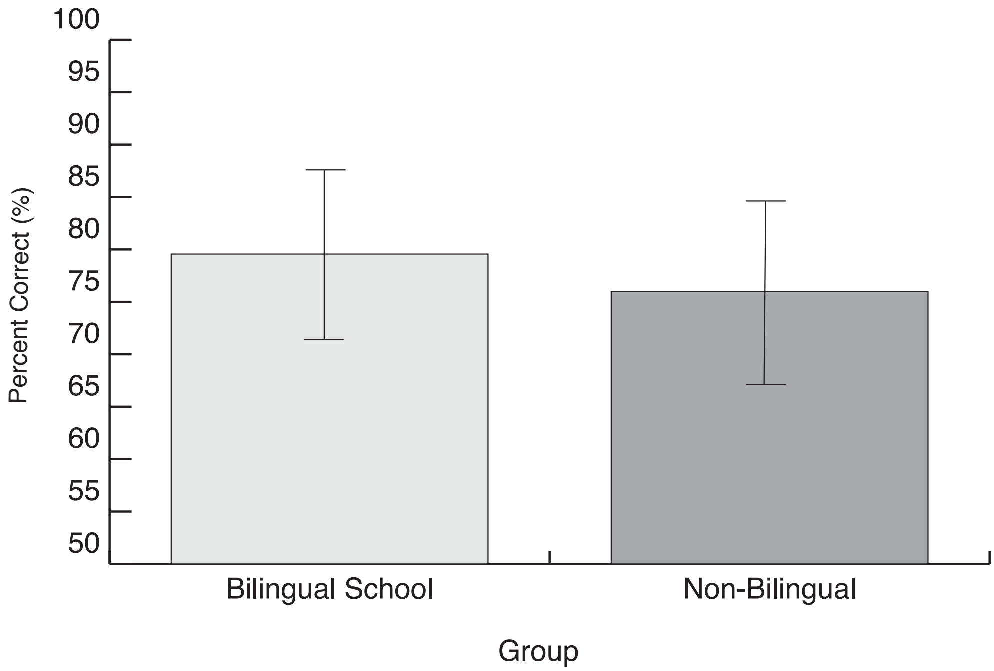


Figure 3. *Comparison of Overall Perception Scores from Discrimination Task by School.* Displays the overall mean score (%) of the AX Categorical Discrimination task that was achieved by each participant from the bilingual and non-bilingual school (*p*=.024).

Effect for grade on overall perception accuracy score

In the 2x3 ANOVA whose results are presented in Table 5, we see that there is no significant main effect of grade [F(2,118)=.155, p=.857] signifying that no individual grade performed significantly differently from the other two grades. Each combined grade's overall performance on the perception task differed from the other two grades by a maximum of 1.37%, or 2.73 points. The bar graph below (Figure 3) illustrates the similarity of performance across grades.

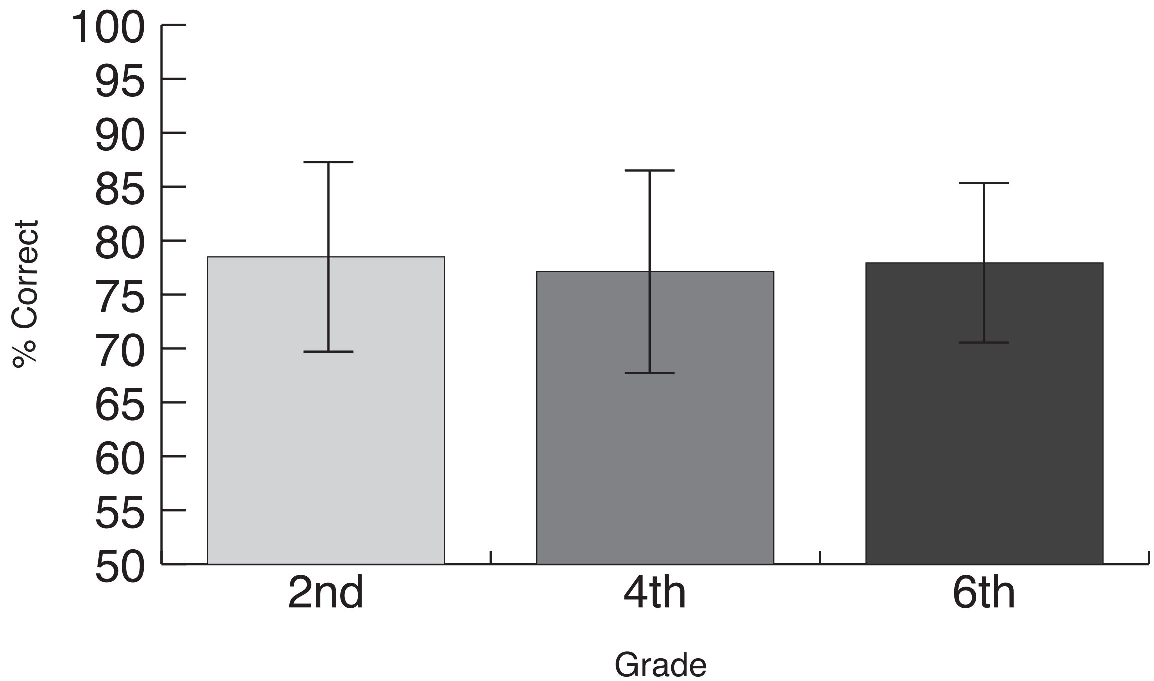


Figure 4. *Overall Accuracy on Discrimination Task by Grade.* When scores across schools are combined, there was no significant difference in any grade's performance on the perception task (*p*=.857)

School-grade Interaction on overall perception accuracy score

As the students from the bilingual school progressed through their English program, their scores from one grade level to the next remained more or less constant or slightly increased. From second to fourth grade, the bilingual students' average score dropped from 158.3 points to 158.05 (out of 200). In contrast, the sixth graders earned an average of 160.91 points – slightly higher than their younger peers. The overall accuracy scores achieved by the non-bilingual school participants, on the other hand, decreased with age: (2^{nd}: 155.18 → 4th: 151.36 → 6th: 149.67). However, regardless of the divergent patterns seen in Figure 5, no significant interaction between school type and grade was found [F(2, 118)=.114, p=.568], indicating that no individual school-grade group performed significantly better or worse than any other, with regards to overall perceptual ability. Similarity in the grades' performance is seen in Figure 4.

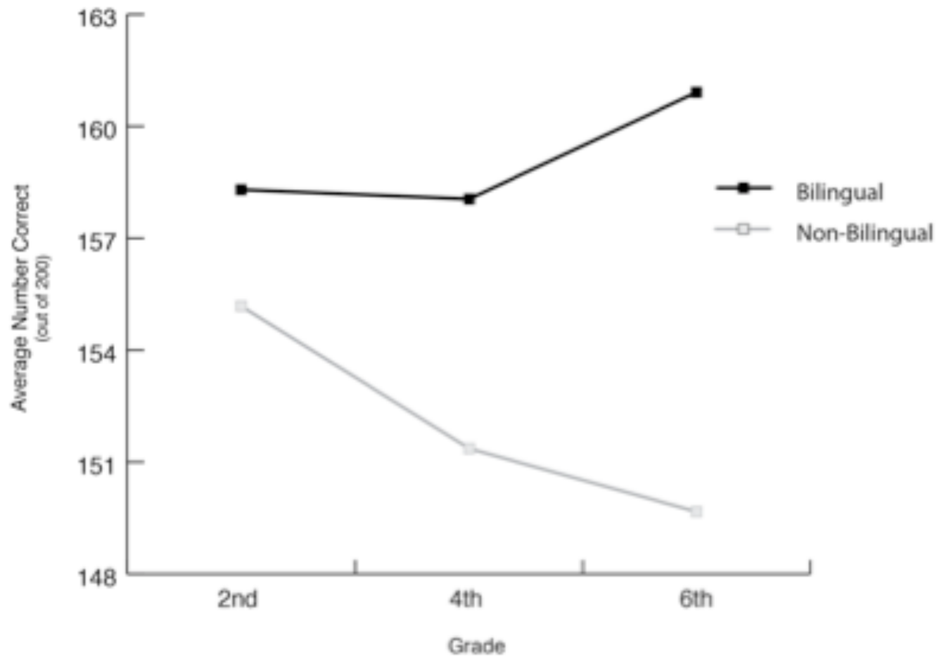


Figure 5. *Average Accuracy Scores of Discrimination Task.* Displays the disparate patterns observed between the two schools in terms of overall perceptual accuracy in the discrimination task.

Q2: Does increased weekly exposure to L2 English improve participants' ability to accurately perceive individual between-category and within-category vowel pairings?

Table 7. *Perception Accuracy of Individual Contrast by School.* (Within-Subjects Effects) There is a main effect for contrast and school - participants from the schools performed significantly differently on at least one vowel pairing (**p*<.05, ****p*<.001)

	Type III Sum of Sq.	df	Mean Sq.	F	Sig.	Part. ETA sq.	Non-Cent Par.	Observed Power
Contrast	41.368	.884	0.650	81.255	.000***	.400	315.610	1.000
School	1.503	.884	.387	2.953	.021*	.024	11.470	.782
Error	62.112	1708	.036					

Knowing that the bilingual school students performed significantly better overall than their non-bilingual school counterparts on the perception task, a repeated measures ANOVA was performed to investigate the effect of school on the perception of individual vowel contrasts. The repeated measures ANOVA (Table 7) determined that there was indeed a significant effect of school [F(3.884,1708)=2.953, p=.021]. Therefore, Duncan's Multiple Range Test (Duncan's

MRT) was performed comparing the mean accuracy scores of the students at each school (between subjects) for the 15 vowel pairings (within subject) to determine which were perceived significantly differently between the schools. The results for seven vowel pairings under examination are displayed in Tables 8 and 9.

Table 8. *Accuracy Score (%) per Contrast by School.* Comparison of mean scores by school. Significance determined through Duncan's Multiple Range Test (*p<.05, **p<.01).

Contrast	i-i	ɪ-ɪ	e-e*	ɛ-ɛ*	æ-æ**	i-ɪ	ɛ-æ**
Bilingual	93.55	88.87	72.07	79.30	85.35	71.29	38.87
Non-Bilingual	87.08	82.92	63.33	69.58	72.71	70.94	51.77

Comparing means through repeated t-tests increases the chance of falsely detecting an effect (Type I error) unless the alpha is divided by the number of pairwise comparisons performed. For the present analysis, the alpha of .05 would have to be divided by the number of comparisons performed, resulting in a new alpha of .007 (.05/7). This method attributed to Bonferroni is often considered to be too extreme as it fails to detect effects that are actually present (Type II error) (Larson-Hall, 2010).

Duncan's MRT, on the other hand, allows one to make a series of layered pairwise comparisons while maintaining an alpha of .05 or .01. This is done by ordering the mean score for each group being compared from smallest to largest and adjusting the value of the critical difference based on the proximity of the two means being compared – adjacent means have a smaller critical value than two means ranked far apart (Bruning & Kintz, 1987).

The Duncan MRT used in Table 9 compared the bilingual and non-bilingual schools' performance on individual vowel pairings. To do this, the mean accuracy scores of each vowel pairing from each school was ordered from lowest to highest, resulting in the ascending enumeration of 30 mean accuracy scores (15 vowel pairings x 2 schools = 30 means)[6]. Next, the difference between every mean accuracy score was calculated, beginning with neighboring means (ex. determining the difference between the lowest mean and the second lowest mean) and increasing the range of comparison until the difference between the highest and lowest mean was found.

Table 9. *Duncan's MRT - Mean Difference per Contrast by School.* Results show no significant difference in performance for /i-i/, /ɪ-ɪ/ and /i-ɪ/. The bilingual school outperformed on /e-e/, /ɛ-ɛ/ and /æ-æ/ but the non-bilingual school outperformed on the /ɛ-æ/ vowel pairing (**p*=.05, ***p*=.01).

α = Error df Error Mean Square Number of Means Critical Range		*p*=05, *p*=.01 1708 (infinity) .036 30 [Range in brackets below each difference]					
Group (% Correct)	Bilingual /i-i/ (93.55%)	Bilingual /ɪ-ɪ/ (88.87%)	Bilingual /e-e/ (72.07%)	Bilingual /ɛ-ɛ/ (79.30%)	Bilingual /æ-æ/ (85.35%)	Bilingual /i-ɪ/ (71.29%)	Bilingual /ɛ-æ/ (38.87%)
Non-Biling. /i-i/ (87.08%)	6.47% [9]						
Non-Biling. /ɪ-ɪ/ (82.92%)		5.95% [5]					
Non-Biling. /e-e/ (63.33%)			8.74%* [6]				
Non-Biling. /ɛ-ɛ/ (69.58%)				9.72%* [9]			
Non-Biling. /æ-æ/ (72.71%)					12.64%** [6]		
Non-Biling. /i-ɪ/ (70.94%)						.35% [2]	
Non-Biling. /ɛ-æ/ (51.77%)							12.9%** [3]

Table 9 focuses on each school's accuracy score of seven contrasts: /i-i/, /ɪ-ɪ/, /e-e/, /ɛ-ɛ/, /æ-æ/, /i-ɪ/ and /ɛ-æ/. The scores for the bilingual school appear across the top and the non-bilingual school's scores appear in the left-most column. The differences in the schools' two

[6] Even though the mean scores for only seven vowel pairings were being compared, Duncan's MRT allows the comparison of all scores involved – in this case 15.

scores appears in the cells at the intersection of the row-column of the same vowel pairing; it is shown as a percent. Using the /i-i/ vowel pairing as an example, we see the bilingual school perceived this contrast correctly 93.55% of the time and the non-bilingual school did so 87.08% of the time. In the cell at the intersection of this row and column we see that the mean difference in performance is 6.47%. The absence of an asterisk indicates that this difference in performance is not significant.

In the same cell where the difference of means appears, there are numbers in square brackets. These numbers indicate the distance/range between the two means after ordering from lowest to highest – neighboring means have a range of two; when comparing the highest and lowest mean accuracy scores here, the range would be 30. Each range corresponds with a critical difference value. In order for two means to be significantly different from each other, the real difference between the means needs to be greater than the critical difference. The critical difference is calculated using the following equation and can be set for significance at .05 and .01, depending on the k values selected[7]:

$$\text{Critical Difference} = \frac{\text{Error Mean Square}}{\text{Avg. number Particiapnts per Group}} \times (\text{k value})$$

Figure 6. *Calculating Critical Difference.*

Results from Duncan's MRT revealed a number of significant differences in performance between the bilingual and non-bilingual schools (Table 9). As a group, the bilingual school

[7] k values increase as the compared means' range increases. There is a set of k values for determining significance at d=.05 and d=.01

participants identified three of the five within-category pairings (ex. *bat-bat*) significantly more accurately than the non-bilingual school students. The mean difference between the schools was significant at the .05 level for two pairings (/e-e/ and /ɛ-ɛ/) and significant at the .01 level for /æ-æ/. There was no significant difference in perception scores for the /i-i/ or /ɪ-ɪ/ within-category vowel pairings.

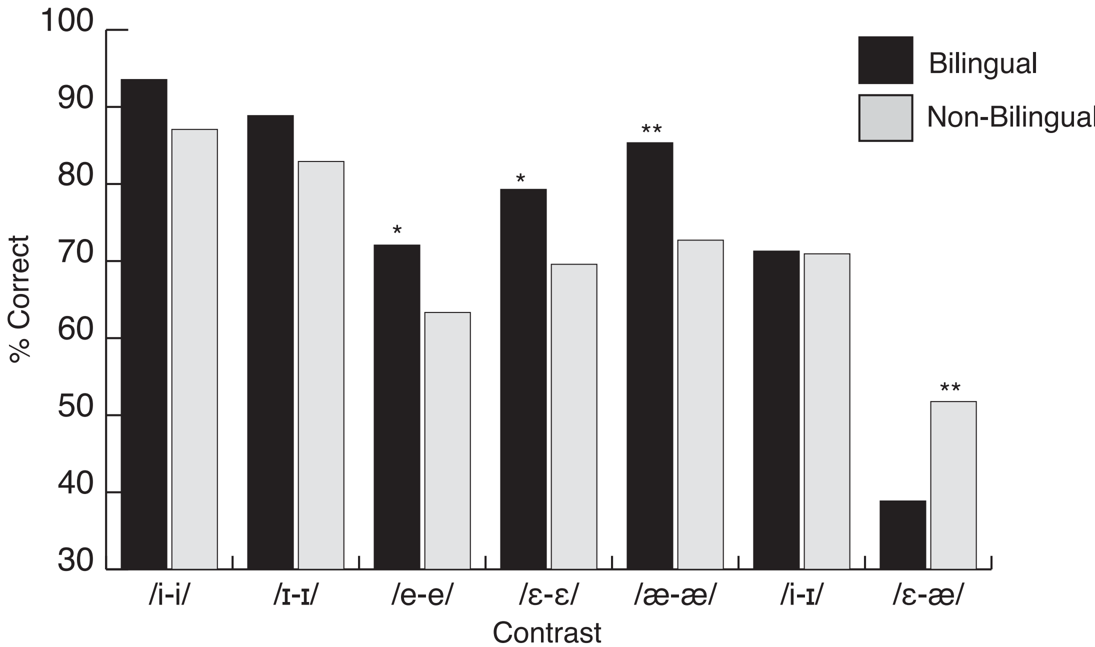


Figure 7. *Accuracy Scores (%) per Contrast by School.* Displays the mean (%) correct response by school for seven vowel pairings. Significance determined through Duncan's MRT (*p=.05, **p=.01)

For between-category vowel contrasts, there was no significant difference in how participants from the two schools performed on the /i-ɪ/ vowel pairing (bil: 93.55%, non: 87.08%). Conversely, the non-bilingual school perceived the /ɛ-æ/ contrast significantly more accurately than the bilingual school at the .01 level (bil: 38.87%, non: 51.77%). For a graphical representation of these mean differences, see Figure 6.

Q3. Does an increase in length of formal instruction improve participants' accuracy in perceiving individual between-category and within-category vowel pairings?

The results from the repeated measure ANOVA (Table 10) reveal that no single grade perceived an individual contrast significantly differently than any other grade. For example, the 6th graders as a group did not perceive the /i-ɪ/ contrast, or any other vowel pairing for that matter, significantly differently than another grade level.

Table 10. *Perception Accuracy of Individual Contrasts by Grade.* (Within-Subject effects) Upon comparing accuracy scores of each individual contrast, there is no main effect for grade (****p*<.001).

	Type III Sum of Sq.	df	Mean Sq.	F	Sig.	Part. ETA sq.	Non Cent. Par.	Observed Power
Contrast	41.304	3.964	0.420	81.091	.000***	.401	321.455	1.000
Grade	1.982	7.928	.250	1.947	.052	.031	15.435	.807
Error	61,632	479.657	.128					

Q4. Do certain school-grade groups perceive specific vowel pairings significantly differently from the other school-grade groups?

Table 11. *Perception Accuracy of Individual Contrast by School and Grade.* (Within-Subject Effects) Upon comparing accuracy scores for each vowel pairing, there is a main effect for school and a significant interaction between school and grade **(*p<.05, ***p<.001).**

	Type III Sum of Sq.	df	Mean Sq.	F	Sig.	Part. ETA sq.	Non Cent. Par.	Observed Power
Contrast	15.161	3.251	4.664	57.583	.000***	.328	187.174	1.000
School	1.567	3.251	.482	5.952	.000***	.048	19.346	.966
Grade	.779	6.501	.120	1.480	.178	.024	9.618	.599
School*Grade	1.262	.501	.194	.397	.024*	.039	15.580	.841
Error	31.068	708	.044					

Even though no school-grade group's overall perception score on the AX Categorical Discrimination task was significantly better or worse than any other, there still exists the possibility that one school-grade group performed significantly differently on specific individual contrasts. For this analysis, participants were grouped together based on two variables – school and grade. This led to comparisons of mean scores between six groups of students: the 2nd grade bilinguals, 4th grade bilinguals, 6th grade bilinguals, 2nd grade non-bilinguals, 4th grade non-bilinguals and 6th grade non-bilinguals. The repeated measures ANOVA (Table 11) that was performed to examine the interaction between school and grade per contrast reached significance [$F(6.501,708)=2.397, p=.024$] and, because of that, further analysis was done to determine where these school-grade differences laid.

Duncan's MRT was used once more to compare the perception scores of each school-grade group against the other five school-grade groups. The mean perception score of the six school-grade groups for each of the 15 vowel pairings resulted in a total of 90 means being ranked and compared. The results for seven of these vowel pairings (/i-i/, /ɪ-ɪ/, /e-e/, /ɛ-ɛ/, /æ-æ/, /i-ɪ/, and /ɛ-æ/) can be found in Tables 12-18.

Table 12. *Duncan's MRT for /i-i/ (Mean difference by group).* Display of school-grade mean comparisons for the /i-i/ vowel pairing (*p=.05, **p=.01)

α Error Degrees of Freedom MS Within Group Error Number of Means Critical Range		p=.05, p=.01 708 .044 90 [in brackets below for each comparison]				
Group (% Correct)	2nd Bilingual (93.48%)	4th Bilingual (89.47%)	6th Bilingual (97.16%)	2nd Non-Bil. (83.09%)	4th-Non-Bil. (87.50%)	6th Non-Bil. (90.28%)
2nd Bilingual (93.48%)	--	4.01 [18]	3.68 [15]	10.39 [30]	5.98 [22]	3.2 [10]
4th Bilingual (89.47%)		--	7.69 [32]	6.38 [12]	1.97 [5]	.81 [9]
6th Bilingual (97.16%)			--	14.07 [40]	9.66 [36]	6.88 [24]
2nd Non-Bil. (83.09%)				--	4.41 [8]	7.19 [20]
4th Non-Bil. (87.50%)					--	2.78 [13]
6th Non-Bil. (90.28%)						--

Table 12 shows that the mean perceptual accuracy scores for /i-i/ ranged from 83.09% (2nd grade non-bilinguals) to 97.16% (6th grade bilinguals) – a difference of 14.07%. However, this difference, just like every other mean difference associated with this vowel pairing did not reach significance. No school-grade group performed significantly differently than any other school-grade group on this vowel pairing.

Table 13. *Duncan's MRT for /ɪ-ɪ/ (Mean difference by group)*. Display of school-grade mean comparisons for the /ɪ-ɪ/ vowel pairing (**p*=.05, ***p*=.01).

α		p=.05, p=.01				
Error Degrees of Freedom		708				
Error Mean Square		.044				
Number of Means		90				
Critical Range		[in brackets below for each comparison]				
Group (% Correct)	2nd Bilingual (90.22%)	4th Bilingual (86.18%)	6th Bilingual (89.77%)	2nd Non-Bil. (75.00%)	4th-Non-Bil. (84.50%)	6th Non-Bil. (88.19%)
2nd Bilingual (90.22%)	--	4.04 [13]	.45 [3]	15.22 [32]	5.72 [15]	2.03 [10]
4th Bilingual (86.18%)		--	3.59 [11]	11.18 [19]	1.68 [3]	2.01 [4]
6th Bilingual (89.77%)			--	14.77 [30]	5.27 [13]	1.58 [8]
2nd Non-Bil. (75.00%)				--	9.5 [17]	13.19 [22]
4th Non-Bil. (84.50%)					--	3.69 [6]
6th Non-Bil. (88.19%)						--

When presented with the /ɪ-ɪ/ vowel pairing (Table 13), the school-grade groups' ability to identify both tokens as belonging to the same L2 English category ranged from 75.00% (2nd grade non-bilinguals) to 90.22% (2nd grade bilinguals) – a difference of 15.22%. However, this difference was not significant. No school-grade group perceived this vowel pairing significantly better or worse than any other school-grade group.

Table 14. *Duncan's MRT for /e-e/ (Mean difference by group).* Display for school-grade mean comparisons for the /e-e/ vowel pairing (**p*=.05, ***p*=.01).

α		p=.05, p=.01				
Error Degrees of Freedom		708				
Error Mean Square		.044				
Number of Means		90				
Critical Range		[in brackets below for each comparison]				
Group (% Correct)	2nd Bilingual (70.11%)	4th Bilingual (67.76%)	6th Bilingual (77.84%)	2nd Non-Bil. (55.88%)	4th-Non-Bil. (68.50%)	6th Non-Bil. (63.19%)
2nd Bilingual (70.11%)	--	2.35 [7]	7.73 [16]	14.23 [18]	1.61 [3]	6.92 [10]
4th Bilingual (67.76%)		--	10.08 [22]	11.88 [12]	.74 [5]	4.57 [4]
6th Bilingual (77.84%)			--	21.96** [34]	9.34 [18]	14.65 [26]
2nd Non-Bil. (55.88%)				--	12.62 [16]	7.31 [9]
4th Non-Bil. (68.50%)					--	5.31 [8]
6th Non-Bil. (63.19%)						--

Table 14 shows that the 6th grade bilingual students identified /e-e/ pairings as containing instances of the same L2 English vowel category significantly better (77.84%) than the 2nd grade non-bilingual students (55.88%). The differences in performance between the all other school-grade groups were not significant.

Table 15. *Duncan's MRT for /ɛ-ɛ/ (Mean difference by group).* Display of school-grade mean comparisons for the /ɛ-ɛ/ vowel pairing (**p*=.05, ***p*=.01).

α	p=.05, p=.01
Error Degrees of Freedom	708
Error Mean Square	.044
Number of Means	90
Critical Range	[in brackets below for each comparison]

Group (% Correct)	2nd Bilingual (75.54%)	4th Bilingual (71.71%)	6th Bilingual (89.77%)	2nd Non-Bil. (59.56%)	4th-Non-Bil. (77.50%)	6th Non-Bil. (68.06%)
2nd Bilingual (75.54%)	--	3.83 [5]	14.23 [28]	15.98 [22]	1.96 [3]	7.48 [14]
4th Bilingual (71.71%)		--	18.06* [32]	12.15 [17]	5.79 [7]	3.65 [10]
6th Bilingual (89.77%)			--	30.21** [50]	12.27 [26]	21.71** [40]
2nd Non-Bil. (59.56%)				--	17.94* [24]	8.5 [8]
4th Non-Bil. (77.50%)					--	9.44 [16]
6th Non-Bil. (68.06%)						--

The 6th grade bilingual students perceived the two tokens of the /ɛ-ɛ/ pairing as the same significantly more frequently (89.77%) than the 2nd and 6th grade non-bilinguals (59.56%, 68.06%) and the 4th grade bilinguals (71.71%) (Table 15). In addition to significantly underperforming when compared to the 6th grade bilinguals, the 2nd grade non-bilingual group also underperformed when compared to the 4th grade non-bilinguals (77.50%). The difference of the mean scores between the other school-grade groups did not reach significance.

Table 16. *Duncan's MRT for /æ-æ/ (Mean difference by group).* Display of school-grade mean comparisons for the /æ-æ/ vowel pairing (**p*=.05, ***p*=.01).

α		p=.05, p=.01				
Error Degrees of Freedom		708				
Error Mean Square		.044				
Number of Means		90				
Critical Range		[in brackets below for each comparison]				
Group (% Correct)	2nd Bilingual (82.61%)	4th Bilingual (78.95%)	6th Bilingual (93.75%)	2nd Non-Bil. (69.12%)	4th-Non-Bil. (71.5%)	6th Non-Bil. (77.78%)
2nd Bilingual (82.61%)	--	3.66 [3]	11.14 [32]	13.49 [22]	11.11 [16]	4.83 [7]
4th Bilingual (78.95%)		--	14.80 [34]	9.83 [19]	7.45 [14]	1.17 [5]
6th Bilingual (93.75%)			--	24.63** [52]	22.25** [47]	15.97 [38]
2nd Non-Bil. (69.12%)				--	2.38 [6]	8.66 [15]
4th Non-Bil. (71.5%)					--	6.28 [12]
6th Non-Bil. (77.78%)						--

Table 16 shows that when presented with the /æ-æ/ pairing, the 6th grade bilinguals (93.75%) identified the two instances of /æ/ as the same at a rate significantly better than the 2nd grade non-bilinguals (69.12%) and the 4th grade non-bilinguals (71.5%). The difference in performance between the other groups was not significant.

Table 17. *Duncan's MRT for /i-ɪ/ (Mean difference by group).* Display of school-grade mean comparisons for the /i-ɪ/ vowel pairing (**p*=.05, ***p*=.01).

α	p=.05, p=.01					
Error Degrees of Freedom	708					
Error Mean Square	.044					
Number of Means	90					
Critical Range	[in brackets below for each comparison]					
Group (% Correct)	2nd Bilingual (74.46%)	4th Bilingual (68.42%)	6th Bilingual (70.45%)	2nd Non-Bil. (81.62 %)	4th-Non-Bil. (58.75%)	6th Non-Bil. (77.78%)
2nd Bilingual (74.46%)	--	6.04 [10]	4.01 [6]	7.16 [13]	15.71 [20]	3.32 [9]
4th Bilingual (68.42%)		--	2.03 [5]	13.20 [22]	9.67 [11]	9.36 [18]
6th Bilingual (70.45%)			--	11.17 [18]	11.7 [15]	7.33 [13]
2nd Non-Bil. (81.62%)				--	22.87* [32]	3.84 [5]
4th Non-Bil. (58.75%)					--	19.03* [28]
6th Non-Bil. (77.78%)						--

In Table 17 we see that the 4th grade non-bilingual students perceived the tokens /i/ and /ɪ/ to be different 58.75% of the time – significantly less often than the 2nd grade non-bilinguals (81.62%) and the 6th grade non-bilinguals (77.78%). No other differences between school-grade groups reached significance.

Table 18. *Duncan's MRT for /ɛ-æ/ (Mean difference by group).* Display of school-grade mean comparisons for the /ɛ-æ/ vowel pairing (*p=.05, **p=.01)

α	p=.05, p=.01					
Error Degrees of Freedom	708					
Error Mean Square	.044					
Number of Means	90					
Critical Range	[in brackets below for each comparison]					
Group (% Correct)	2nd Bilingual (36.68%)	4th Bilingual (41.45%)	6th Bilingual (38.92%)	2nd Non-Bil. (60.29%)	4th-Non-Bil. (41.75%)	6th Non-Bil. (57.64%)
2nd Bilingual (36.68%)	--	4.77 [4]	2.24 [2]	23.61** [16]	5.07 [5]	20.96** [12]
4th Bilingual (41.45%)		--	2.53 [3]	18.84* [13]	.30 [2]	16.19* [9]
6th Bilingual (38.92%)			--	21.37** [15]	2.83 [4]	18.72* [11]
2nd Non-Bil. (60.29%)				--	18.54* [12]	2.65 [5]
4th Non-Bil. (41.75%)					--	15.89* [8]
6th Non-Bil. (57.64%)						--

Table 18 shows that the 2nd and 6th grade non-bilingual participants correctly perceived the /ɛ-æ/ vowel pairing as different significantly more frequently than the other four groups, but did not differ significantly from each other. The 2nd, 4th, and 6th grade bilinguals, along with the 4th grade non-bilinguals, did not differ significantly in their performance on this vowel pairing.

Overall Trends from Within-Category Pairings. Through the analysis of within-category vowel pairings, two patterns emerged – one of similarity and the other of divergence. First, the variance in the six school-grade groups' perception scores for the /i-i/ and /ɪ-ɪ/ vowel pairings did not reach significance, suggesting that each school-grade group can identify members of these two L2 categories equally well. Second, in each of the three remaining within-category contrasts (/e-e/, /ɛ-ɛ/ and /æ-æ/), at least one school-grade group performed significantly better than another. Most notably, the 6th grade bilingual group performed significantly better than at least one other school-grade group in each of these vowel pairings: (1) they perceived the /e-e/ pairing more accurately than the 2nd grade non-bilinguals, (2) they performed significantly better than the 2nd grade non-bilinguals, the 6th grade non-bilinguals, and the 4th grade bilinguals on the /ɛ-ɛ/ contrast and (3) they perceived /æ-æ/ significantly more accurately than the 2nd and 4th non-bilingual groups. Conversely, the 2nd grade non-bilinguals scored significantly worse on each of these three vowel pairings. They performed significantly worse than the 6th grade bilinguals on all three of these within-category vowel pairings and worse than the 4th grade non-bilinguals on the /ɛ-ɛ/ contrast. Comparisons of the remaining group mean differences did not reach significance.

Overall Trends from Between-Category Pairings. According to the Duncan's MRT, five school-grade groups did not differ significantly in their perception of /i/ and /ɪ/ as separate phonemes. The 4th grade non-bilingual students, however, were significantly less accurate than the 2nd grade non-bilinguals, and the 6th grade non-bilinguals at perceiving this contrast. Regarding the /ɛ-æ/ contrast, the 2nd and 6th grade non-bilingual participants perceived this vowel pairing as containing distinct L2 phones significantly more frequently than the other four groups, but did not differ significantly from each other. The other four groups (2nd, 4th, and 6th grade bilinguals and the 4th grade non-bilinguals) did not differ significantly from each other.

Q5. Does the order that the English words are presented in affect the accuracy of perception of that vowel pairing?

The Natural Referent Vowel theory states that the order in which two vowels are presented affects people's perception of those vowels. More specifically, when presented with two vowels, a person will discriminate the contrast with increased reliability when the more central vowel is presented before the more peripheral vowel. Regardless of language background, people will hear a difference between the English phones /i/ and /ɪ/ more when they are presented first with /ɪ/ and then /i/ than when presented with the same phones in reverse order.

To test this hypothesis, a series of paired t-tests was performed comparing the participants' mean perception accuracy score in two scenarios (1) when presented with a vowel contrast where the peripheral vowel was presented first and (2) when the peripheral vowel was presented second (ex. /æ-ɛ/ vs. /ɛ-æ/, respectively). Because this order effect has been found with participants of varying experience (Polka & Bohn, 1996; Polka, Molnar, Baum, Menard, & Steinhauer, 2009) the averages from all 124 participants were grouped together. Results of these comparisons are found in Table 19.

Of the ten vowel pairings, there was a significant difference in participants' perception scores based on presentation order for two vowel pairings, one of which supported the NRV theory. The vowel pair that affected participants' perception in support of the NRV contained the phones /ɪ/ and /ɛ/. Here, their was a significant improvement in accuracy when these vowels were presented as /ɛ-ɪ/ (M=.8196, SD=.1980) than when presented in the opposite order /ɪ-ɛ/ (M=.6996, SD=.2723), t(123)=-5.883, p<.001. The other vowel pair that significantly affected perception depending on the order in which the phones were presented contained /ɛ/ and /æ/ - but this order affect did not support the NRV: participants perceived a difference between /æ-ɛ/ (M=.5000, SD=.2943) significantly more than when the order was reversed, /ɛ-æ/ (M=.4022, SD=.2773); t(123)=-4.569, p<.001.

Table 19. *Nartural Referent Vowel (NRV) All Participants*. (Paired samples t-test). Comparing accuracy scores within vowel pairings to test for order effects on perception of phones (***p*<.001).

	Vowel Pairing	Mean Accuracy	Std. Deviation of Indiv. Contrast	Paired Differences			t	df	Sig. (2-tailed)
				Mean Difference	Std. Deviation	Std. Error Mean			
1	i-ɪ	.7218	.2833	.02117	.19891	.01786	1.185	123	.238
	ɪ-i	.7006	.3057						
2	i-e	.6754	.3251	.02319	.17918	.01609	1.441	123	.152
	e-i	.6522	.3273						
3	i-ɛ	.9113	.1502	-.02319	.12588	.01130	-2.051	123	.042
	ɛ-i	.9345	.1166						
4	i-æ	.9294	.1268	-.01109	.11714	.01052	-1.054	123	.294
	æ-i	.9405	.1051						
5	ɪ-e	.5353	.2782	.02923	.22548	.02025	1.444	123	.151
	e-ɪ	.5060	.2811						
6	ɪ-ɛ	.6996	.2723	-.11996	.22900	.02056	-5.833	123	.000**
	ɛ-ɪ	.8196	.1980						
7	ɪ-æ	.9385	.1155	.00605	.11588	.01041	.581	123	.562
	æ-ɪ	.9325	.1239						
8	e-ɛ	.8911	.1520	.00242	.17111	.01537	.157	123	.875
	ɛ-e	.8887	.0191						
9	e-æ	.9415	.1074	.00504	.12181	.01094	.461	123	.646
	æ-e	.9365	.1280						
10	ɛ-æ	.4022	.2773	-.09778	.23833	.02140	-4.569	123	.000**
	æ-ɛ	.5000	.2943						

Q6. Do the adult L1 Spanish speakers exhibit evidence of perceptual learning when their overall performance on the AX categorical discrimination task is compared to those of the participants from the bilingual and non-bilingual school?

Table 20. *Overall Average Scores of Perception Task*. Displays the average scores on the AX Categorical Discrimination task by the children and adults.

Group	**Number**	**Mean Score (%)**	**Std. Deviation**
Bilingual School	64	79.56	8.01
Non-Bilingual School	60	75.97	8.74
Adults	11	91.18	4.64
Total	135	78.91	9.04

In Table 5 (page 38), a 2x3 ANOVA determined that the participants from the bilingual school perceived L2 English front vowels significantly better than the participants from the non-bilingual school. In this section, the mean perception scores from the bilingual and non-bilingual schools (79.56% and 75.97%, respectively) were compared against the perception scores of eleven adult L1 Spanish speakers' (91.18%) from a previous perception study. A One-Way ANOVA was performed to test for a main effect of group (Table 21).

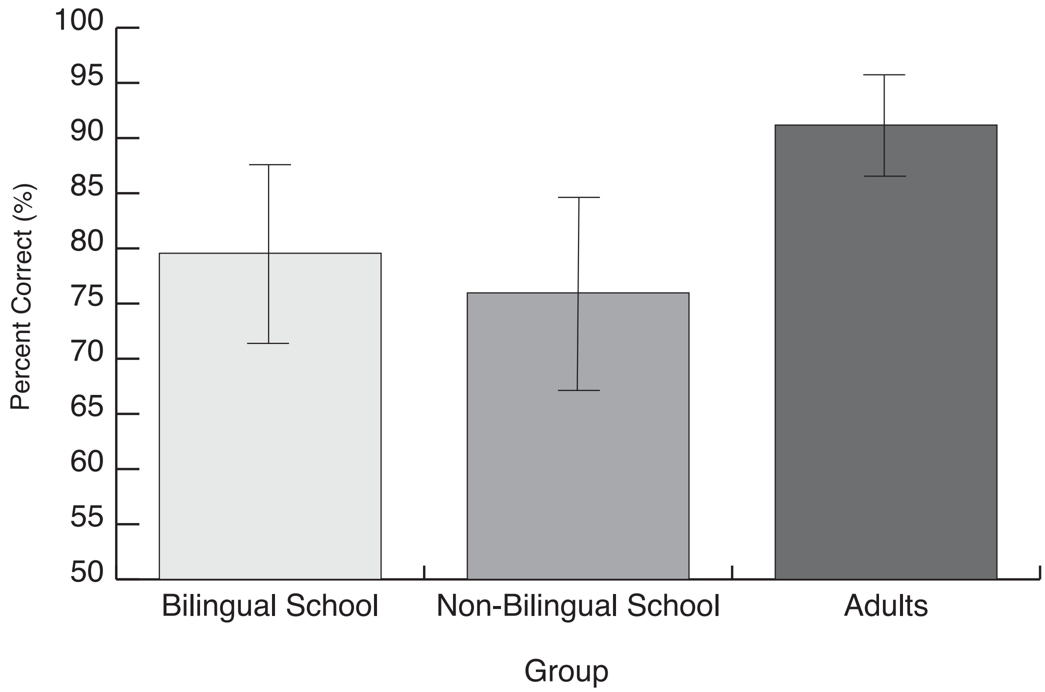


Figure 8. *Comparison of Overall Perception Scores from AX Categorical Discrimination Task.* A display of average perception score by school group - including the adult group.

Table 21. *Comparison of Overall Perception Accuracy Scores by School and Adults.* (One-Way ANOVA). This table displays the mean accuracy scores for the three groups of participants. There is a significant effect for group (***p<.001).

	Type III Sum of Squares	Df	Mean Sq.	F	Sig.	Partial Eta Sq.	Noncent. Param.	Observed Power
Intercept	49.404	1	49.404	7437.4	.000	.983	7437.4	1.0
Group	.220	2	.110	16.587	.000***	.201	33.174	1.0
Error	.877	132	.007					

Because the One-Way ANOVA revealed a significant effect for group [F(2,132)=16.587,p<.001] (Table 21), a post-hoc was then performed on the overall accuracy of the three groups to determine which groups achieved significantly different overall accuracy scores (Table 22). The results of the Bonferroni post-hoc showed that each group performed significantly differently from the other groups in the discrimination task. The adult group achieved significantly higher overall perception scores (91.18%) than the children at the bilingual ($p<.001$) and non-bilingual schools ($p<.001$) (79.56% and 75.97%, respectively).

Table 22. *Comparison of Overall Perception Accuracy Scores (by School and Adults).* (Post-Hoc). Shows a significant main effect for group - the adults performed significantly better than the other two groups ($p<.001$) and the bilingual school performed significantly better than the non-bilingual school ($p=.046$)

Group (I)	Group (II)	Mean Diff.	Std. Error	Sig.	95% Confidence Interval	
					Lower	**Higher**
Adult	Bilingual	.11619	.026602	.000***	.05169	.18070
	Non-Bilingual	.15215	.026732	.000***	.08733	.21697
Bilingual	Non-Bilingual	.03596	.014646	.046*	.00044	.07147

2.7 DISCUSSION – PERCEPTION

2.7.1 Overall perception scores

Experience with an L2 has been shown to positively affect perception of L2 sounds (Flege et al, 1997; Højen & Flege, 2006; Levy & Strange, 2008). In the present study, the bilingual school participants earned significantly higher accuracy scores than their non-bilingual school counterparts in the AX Categorical Discrimination task containing American English front vowels. Because students from both schools shared the same L1 and were both introduced to English as a foreign language in first grade, this suggests that an increased amount of exposure to L2 English in the classroom (approximately 14 hours per week versus 4 hours per week) beneficially impacted the bilingual students' perception of these English vowels. It is important to note though, that the effect size was small (η^2=.08) – so, even though the participants from the bilingual school benefitted from the additional exposure, its contribution was rather small.

Even though the difference in amount of exposure had an effect on overall accuracy by school, there was no significant effect for grade. No grade, regardless of school type, performed significantly better than any of the others - Figure 4 and Figure 5 provide insight as to why. Figure 5 displays the beginning of divergent L2 perception patterns – the bilingual school participants' scores remained constant or increased with age (79.15% → 80.46%) and the non-bilingual school participants' scores decreased with age (77.59% → 74.84%). By taking the average score of these opposing trends at each grade level we see that the upward movement from second to sixth grade at the bilingual school is off set by the decreasing scores of each corresponding grade at the non-bilingual school. This resulted in average scores by grade that varied by only 1.71% (2nd 78.49%, 4th 77.13%, 6th 77.93%) (Figure 4).

Despite the increasingly divergent perception scores present between schools, the interaction between school and grade on overall performance was not significant. This result was unexpected, especially taking into consideration the overall effect for school. It is surprising that with continued exposure to the L2, the 6th grade bilingual school students did not perform significantly better than any of the other groups. Jia and Aaronson (2003) found that significant improvement in phonological skills of immigrant children after three years immersed in the target language environment, but perhaps there are limitations to the beneficial impact of L2 exposure when it is restricted solely to the classroom environment (White & Genesee, 1996).

Donato and Tucker (2010) stated that high-quality foreign language programs in elementary schools are a prerequisite for reaching high levels of proficiency later on. It may be that the additional 14 hours per week over four years experienced by the 6th grade bilingual school participants was not enough time or exposure to significantly impact their perception scores and set them apart from their peers. As we know, it takes years to establish L1 phonetic categories (Hazan and Barrett, 2000) and that L2 phonetic categories improve with increased experience, so it is possible that if the patterns seen in Figure 5 were to continue into middle and high school, we would see the emergence of a significant interaction between school and grade on overall performance. Further research will have to be conducted to see if these patterns continue into adolescence.

The non-significant differences in performance between the grades across schools in conjunction with the non-significant school-grade interaction support the conclusion made by White and Genesee (1996) that length of study in a formal environment is not analogous to AOA in a natural setting. We cannot expect to see the same perceptual improvement from 4-14 hours of L2 exposure per week as we do with the exposure immigrant participants experience through living in a target language community over the same amount of time (approximately 4 years).

To test the claims that L2 perceptual learning can indeed become more native-like with experience (Gottfried, 1984; Neufeld, 1988; Fabra & Romero, 2012), the overall perception scores from the children at the bilingual and non-bilingual schools were compared to the overall perception scores of 11 L1 Spanish-speaking adults. The adults were late onset learners, had only attended non-bilingual schools growing up, and had been living in the US for an average of 3.8 years at the time of the study.[8] The results of the One-Way ANOVA (Table 22) comparing group means revealed that the adults performed significantly better (91.18%) than the non-bilingual (75.97%; *p*<.001) and bilingual (79.56%; p<.001) groups on this perception task. This finding supports the claim that with sufficient exposure, even late-learners can improve their understanding of L2 phonology.

[8] For more details of the adult L1 Spanish speakers, see **Table 3**

2.7.2 Individual contrasts – by school

2.7.2.1 /i/ and /ɪ/

In addition to investigating how additional exposure affected children's overall performance on this discrimination task, analyses looking into how the amount of exposure affected performance on individual vowel pairs were also conducted. Of the seven contrasts that were the focus of this dissertation (/i-i/ /ɪ-ɪ/ /e-e/ /ɛ-ɛ/ /æ-æ/ /i-ɪ/ /ɛ-æ/), participants from the bilingual and non-bilingual schools perceived three vowel pairings at statistically similar rates: /i-i/ (93.55% vs. 87.08%), /ɪ-ɪ/ (88.87% vs. 82.92%) and /i-ɪ/ (71.29% vs. 70.94%). Because previous studies have found that L1 Spanish speakers perceive English /i/ to be perceptually similar to Spanish /i/ (Flege, 1991; Iverson & Evans, 2009), it is not a surprise that participants from both schools were able to identify when two tokens of English /i/ were presented together. In addition to being perceptually similar, Spanish and English /i/ are also acoustically similar (Bradlow 1995; Mendez, 1982). Because of these cross-linguistic similarities, the Speech Learning Model predicts that the /i/ phone from both languages likely have collapsed into one phonetic category. The high frequency with which participants from both schools were able to identify within-category tokens of /i/ support this prediction.

L1 Spanish speakers' perception of English /ɪ/ is less straightforward. In the L2-L1 assimilation task conducted by Flege (1991) (Figure 2), experienced and inexperienced L1 Spanish speakers of English perceived /ɪ/ to be perceptually most similar to Spanish /i/ 36% of the time, to Spanish /e/ 39% of the time, and indicated that it wasn't similar to any Spanish category 21% of the time. This variation suggests that there is not a good Spanish equivalent for English /ɪ/ and that this could indicate the potential for L1 Spanish speakers to establish a *new phonetic category*, according the SLM. When L2 learners form a new phonetic category, their sensitivity to within-category differences decreases, allowing them to disregard individual differences in productions of /ɪ/ from multiple speakers. The 124 participants have an average accuracy score of 85.99% on the /ɪ-ɪ/ vowel pairing – relatively high. However, examining the accuracy with which these participants disregard within-category differences is not tantamount to their having formed a new phonetic category for English /ɪ/. To explore this possibility, we must also examine their performance perceiving the difference between /ɪ/ and its closest acoustic neighbor, /i/.

The bilingual and non-bilingual schools perceived the /i-ɪ/ vowel pairing correctly at same frequency (*M*=71.12%). Even though this is lower than their combined perception scores for the individual within-category vowel pairings, it is still relatively high. In fact, following the L2-L1 assimilation patterns from Flege (1991), PAM-L2 predicts that this contrast would be perceived by L1 Spanish speakers "relatively easily." In Cebrian (2006), the author concluded that the frequency with which his participants perceived a difference between /i-ɪ/ (59%) fell into the realm of "relatively easy". The current participants' combined average accuracy score on /i-ɪ/ is well above this boundary used by Cebrian, and therefore, supports PAM-L2's prediction. The participants' relatively high performance distinguishing these two English phones goes against many previous research studies that have found the /i-ɪ/ contrast to be very difficult for L1 Spanish speakers to perceive (Bohn, 1995; Escudero, 2006; Escudero & Boersma). It is even possible that they have formed a new phonetic category for English /ɪ/.

There is the possibility that the native Spanish speakers were discriminating between English /i/ and /ɪ/ using durational cues instead of spectral cues. It has been found that L1 Spanish speakers often rely on durational cues to distinguish between English tense-lax vowels even though L1 English speakers primarily rely on acoustic cues to make the same discrimination (Barrios et al., 2016; Hillenbrand et al., 2000). If the participants in the current study are indeed using duration to make this distinction instead of acoustic cues, we should see that they produce both /i/ and /ɪ/ in the same way. This possibility will be addressed in the production task described in section 3.2.

2.7.2.2 /e/-/ɛ/-/æ/

Participants from the bilingual school identified the remaining within-category vowel pairings as containing to instances of the same L2 phone significantly more frequently than participants from the non-bilingual (/e-e/ 72.07% vs 63.33%, /ɛ-ɛ/ 79.30% vs. 69.58%, /æ-æ/ 85.35% vs. 72.71%). At first glance it looks like the participants from the bilingual school may have categories established for these L2 phones because of their strong performance identifying phones from the same category, but upon examination of their poor performance distinguishing the /ɛ-æ/ contrast, this doesn't seem as likely.

Results on the /ɛ-æ/ contrast were unexpected for two reasons: first, the non-bilingual school students were the ones who performed significantly better (51.77%) than students from the bilingual school (38.87%) (p<.01). It was predicted that the students from the bilingual school would have had significantly more accurate perception due to increased exposure to L2 sounds. The second unexpected result is based on the predicted versus actual levels of discrimination difficulty as projected by PAM-L2. Based on previous vowel assimilation task (Flege, 1991) (Table 2) PAM-L2 predicts that L1 Spanish speakers will be able to discriminate the two sounds relatively easily because English /ɛ/ and /æ/ assimilate onto Spanish /a/ in a way that matches Best and Tyler's (2007) description of a *Category-Goodness Pattern*. English /ɛ/ assimilates to Spanish /a/ 39% of the time and English /æ/ maps onto Spanish /a/ 82% of the time (Category-Goodness Pattern). However, as mentioned above, students from both schools had a difficult time perceiving this distinction. The SLM predicts the combined accuracy score for both schools was below chance (45.11%).

A potential explanation for why the non-bilingual students perceived the /ɛ-æ/ contrast more accurately than the bilingual students deals with their differing stages of interlingual category formation. Having concluded that experience with an L2 results in more accurate perception of the L2 sound system, it is logical to assume that the bilingual school students have interlingual categories for /ɛ/ and /æ/ that more closely resemble those of a native English speaker. Conversely, the non-bilingual school students' interlingual categories more closely align with the closest L1 counterparts, /e/ and /a/. In other words, the non-bilinguals' new interlingual /ɛ/ and /æ/ categories are further apart from each other in perceptual space than the bilinguals', presumably resulting in less perceptual confusion between /ɛ/ and /æ/. This view is in line with Flege and Bohn's (1989) conclusion that L1 Spanish speakers could accurately identify L2 English /ɛ/ and /æ/ along a continuum, not because of correctly established L2 categories, but because they had strongly assimilated English /æ/ onto L1 Spanish /a/, a phone that is produced further back in the vocal tract.

2.7.3 Individual contrasts – by grade

Similar to how there was no main effect for grade (i.e., length of exposure) on the participants' overall perception scores, there was no effect of grade on any of the individual vowel pairings either. This suggests that there is no acoustic feature in these L2 English vowels that the participants become more or less attuned to as they progress through their programs. Once again, this could be due to an insufficient amount of exposure (White & Genesee, 1996) in both schools or it could be that the insufficient amount of exposure from the non-bilingual school is occluding the perceptual changes to individual phones experienced by the bilingual students. To see if this is indeed the case, we must look at how each school-grade group performed on these seven vowel pairings.

2.7.4 Individual contrasts – school*grade interaction

2.7.4.1 /i/ and /ɪ/

There was no significant school*grade interaction for the within-category vowel pairings /i-i/ and /ɪ-ɪ/, thereby mirroring the results of the comparison of individual contrasts by school (Section 2.6.3). Each school-grade group perceived these vowel pairings equally as well. However, there was a significant interaction of school and grade for the between-category contrast /i-ɪ/. The 4th grade non-bilingual school students identified the tokens of this contrast significantly less frequently (58.75%) than the 2nd grade non-bilinguals (81.62%) and the 6th grade non-bilinguals (77.78%). The performance of the 4th grade non-bilinguals is unexpected, considering their consistently accurate perception of /i-i/ and /ɪ-ɪ/ (87.5% and 84.5%, respectively). Regardless of the statistically significantly worse performance than their other non-bilingual school classmates, all six school-grade groups support PAM-L2's prediction that this contrast would be discriminated "relatively easily".

One possible explanation for the 4th grade non-bilingual school participants' worse performance compared to that of the 2nd and 6th graders from the same school could be due to overregularization. Overregularization (i.e. U-Shaped Learning) is a relatively common occurrence in L1 and L2 acquisition. It references the phenomenon where an individual starts out using a linguistic structure in a native-like way, then replaces it with a non-native-like form or pronunciation only to return to the original manner of production at a later time (e.g., Marcus, Pinker, Ullman, Hollander, Rosen, Xu, & Clahsen, 1992). The quintessential example of this is the overgeneralization of the *-ed* suffix used to indicate past tense in English. Learners have been shown to acquire irregular past tense before noticing the pattern for creating regular past tense verbs. However, upon this recognition, learners will add -*ed* to the base of a verb that should be irregular, thereby creating ungrammatical forms like *goed* or *comed* instead of *went* and *came*. With more exposure and practice, users revert back to using the native-like form.

There are fewer studies that show overregularization of phonological elements of language and the ones that do usually focus on a very limited number of participants' production shortly after having arrived in an L2 environment (e.g., Abrahamsson, 1999, 2003; Sato, 1987). These studies observed this u-shape learning in L2 learners' production of consonant clusters and suggested that the decrease in pronunciation accuracy was connected to an increase in proficiency - as the learners attempted to relay a message instead of simply producing simple words or phrases in isolation, they committed more errors. Since many studies have linked perception and production, it is conceivable that the u-shape learning present in these longitudinal production studies also impacted the language learners' perception of sounds they heard in natural speech, although it must be noted that perception was not examined in these studies.

This does, however, go against the general principle that increased exposure to a language improves perception and production, but it may actually illustrate a natural progression in language development. Fabra (2005) found that native Catalan speakers were producing English /ɪ/ more accurately than English /i/, even though English /i/ and Catalan /i/ are acoustically similar. Cebrian (2007) claimed that in the process of creating a new L2 category (i.e. coming to realize the cues needed to differentiate /i/ for /ɪ/) there may be a period in which the established category worsens. We may only see this occur in the non-bilingual school if this process already occurred in the bilingual school before second grade.

2.7.4.2 /e/-/ɛ/-/æ/

Once again, there is more significant variation in the performance of the school-grade groups on these three phones, much like there was when comparing the scores of each school. The sixth grade bilinguals performed significantly better than at least one non-bilingual school grade on each of these three within-category vowel pairings. For /e-e/ they performed significantly better than the 2nd grade non-bilinguals (77.84% vs. 55.88%), for /ɛ-ɛ/ they were significantly more accurate (89.77%) than the 4th grade bilinguals (71.71%) and the 2nd (59.56%) and 6th non-bilinguals (68.06%), and on the /æ-æ/ vowel pairing the 6th grade bilinguals (93.75%) performed significantly better than the 2nd and 4th grade non-bilinguals (69.15% and 71.5%, respectively).

The 6th grade bilingual school students' superior performance on these within-category contrasts leads one to believe that they may have established L2 phonetic categories for /ɛ/ and /æ/, but their performance discriminating these phones from one another does not support this (38.92%). The 6th grade and the 2nd grade non-bilinguals, on the other hand, performed significantly better at discriminating /ɛ/ from /æ/ than each grade from the bilingual school and the 4th grade non-bilinguals.[9]

PAM-L2 predicted that this contrast would be relatively easy to perceive because the native Spanish speakers in Flege (1991) assimilated English /ɛ/ and /æ/ onto the same category at different frequencies. Flege's participants perceived English /æ/ to be most similar to Spanish /a/ 82% of the time, but they associated English /ɛ/ with Spanish /a/ 39% of the time, with Spanish

[9] For a possible explanation as to why the 4th graders at the non-bilingual school performed worse than their 2nd and 6th grade classmates, refer to the parts of section 2.7.4.1 referring to overregularization.

/e/ 44% of the time, and to no L1 vowel 13% of the time. If the bilingual school students were to assimilate English /ɛ/ and /æ/ more to the *none* category than the non-bilingual school students, this would be one explanation as to why their ability to discriminate the two sounds is worse. If the non-bilingual school students perceive /ɛ/ and /æ/ to be more similar to Spanish /e/ and /a/, they would have an easier time discriminating the two sounds. In a case like this, the ability to perceive differences between L1 and L2 phones may be an inhibiting factor in the establishment of new L2 phonetic categories.

3.0 EXPERIMENT 2 – PRODUCTION

3.1 METHODS

3.1.1 Participants

All of the child participants from the perception task also participated in the production task.

3.1.2 Stimuli

3.1.2.1 English

Images presented to the children for the picture-naming task were selected after holding discussions with the English teachers at the bilingual and non-bilingual schools. The vocabulary chosen for this task were based on the teachers' prediction that their students would know the words in English.

In the end, two pictures were shown for every vowel under investigation in this dissertation: /i/ *teacher, green;* /ɪ/ *sister*, *big;* /e/ *baby*, *paper*; /ɛ/ bed, pen /æ/ *cat*, *hand* and the low back vowel /ɑ/ *father, hot* (Table 3). Although two images were shown, only the participants' productions of *teacher, sister, baby, pen, cat,* and *father* were used for analysis. The production of these words were chosen because (1) their consonantal environments were 'cleaner', that is to say the stressed vowel was not adjacent to a rhotic and, therefore, these words would exhibit fewer effects of coarticulation and (2) more students were able to recall these English words, which then led to more data points being collected.

Pictures are often used in L2 production tasks to limit the potential orthographic effects on the production of target words (e.g., Morrison, 2002). In English, for example, the grapheme ‘i' often corresponds to /ɪ/ in stressed syllables (ex. *bit, children*) but in Spanish the grapheme ‘i' always corresponds to the phone /i/ (ex. *pito, giro*). To prevent the possibility of confusion that could arise from a shared grapheme mapping onto different phonemes in English and Spanish, only pictures were used to elicit the production of English words.

3.1.2.2 Spanish

To examine the children’s production of the Spanish vowels most closely corresponding to the English vowels under investigation, the Spanish participants produced the Spanish phonemes /i/, /e/ and /a/ within the words words *pito*, *peto*, and *pato.* Because Spanish has an orthography that matches one grapheme to one phoneme, the orthography’s effect on the participants’ production should be minimal and shouldn’t impact how the L1 Spanish participants produce the Spanish words they are reading. This should allow for comparisons to be drawn between the English and Spanish production tasks.

3.2 PICTURE-NAMING TASK (CHILDREN)

A week after participating in the perception task, the students participated in two production tasks – one in English and the other in Spanish (section 3.1.2). The two tasks took a total of eight minutes to complete and were done in the same rooms in which the perception tasks took place. The same headphones as the previous task were used because they had a built in microphone. The participants' utterances were recorded onto the PI's personal computer using Garage Band and were later analyzed using the speech analysis software Praat.

English - The participants were shown a series of 12 pictures on a computer screen one at a time (Appendix A). Each picture was presented on the computer four times in a random order. The participants were instructed to say the name of each picture in English when it appeared on the screen. The pictures were displayed on the screen for five seconds, followed by two seconds of a blank screen before the next picture was presented. The pictures displayed to the participants were chosen because each one contained one of the focus vowels, /i ɪ e ɛ æ ɑ/, and because the English instructors at the schools predicted that their students would know these words in English.

Spanish – One at a time, three Spanish words (*pito, peto, pato*) were shown to the participants on the computer screen. Each word appeared three times for a period of five seconds. The participants were instructed to read the Spanish word aloud every time the word on the screen changed.

Table 23. *Production Task - Elicited Words*. The participants were instructed to produce the following words in (1) a picture-naming task [English] and (2) a word-reading task [Spanish].

Vowel Sound	English Word	Spanish Word
i	teacher, green	pito
ɪ	sister, big	--
e	paper, baby	peto
ɛ	bed, pen	--
æ	cat, hand	--
ɑ	father, hot	--
a	--	pato

3.3 RESEARCH QUESTIONS AND PREDICTIONS

*Q1. Do native Spanish speakers produce the L2 English front vowels /*i ɪ e ɛ æ*/ in significantly different ways?*

Of the seven vowel pairings under investigation, participants from both schools perceived six at an averaged rate above 70% (/i-i/, /ɪ-ɪ/, /e-e/, /ɛ-ɛ/, /æ-æ/ and /i-ɪ/) and one with an accuracy of 52% (/ɛ-æ/). Based on the notion that accurate perception precedes accurate production put forth by SLM and PAM-L2, we should expect to see /i/ and /ɪ/ produced differently from each other because participants from both schools were able to consistently identify the within-category and between-category presentations using this phones. Furthermore, SLM hypothesizes that with experience, /ɪ/ could be considered a new sound, thereby aiding in the establishment of a new phonetic category that is perceived and produced in a more native-like manner (Flege, 1995). However, as stated in section 1.5.2, perception and production do not always match: Flege, Bohn and Jang (1997) found that even though a group of adult L1 Spanish speakers could identify the boundary of the L2 English phone /i/, they could not produce English /i/ and /ɪ/ acoustically differently, and the two were often confusable when presented to L1 English speakers. It may be that there is a lag between what can be perceived and what can be produced.

Based solely on the perception data from Section 2, if the participants were to have difficulty producing a difference between any of the English front vowels it would be /ɛ/ and /æ/ because of the overwhelming difficulty both groups had in hearing a difference between these two sounds (38.87% and 51.77%, respectively). In a study by Jeske (2012), a group of L1 Spanish speakers were unable to perceive a difference between English /ɛ/ and /æ/ at a rate better than chance, but were able to produce the two phones in distinct ways based on spectral characteristics. These productions, though, were elicited via an imitation production task, which requires less processing than picture-naming, so a direct comparison between these results and those from the present study would be tenuous.

Q2. Is there a significant difference in how the bilingual school and non-bilingual school produce these English vowels?

As stated in section 1.5.1, it is often believed that perception necessarily precedes production (e.g., Flege et al., 1999). Because there was no significant difference between how the bilingual and non-bilingual school participants perceived vowel pairings containing /i/ and /ɪ/, these two L2 vowels could be produced similarly by both groups. However, it has been shown that L1 Spanish adults' production of /ɪ/ benefitted marginally from more exposure to English even though they did not differ from the inexperienced group in a perception task (Flege et al., 1997). If this is the case for the current groups of participants, we may see the English /ɪ/ produced by the bilingual school to be slightly more native-like than the production of English /ɪ/ by their non-bilingual counterparts.

Even though there was no significant variation between the perception of /i/-/ɪ/ vowel pairings by school, there was considerable variation between how the two groups perceived vowel pairings containing /ɛ/ and /æ/ - the bilingual students were more accurate perceiving similarities in the within-category pairings, but the non-bilingual students were significantly better at identifying the differences in the between-category pairing: /ɛ-æ/. Because of this significant difference in performance with these two vowels, we may see significant differences between the groups' productions as well.

The bilingual school students may have performed significantly worse discriminating between /ɛ/ and /æ/ because they are in the process of establishing new phonetic categories for the two English sounds. If the bilingual school participants' additional exposure to L2 English provided them with enough input to begin assimilating English /ɛ/ and /æ/ less strongly onto Spanish categories /e/ and /a/, respectively, their productions of these two English sounds may be produced further away from their nearest acoustic and perceptual L1 counterparts than the participants of the non-bilingual school. However it should be noted that the non-bilingual school also discriminated the /ɛ-æ/ vowel pairing poorly (51.77%), indicating they may also be in the beginning stages of L2 category formation as well, though it is predicted that the non-bilingual school participants' productions of English /ɛ/ and /æ/ will more closely resemble Spanish /e/ and /a/ because of their significantly worse perception in the /ɛ-ɛ/ and /æ-æ/ contrasts.

Q3. Is there a significant difference in how grade levels produce English vowels?

Because there was no significant effect for grade on the perception task, I predict that the grades will produce the English vowels in statistically similar ways.

Q4. Do the L1 Spanish speakers produce English vowels significantly differently from those vowels' closest Spanish counterpart?

L2 perception studies have noted the acoustic similarities between English and Spanish /i/ (e.g., Mendez, 1982) and others have noted the perceptual similarities as well (e.g., Flege, 1991; Iverson & Evans, 2009). Taking both of these into account, it is likely that L1 Spanish speakers learning English will have created a diaphone – a category that is the same in both languages (Flege, 1995). It is unlikely, therefore, that this dissertation's L1 Spanish speakers will produce L1 Spanish /i/ and L2 English /i/ significantly differently.

The SLM suggests that English /ɪ/ could be a new sound after enough exposure to the L2 because it does not assimilate strongly onto any L1 category (Bohn & Flege, 1997; Flege, 1991; Flege, 1995). This, coupled with the current participants' ability to discriminate between English

/ɪ/ and /i/ suggests that they will be able to produce English /ɪ/ significantly differently from Spanish /i/.

/ɛ/ is not assimilated strongly onto any L1 sounds either (Flege, 1991), which may indicate there will not be strong negative L1 effects that inhibit the formation of a new category (Flege, 1995). Flege et al. (1997) found adult L1 Spanish speakers consistently produced highly intelligible examples of /ɛ/ and suggested that this could be due to [ɛ] being an allophone of Spanish /e/ in certain consonantal environments (Dalbor, 1980). If the participants in this dissertation are able to make the same distinctions between English /ɛ/ and its nearest L1 counterparts or recognize the similarities between English /ɛ/ and the acoustically similar Spanish [ɛ], they may be able to produce this L2 phone significantly differently than any L1 sound.

English /æ/ assimilates onto Spanish /a/ quite strongly (Flege, 1991) and L1 Spanish speakers' production of this L2 phone has been identified as sounding like English /ɑ/ (Flege, Bohn & Jang, 1997). An L2-L1 assimilation task by Iverson and Evans (2009) found that English /ɑ/ also assimilated relatively strongly onto Spanish /a/. With both English /æ/ and /ɑ/ mapping to a single L1 category PAM-L2 predicts that it will be hard for native Spanish speakers to discriminate between the two sounds, which suggests a tendency to produce them similarly also. However, it is difficult to make a firm prediction on their production of English /æ/ because this phone was also confused with English /ɛ/, suggesting that the production could be more acoustically similar to English /ɛ/ than it is to Spanish /a/ or English /ɑ/.

3.4 RESULTS

The L1 Spanish speakers' vowel production was analyzed using Praat, version 5.4.02. The PI outlined each vowel from the peak of the first repeated waveform to the peak of the last repeated waveform. To measure the F1, F2 and F3 of each of these vowel segments at the midway point, the PI used a Praat script written by David Mortenson while at the University of Pittsburgh (Mortenson, 2011). Upon running the script, if the F1 or F2 value was calculated to be more than one standard deviation away from the mean for that vowel, it was categorized as a bad token and the PI measured the formants by hand. When working by hand, the PI determined the location of the vowel's midway point and retrieved the value of the first three formants as calculated by Praat. The PI then verified the program's measurements by placing the cursor on each of the first three formants present in the spectrogram and taking the measurement manually. If there was a discrepancy between the measurements taken by Praat and the PI, the PI's measurements were used.

Table 24. *Mean Barks Measurements of English Vowels by L1 Spanish Speakers*. Mean barks measurements for all 124 native Spanish speaking participants.

Vowel	Vertical Placement (B1)		Horizontal Placement (B2)	
	Mean	SD	Mean	SD
/i/	11.471	.5677	1.451	.4972
/ɪ/	11.233	.5054	1.5263	.4073
/e/	10.162	.5050	1.510	.4858
/ɛ/	9.271	.6178	1.4102	.4650
/æ/	6.647	.7032	2.386	.5618
/ɑ/	7.006	.6406	2.927	.6582

Initially the first three vowel formants were measured in Hertz (Hz), the number of repeated cycles per second. However, Hertz measurements from multiple people will vary due to physiological differences, such as those that result from age and gender. To account for these differences in the current set of participants, the Hertz values were converted into Barks, a unit of measurement that normalizes physiological differences between individual speakers and mimics "the cognitive processes that allow human listeners to normalize vowels uttered by different speakers" (Thomas & Kendall, 2007). A vowel normalization website created by Thomas and Kendall (2007) at the University of Oregon was used to convert the Hertz into Barks, the output of which provides the vertical (B1) and horizontal (B2) location of the vowel on a vowel chart. Large B1 measurements correspond to high vowels whereas large B2 measurements correspond to back vowels. Figure 9, below, displays the distribution of vowels for the L1 English speakers who produced the English stimuli and the L1 Spanish vowels as produced by all 124 participants.

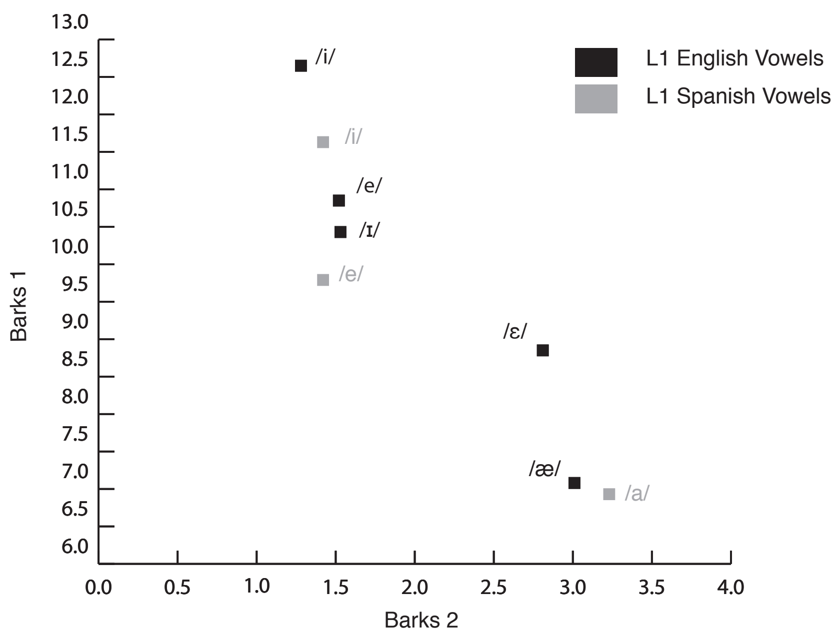


Figure 9. *L1 English Vowels and L1 Spanish Vowels.* This chart displays the distribution of vowels as they are produced by L1 English speakers and L1 Spanish speakers, measured in Barks.

Q1. Do native Spanish speakers produce the L2 English front vowels /i ɪ e ɛ æ/ and the English back vowel /a/ in significantly different ways?

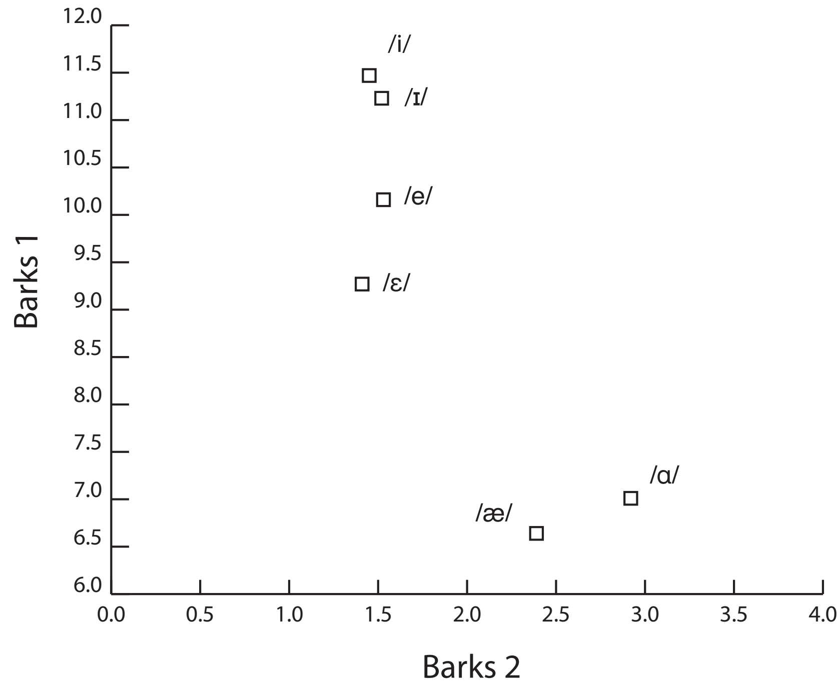


Figure 10. *Compiled English Vowel Measurements by L1 Spanish Speakers.* Graph displays the mean Barks measurements for six English vowels as produced by 124 L1 Spanish speaking participants, not distinguishing between school or grade.

Multiple repeated measures ANOVAs were performed comparing the B1 and B2 measurements of the L2 English vowels /i ɪ e ɛ æ ɑ/ to see if each vowel was produced in a significantly different way. For these comparisons, the vowels were divided into two groups based on their location in the vowel space (Figure 10): /i ɪ e ɛ/ and /ɛ æ ɑ/. English /ɛ/ was included in both groups because (1) it is the lax counterpart of /e/ and (2) because it was often confused with /æ/ in the perception task. The within-subject effects from the repeated measures ANOVA can be seen in Tables 25 and 28.

Table 25. *Comparison Barks 1 and Barks 2 Measurements in the Production of /i/-/ɪ/-/e/-ɛ/ (All participants).* (Within-Subject Effects). Two repeated measures ANOVAs were conducted. The first compared the B1 measurement of /i ɪ e ɛ/ and the second compared the B2 measurements. The results have been compiled into a single table. (*p=.05, **p=.01, ***p<.001).

Vertical Placement (B1)								
	Type III sum of Sq	df	Mean Sq.	F	Sig.	Part. Eta Sq.	Noncent. Param.	Obs. Power
Vowel	376.29	2.582	145.7	623.172	.000***	.841	1609.1	1.0
School	.307	2.582	.119	.508	.649	.004	1.313	.145
Grade	3.040	5.164	.589	2.517	.028*	.041	13.00	.793
School*Grade	1.028	5.164	.199	.851	.517	.014	4.396	.310
Error	71.253	304.69	.234					
Horizontal Placement (B2)								
Vowel	1.242	2.861	.434	2.877	.039*	.024	8.231	.670
School	2.852	2.861	.997	6.609	.000***	.053	18.905	.967
Grade	2.622	5.721	.458	3.038	.008**	.049	17.381	.898
School*Grade	.393	5.721	.069	.456	.833	.008	2.608	.182
Error	50.923	337.568	.151					

The repeated measures ANOVA comparing the Barks measurements of /i ɪ e ɛ/ (Table 25) revealed that there was a significant effect of vowel, meaning that the participants produced at least one of the English vowels with a significantly different B1 (p<.001) and B2 (p=.039). To see which of these vowels were pronounced significantly differently than the others, a series of paired t-tests were performed comparing each vowel's B1 and B2 measurements against the other's.

The first series of paired t-tests compared the vertical placement (B1) of /i/, /ɪ/, /e/ and /ɛ/ (Table 26). The results from these paired t-tests showed that each of these vowels was produced with a significantly different B1 measurement (p<.001). For example, L2 English /i/ was produced significantly higher in the vowel space than was /ɪ/, which was significantly higher than /e/, which was, in turn, significantly higher than /ɛ/. Conversely, the paired t-tests comparing the horizontal placement of these vowels (B2) revealed that, as one group, the native Spanish speakers did not significantly change the horizontal placement of English /i/, /ɪ/, /e/ and /ɛ/ when speaking (Table 27).

Table 26. *Comparison of Vertical Placement (B1) Measurements of English /i/-/ɪ/-/e/-/ɛ/.* (Paired t-test). Comparison of the B1 measurements from all 124 participants across both schools and all grades for the vowels /i/ /ɪ/ /e/ /ɛ/ (*Significance set at .0083).

	Mean Difference	Std. Dev.	Std. Error	t	df	Sig.
/i-ɪ/	.2380	.5059	.0454	5.239	123	.000*
/i-e/	1.3088	.6167	.0554	23.634	123	.000*
/i-ɛ/	2.2001	.7708	.0692	31.783	123	.000*
/ɪ-e/	1.0708	.5495	.0493	21.7	123	.000*
/ɪ-ɛ/	1.9622	.6572	.0590	33.249	123	.000*
/e-ɛ/	.8913	.6979	.0627	14.222	123	.000*

Table 27. *Comparison of Horizontal Placement (B2) Measurements of English /i/-/ɪ/-/e/-/ɛ/.* (Paired t-test). Comparison of the B2 measurements from all 124 participants across schools and grades (Significance set at .0083).

	Mean Difference	Std. Dev.	Std. Error	t	df	Sig.
/i-ɪ/	-.0754	.5230	.0470	-1.605	123	.111
/i-e/	.0801	.5792	.0520	-1.540	123	.126
/i-ɛ/	.0407	.6273	.0563	.723	123	.471
/ɪ-e/	-.0048	.5002	.0449	-.106	123	.916
/ɪ-ɛ/	.1161	.5623	.0505	2.299	123	.023
/e-ɛ/	.1208	.5156	.0463	2.609	123	.010

A repeated measures ANOVA comparing the Barks measurements of /ɛ æ ɑ/ (Table 28) revealed that there was a significant effect of vowel, meaning that the participants produced at least one of the English vowels with a significantly different B1 (p<.001) and B2 (p<.001). Next, a series of paired t-tests were performed to see which of these vowels were pronounced significantly differently than the others. The results of these analyses (Tables 29 and 30) indicate that, as a group, the 124 L1 Spanish speakers produced /ɛ/, /æ/ and /ɑ/ at significantly different heights (B1, p<.001) and with significantly different degrees of backness (B2, p<.001).

Table 28. *Comparison Barks 1 and Barks 2 Measurements in the Production of /ɛ/-/æ/-/a/.* (Within-Subject effects). Two repeated measures ANOVAs were conducted. The first compared the B1 measurement of /ɛ/-/æ/-/ɑ/ and the second compared the B2 measurements (***p*=.01, ****p*<.001).

Vertical Placement (B1)								
	Type III sum of Sq.	df	Mean Sq.	F	Sig.	Part. Eta Sq.	Noncent. Param.	Obs. Power
Vowel	485.642	1.791	271.14	962.822	.000***	.891	1724.520	1.0
School	8.016	1.791	4.476	15.893	.000***	.119	28.467	.999
Grade	8.830	3.582	2.465	8.753	.000***	.129	31.354	.998
School*Grade	.834	3.582	.233	.827	.498	.014	2.961	.248
Error	59.519	211.351	.282					
Horizontal Placement (B2)								
Vowel	142.642	1.773	80.437	348.834	.000***	.747	618.602	1.0
School	2.850	1.773	1.607	6.969	.002**	.056	12.358	.899
Grade	8.780	3.547	2.476	10.736	.000***	.154	38.077	1.0
School*Grade	1.394	3.547	.393	1.705	.158	.028	6.047	.485
Error	48.252	209.254	.231					

Table 29. *Comparison of Vertical Placement (B1) of English /ɛ/-/æ/-/a/.* (Paired t-test). Comparison of the B1 measurements from all 124 participants across schools and grades for the vowels /ɛ/-/æ/-/ɑ/. Significance set at .0167 (****p*<.001).

	Mean Difference	Std. Dev.	Std. Error	t	df	Sig.
/ɛ-æ/	2.6240	.8536	.0767	34.229	123	.000***
/ɛ-ɑ/	2.2649	.8977	.0806	28.094	123	.000***
/æ-ɑ/	-.3591	.5833	.0524	-6.855	123	.000***

Table 30. *Comparison of Horizontal Placement (B2) of English /ɛ/-/æ/-/a/.* (Paired t-test). Comparison of the B2 measurements from all 124 participants across schools and grades for the vowels /ɛ/-/æ/-/ɑ/. Significance set at .0167 (****p*<.001).

	Mean Difference	Std. Dev.	Std. Error	t	df	Sig.
/ɛ-æ/	-.9754	.7062	.0634	-15.380	123	.000***
/ɛ-ɑ/	-1.5170	.8314	.0747	-20.317	123	.000***
/æ-ɑ/	-.5416	.5532	.0497	-10.902	123	.000***

The results from Tables 25 and 28 show that, as a single group, the 124 L1 Spanish-speaking students produced each L2 English vowel significantly differently from every other L2 English vowel in at least one Barks measurement. The paired t-tests revealed that they produced the English vowels /i/, /ɪ/, /e/ and /ɛ/ with significantly different B1 measurements, meaning that each vowel was produced at a significantly distinct height. The L1 Spanish participants did not, however, produce any of these vowels with a significantly different B2 meaning that /i/, /ɪ/, /e/, and /ɛ/ were produced with only negligible differences in backness. Unlike the aforementioned vowels, /ɛ/, /æ/, and /ɑ/ were all produced with significantly different B1 and B2 measurements. The L1 Spanish speaking participants produced each of these three vowels with significantly distinct height and backness.

Q2. Is there a significant difference in how the bilingual school and non-bilingual school produce these seven English vowels?

Table 31. *Mean Barks Measurements by School.* A display of the mean Barks measurements of each vowel by school.

		Bilingual School		Non-Bilingual School	
		Mean	SD	Mean	SD
Barks 1 (Vertical)	/i/	11.5817	.5891	11.3535	.5235
	/ɪ/	11.3647	.5507	11.0920	.4119
	/e/	10.2886	.4669	10.0269	.5128
	/ɛ/	9.3535	.6214	9.1822	.6066
	/æ/	6.4243	.6515	6.8838	.6832
	/ɑ/	6.7970	.6678	7.2284	.5307
Barks 2 (Horizontal)	/i/	1.6339	.5002	1.2559	.4161
	/ɪ/	1.5677	.4115	1.4821	.4015
	/e/	1.5387	.4829	1.522	.4927
	/ɛ/	1.4360	.4403	1.3827	.4923
	/æ/	2.2645	.5572	2.5147	.5418
	/ɑ/	2.7249	.6694	3.1430	.5772

Tables 25 and 28 also revealed a significant effect of school for the B2 of /i/, /ɪ/, /e/, and /ɛ/ [$F(2.861,337.568)=6.609, p<.001$], and the B1 [$F(1.791,211.351)=15.593, p<.001$] and B2 [$F(1.773,209.254)=6.969, p=.002$] of /ɛ/, /æ/, /ɑ/. The difference in how the two schools produced the B1 of the first group of vowels was not significant [$F(2.582,304.69)=.508, p=.649$]. Individual repeated measure ANOVAs were performed with *school* as the sole Independent Variable to verify the results from Tables 25 and 28 but, primarily, to provide the necessary Error values needed to conduct the following sets of Duncan MRTs. The following results from the Duncan MRTs will identify the B1 and B2 measurements of each vowel that the two schools produce significantly differently.[10]

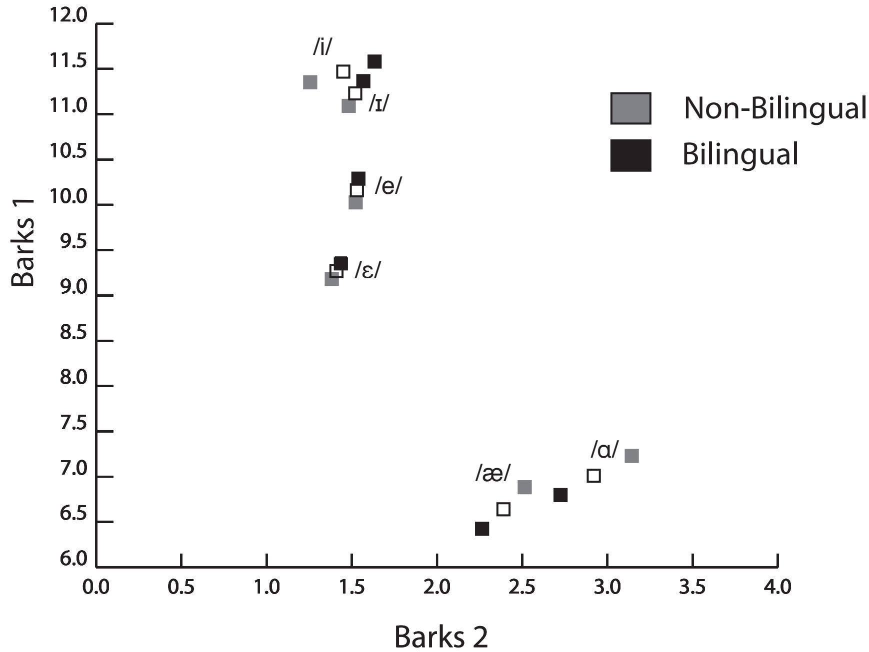


Figure 11. *Mean Barks Measurements of English Vowels by L1 Spanish Speakers by School.* This chart displays the average B1 and B2 measurements of the 124 L1 Spanish speaking participants by school and the combined average.

The first repeated measures ANOVA with *school* as the sole Independent Variable is the B2 of /i/, /ɪ/, /e/, and /ɛ/ (Table 32). We see that school continues to have a significant effect on B2 production of these vowels [$F(2.847,366)=3.029, p=.001$], so Duncan's MRT was performed (Table 33) to see how many B2 measurements were produced significantly differently by school.

---

[10] Because the repeated measures ANOVA for school*grade revealed that there was no within-subjects main effect for school in the B1 production of /i/, /ı/, /e/, and /ɛ/ (p=.649), there is no need to conduct a Duncan's MRT on those values.

Table 32. *Comparison of Barks 2 Production by School /i/-/ɪ/-/e/-/ɛ/.* (Within-Subject effects). Results from a repeated measures ANOVA where the only IV is *school*. This test was performed to provide the PI with the error values needed to conduce Duncan's MRT.

	Type II Sum of Sq.	df	Mean Sq.	F	Sig.	Part. Eta Sq.	Noncent. Par.	Obs. Power
Vowel	1.338	2.847	.470	3.029	.032*	.024	8.623	.694
School	2.549	2.847	.895	5.773	.001**	.045	16.433	.941
Error	53.871	366	.147					

Table 33. *Comparison of Barks 2 by School of English Vowels /i/-/ɪ/-/e/-/ɛ/* (Duncan's MRT). This table shows that the bilingual and non-bilingual schools produce the L2 English /i/ with significantly different B2 values (**p*=.05, ***p*=.01)

α Error Degrees of Freedom Error Mean Square Number of Means Critical Range	p=.05, p=.01 366 .147 8 [in brackets below for each comparison]			
School (Barks1 Meas.)	Non-Bilingual /i/ - (1.2559)	Non-Bilingual /ɪ/ - (1.4821)	Non-Bilingual /e/ - (1.522)	Non-Bilingual /ɛ/ - (1.3827)
Bilingual /i/ - (1.6339)	.378** [8]			
Bilingual /ɪ/ - (1.5677)		.0856 [4]		
Bilingual /e/ - (1.5387)			.0167 [2]	
Bilingual /ɛ/ - (1.4360)				.0533 [2]

There is a significant difference in the B2 of the bilingual and non-bilingual schools in the production of L2 English /i/. The non-bilingual school participants produce it significantly more forward than the bilingual school participants. The B2 values of /ɪ/, /e/ and /ɛ/ were not produced significantly differently by members of the two schools.

Table 34. *Comparison of Barks 1 Production by School /ɛ/-/æ/-/a/.* (Within-Subjects effects). Results from a repeated measures ANOVA where the only IV is *school*. This test was performed to provide the PI with the error values needed to conduct Duncan's MRT.

	Type II Sum of Sq.	df	Mean Sq.	F	Sig.	Part. Eta Sq.	Noncent. Par.	Obs. Power
Vowel	497.400	1.720	289.214	879.411	.000***	.878	1512.437	1.0
School	7.867	1.720	4.574	13.909	.000***	.102	23.921	.996
Error	69.004	244	.283					

Table 35. *Comparison of Barks 1 by School of Enlgish Vowels /ɛ/-/æ/-/a/.* (Duncan's MRT). This table shows that the bilingual and non-bilingual schools produce L2 English /æ/ and /ɑ/ with significantly different B1 values (*p=.05, **p=.01)

α Error Degrees of Freedom Error Mean Square Number of Means Critical Range	p=.05, p=.01 211.351 .283 6 [in brackets below for each comparison]		
School (Barks1 Meas.)	Non-Bilingual /ɛ/ - (9.1822)	Non-Bilingual /æ/ - (6.8838)	Non-Bilingual /ɑ/ - (7.2284)
Bilingual /ɛ/ - (9.3535)	.1713 [2]		
Bilingual /æ/ - (6.4243)		.4595** [3]	
Bilingual /ɑ/ - (6.797)			.4314** [2]

The second repeated measures ANOVA (Table 34) revealed that there is a significant effect of school on the vertical placement of /ɛ/, /æ/ or /ɑ/ [F(1.72,244)=13.909,p<.001]. The subsequent Duncan MRT (Table 35) revealed that there is no significant difference in the vertical placement of /ɛ/ between the two schools. On the other hand, the participants from the bilingual and non-bilingual schools produce /æ/ and /ɑ/ significantly differently – namely, the bilingual school produces both vowels with a significantly lower B2. The smaller B2 means that these vowels are located in a lower position in the vowel space.

Table 36. *Comparison of Barks 2 Production by School /ɛ/-/æ/-/a/.* (Within-Subjects effects). Results from a repeated measures ANOVA where the only IV is *school*. This test was performed to provide the PI with the error values needed to conduct Duncan's MRT.

	Type II Sum of Sq.	df	Mean Sq.	F	Sig.	Part. Eta Sq.	Noncent. Par.	Obs. Power
Vowel	147.882	1.708	86.577	312.130	.000***	.719	533.147	1.0
School	3.536	1.708	2.070	7.462	.001**	.058	12.747	.911
Error	57.802	244	.237					

Table 37. *Comparison of Barks 2 by School of English Vowels /ɛ/-/æ/-/a/.* (Duncan's MRT). This table shows that participants from the non-bilingual school produce /æ/ and /ɑ/ significantly further back in the mouth than the participants from the bilingual school (**p*=.05, ***p*=.01).

α Error Degrees of Freedom Error Mean Square Number of Means Critical Range	*p*=.05, *p*=.01 209.254 .237 6 [in brackets below for each comparison]		
School (Barks1 Meas.)	Non-Bilingual /ɛ/ - (1.3827)	Non-Bilingual /æ/ - (2.5147)	Non-Bilingual /ɑ/ - (3.143)
Bilingual /ɛ/ - (1.436)	.0533 [2]		
Bilingual /æ/ - (2.2645)		.2502** [2]	
Bilingual /ɑ/ - (2.7249)			.4181** [2]

The final repeated measures ANOVA (Table 36) examined the horizontal placements (B2) of /ɛ/, /æ/ and /ɑ/ and found a significant effect for school. When these B2 values were examined through the results of another Duncan MRT (Table 37), there was no significant difference in B2 measurements between schools for the L2 English vowel /ɛ/. However, it was revealed that the bilingual school produced /æ/ and /ɑ/ with significantly smaller B2 measurements, meaning that the bilingual school participants produced these L2 English categories more forward in the mouth.

Overview. When separated into groups by school, some differences in the production emerge. Of the vowels /i/, /ɪ/, /e/ and /ɛ/, the schools only differed significantly in their horizontal placement of the vowel /i/ - the non-bilingual school participants produced this sound at the same height as the bilingual school participants, but significantly more forward. The second set of Duncan MRTs revealed that participants from non-bilingual school produced /æ/ and /ɑ/ significantly higher and further back than the participants from the bilingual school. Figure 11 displays the L2 English productions in relation to the L1 English speakers that provided the stimuli for the perception task.

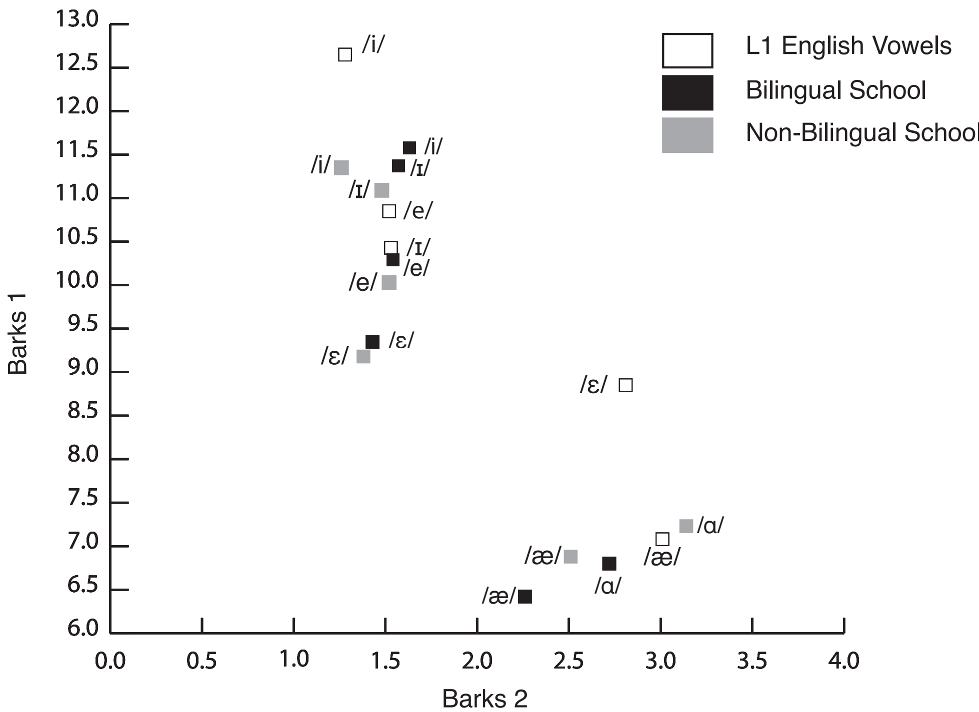


Figure 12. *L1 English Vowels and L2 English Vowels by School*. This figure illustrates the distribution of English vowels as produced by L1 English speakers and the L2 English participants by school.

Q3. Is there a significant difference in how grade levels produce English vowels?

Table 38. *Mean Barks Measurements by Grade.* Display of the mean B1 and B2 measurements for 2nd, 4th, and 6th graders across schools.

		2nd Grade		4th Grade		6th Grade	
		Mean	SD	Mean	SD	Mean	SD
/i/	Barks 1	11.3117	.4963	11.6553	.5244	11.4268	.6328
	Barks 2	1.56201	.5051	1.4421	.4990	1.3495	.4760
/ɪ/	Barks 1	11.1521	.4766	11.3577	.4903	11.1761	.5347
	Barks 2	1.6276	.2928	1.5696	.4034	1.3773	.4719
/e/	Barks 1	9.9175	.6010	10.3529	.4069	10.1963	.3963
	Barks 2	1.7286	.5294	1.3755	.3420	1.5046	.5166
/ɛ/	Barks 1	9.3193	.6395	9.3478	.6389	9.1371	.5633
	Barks 2	1.5244	.5391	1.2787	.3986	1.4406	.4269
/æ/	Barks 1	6.4160	.6964	6.6493	.6480	6.8744	.7097
	Barks 2	2.2281	.4717	2.3711	.4791	2.5590	.6806
/ɑ/	Barks 1	6.6350	.4521	7.0407	.4852	7.3380	.7581
	Barks 2	2.5080	.4530	2.9677	.4720	3.3018	.7680

Tables 25 and 28 also showed a significant effect of grade on the participants' production of B1 and B2 for both sets of vowels (/i ɪ e ɛ/ and /ɛ æ ɑ/). This means that certain grades produced at least one aspect of one of these vowels significantly differently from the other grades. Individual repeated measures ANOVA were performed with grade as the sole Independent Variable to verify the results from Tables 25 and 28 and provide the necessary Error values needed to conduct the sets of Duncan MRTs found below. The results from these Duncan MRTs identify the B1 and B2 measurements of each vowel that were produced significantly differently by at least one grade. To see a mapping of the grades' vowels, see Figure 13 below.

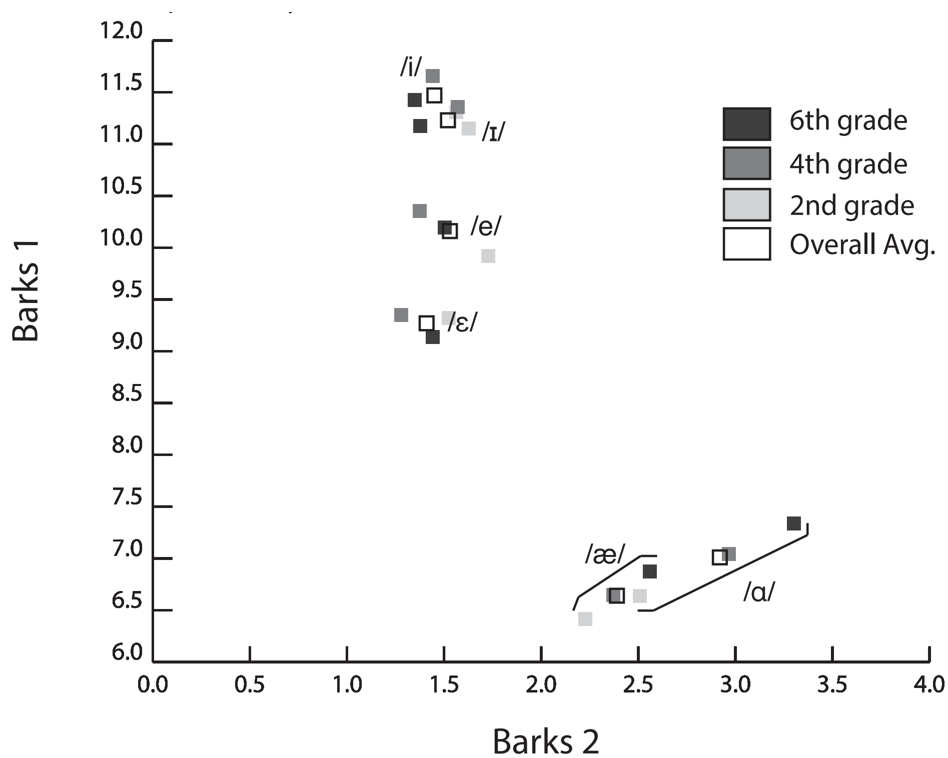


Figure 13. *Mean Barks Measurements of English Vowels by L1 Spanish Speakers by Grade.* Graphical representation of the mean B1 and B2 for the 2nd, 4th, and 6th graders. As a reference, overall average is provided.

Table 39. *Comparison of Barks 1 Production by Grade /i/-/ɪ/-/e/-/ɛ/.* (Within-Subjects effects) Results from a repeated measures ANOVA where the only IV is *grade* (**p*=.05, ***p*=.01).

	Type II Sum of Sq.	df	Mean Sq.	F	Sig.	Part. Eta Sq.	Noncent. Par.	Obs. Power
Vowel	382.917	2.588	147.980	638.498	.000***	.841	1652.193	1.0
Grade	2.810	5.175	.543	2.343	.039*	.037	12.126	.759
Error	72.565	363	.200					

Table 40. *Comparison of Barks 1 Production by Grade of English Vowels /i/-/ɪ/-/e/-/ɛ/* (Duncan's MRT). This table displays the comparison of each grade's B1 value against the other grades'. From this we see where significant difference in production occur (**p*=.05, ***p*=.01)

α	p=.05, p=.01
Error Degrees of Freedom	363
Error Mean Square	.200
Number of Means	12
Critical Range	[in brackets below each comparison value]

		i		ɪ		e		ɛ	
	Grade (Bark meas.)	4th (11.66)	6th (11.42)	4th (11.36)	6th (11.18)	4th (10.35)	6th (10.19)	4th (9.35)	6th (9.14)
i	2nd (11.31)	.344** [5]	.115 [3]						
	4th (11.66)		.229* [2]						
ɪ	2nd (11.15)			.206 [4]	.024 [2]				
	4th (11.36)				.1816 [3]				
e	2nd (9.92)					.435** [3]	.279** [2]		
	4th (10.35)						.1566 [2]		
ɛ	2nd (9.32)							.028 [2]	.182 [2]
	4th (9.35)								.211* [3]

The results from Table 39 confirmed the significant within-subjects effect for grade [F(5.175,363)=2.343,p=.039], meaning that a series of Duncan MRTs could be performed to determine which aspects of these vowels were produced significantly differently by grade. Table 40, which compares the B1 measurements by vowel and grade, shows that the 4th graders produce a significantly higher B1 than the 2nd and 6th graders on the L2 English vowel /i/. There was no difference in the B1 production of vowel /ɪ/ amongst the grades. The 2nd graders, however, did produce an L2 English /e/ that was significantly lower than the 4th and 6th graders. In turn, the 6th graders produced /ɛ/ with a significantly lower B1 value than the 4th graders.

Table 41. *Comparison of Barks 2 Production by Grade /i/-/ɪ/-/e/-/ε/* (Within-Subjects Effects) Results from a repeated measures ANOVA where the only IV is *grade* (**p*=.05).

	Type II Sum of Sq.	df	Mean Sq.	F	Sig.	Part. Eta Sq.	Noncent. Par.	Obs. Power
Vowel	1.270	2.818	.451	2.836	.041*	.023	7.991	.659
Grade	2.206	5.637	.391	2.462	.027*	.039	13.875	.810
Error	54.214	341.017	.159					

Table 42. *Comparison of Barks 2 by Grade of English Vowels /i/-/ɪ/-/e/-/ε/* (Duncan's MRT). This table displays the results of Duncan's MRT that compared each grade's B2 values against every other grade's B2 values. From this we see where significant differences in production occur (**p*=.05, ***p*=.01).

α Error Degrees of Freedom Error Mean Square Number of Means Critical Range	p=.05, p=.01 341 (infinity) .159 12 [in brackets below each comparison value]

		i		ɪ		e		ε	
Grade (Bark meas.)		4th (1.44)	6th (1.35)	4th (1.57)	6th (1.38)	4th (1.38)	6th (1.51)	4th (1.28)	6th (1.44)
i	2nd (1.56)	120 [4]	.213* [8]						
i	4th (1.44)		.093 [5]						
ɪ	2nd (1.63)			.058 [2]	.250** [8]				
ɪ	4th (1.57)				.192 [7]				
e	2nd (1.73)					.353** [10]	.224* [6]		
e	4th (1.38)						.129 [5]		
ε	2nd (1.52)							.246* [8]	.084 [4]
ε	4th (1.28)								.162 [5]

The results from Table 41 confirm that there is a significant effect for grade on the B2 productions of the first set of L2 English vowels: /i/, /ɪ/, /e/, and /ε/ [F(5.637,341.017)=2.462,p=.027]. The results from Duncan MRTs (Table 42) revealed that the 2nd graders produced significantly higher B2 measurements that the 6th graders on the English vowels /i/, /ɪ/ and /e/, meaning that the 6th graders produced these English vowels more forward in the mouth than the 2^{nd} graders. The 2^{nd} graders also produced English /e/ and /ɛ/ significantly further back than the fourth graders.

Table 43. *Comparison of Barks 1 Production by Grade /ɛ/-/æ/-/a/.* (Within-Subjects effects). Results from a repeated measures ANOVA where the only IV is *grade* (****p*<.001).

	Type II Sum of Sq.	df	Mean Sq.	F	Sig.	Part. Eta Sq.	Noncent. Par.	Obs. Power
Vowel	500.107	1.717	291.351	885.733	.000***	.880	1520.37	1.0
Grade	8.551	3.433	2.491	7.572	.000***	.111	25.996	.993
Error	68.32	242	.282					

Table 44. *Comparison of Barks 1 by Grade of English Vowels /ɛ/-/æ/-/a/* (Duncan's MRT). This table displays the results of Duncan's MRT comparing each grade's B1 values against the other grades' B1 production. From this we see where significant differences in production occur by grade (**p*=.05, ***p*=.01).

<table>
<tr><td colspan="3">α
Error Degrees of Freedom
Error Mean Square
Number of Means
Critical Range</td><td colspan="5">p=.05, p=.01
242 (infinity)
.282
9
[in brackets below each comparison value]</td></tr>
<tr><td colspan="2"></td><td colspan="2">ɛ</td><td colspan="2">æ</td><td colspan="2">ɑ</td></tr>
<tr><td colspan="2">Grade
(Bark meas.)</td><td>4^{th}
(9.348)</td><td>6^{th}
(9.137)</td><td>4^{th}
(6.649)</td><td>6^{th}
(6.874)</td><td>4^{th}
(7.041)</td><td>6^{th}
(7.338)</td></tr>
<tr><td rowspan="2">ɛ</td><td>2^{nd}
(9.319)</td><td>.0284
[2]</td><td>.1823
[2]</td><td></td><td></td><td></td><td></td></tr>
<tr><td>4^{th}
(9.348)</td><td></td><td>.2108
[3]</td><td></td><td></td><td></td><td></td></tr>
<tr><td rowspan="2">æ</td><td>2^{nd}
(6.416)</td><td></td><td></td><td>.2334
[3]</td><td>.4585**
[4]</td><td></td><td></td></tr>
<tr><td>4^{th}
(6.649)</td><td></td><td></td><td></td><td>.2251
[2]</td><td></td><td></td></tr>
<tr><td rowspan="2">ɑ</td><td>2^{nd}
(6.635)</td><td></td><td></td><td></td><td></td><td>.4057**
[4]</td><td>.7030**
[5]</td></tr>
<tr><td>4^{th}
(7.041)</td><td></td><td></td><td></td><td></td><td></td><td>.2973*
[2]</td></tr>
</table>

The results from Table 43 confirm that there is a significant effect for grade on the B2 productions of the first set of L2 English vowels: /ɛ/, /æ/ and /ɑ/ [F(3.433,242)=7.572,p<.001]. The subsequent Duncan's MRT (Table 44) found that there was no significant difference in the

B1 measurements of the vowel /ɛ/[11]. However, the 2^{nd} graders produced significantly smaller B1 values for the vowel /æ/ when compared to the 6^{th} graders and the all three grades produced significantly different B1 values for the vowel /ɑ/. Here the 2^{nd} graders produced /ɑ/ with significantly smaller B1 values than the 4^{th} graders (6.635 vs. 7.041, respectively), who, in turn, produced significantly smaller B1 values than the 6^{th} graders (7.338).

Table 45. *Comparison of Barks 2 Production by Grade /ɛ/-/æ/-/a/.* (Within-Subjects effects). Results from a repeated measures ANOVA where the only IV is *grade*.

	Type II Sum of Sq.	df	Mean Sq.	F	Sig.	Part. Eta Sq.	Noncent. Par.	Obs. Power
Vowel	145.156	1.739	83.471	334.079	.000***	.734	580.968	1.0
Grade	8.763	3.478	2.520	10.084	.000***	.143	35.073	.999
Error	52.574	242	.217					

Table 46. *Comparison of Barks 2 by Grade of English Vowels /ɛ/-/æ/-/a/.* (Duncan's MRT). This table displays the results of Duncan's MRT comparing each grade's B1 vales against the other grades' productions. From this we see where significant differences in production occur by grade (**p*=.05, ***p*=.01)

α	*p*=.05, *p*=.01
Error Degrees of Freedom	242 (infinity)
Error Mean Square	.217
Number of Means	9
Critical Range	[in brackets below each comparison value]

		ɛ		æ		ɑ	
Grade (Bark meas.)		4^{th} (1.279)	6^{th} (1.441)	4^{th} (2.371)	6^{th} (2.559)	4^{th} (2.968)	6^{th} (3.302)
ɛ	2^{nd} (1.524)	.246* [3]	.084 [2]				
	4^{th} (1.279)		.162 [2]				
æ	2^{nd} (2.228)			.143 [2]	.331** [4]		
	4^{th} (2.371)				.188 [3]		
ɑ	2^{nd} (2.508)					.460** [3]	.794** [4]
	4^{th} (2.968)						.334** [2]

[11] When compared with the vowels /i/, /ɪ/ and /e/ in Table 40, it was determined that the B1 values of /ɛ/ between the 2^{nd} and 6^{th} graders were significantly different.

Results from Table 45 confirm that there is a significant effect for grade [F(3.478,242)=10.084,p<.001]. Results from the subsequent Duncan's MRT (Table 46) revealed that the 4th graders produced /ɛ/ with a significantly smaller B1. The 6th graders produced /æ/ with a significantly larger B2 and every grade produced the B2 of the English vowel /ɑ/ significantly differently than the others – 2nd graders produced /ɑ/ with significantly smaller B2s than the 4th graders (2.508 vs 2.968, respectively), who, in turn, produced /ɑ/ with significantly smaller B2s than the 6th graders (3.302).

Overview. Through the variance of the B1 and B2 measurements between the 2nd, 4th and 6th graders, one pattern emerged more prominently than the rest – as a group, the 2nd graders appear to have a narrower vowel space than the other grades. The 2nd graders produce the mid and high front vowels further back than the 6th graders (ex. /i/ and /ɪ/), the 4th graders (ex. /ɛ/) or the 4th and the 6th graders (ex. /e/). Moreover, in their production of the low vowels /æ/ and /ɑ/, the B2 measurements were significantly more forward than just the 6th graders (ex. /æ/) or the 4th and 6th graders (ex. /ɑ/). In addition to the 2nd graders' low vowels being produced more forward, these two vowels were also produced significantly closer together than the 4th and 6th graders' vowels.[12]

A more tenuous pattern can be seen with the 2nd graders' production of B1. The 2nd graders' B1 is significantly smaller, signifying a lower placement in the vowel space, than the 4th graders (ex. /i/) or the 4th and 6th graders (ex. /e/, /æ/, and /ɑ/). In turn, the 4th and 6th graders B1 productions switched from between being non-significant (ex. /ɪ/ /e/ /æ/) to the 4th graders producing larger B1s (ex. /i/) to the 6th graders producing larger B1 (ex. /ɛ/ /ɑ/).

---

[12] The table for this repeated measures ANOVA comparing distance between /ɑ/ and /æ/ by grade can be found in Appendix B.

*Q4. Will individual school*grade groups produce certain vowels?*

The repeated measures ANOVAs that compared the B1 and B2 values of the seven L2 English vowels under investigation (Tables 25 and 28) revealed that there was no significant interaction between school and grade and the production of B1 or B2 on any individual vowel – for the vowels /i/-/ɪ/-/e/-/ɛ/ B1 [F(5.164,304.69)=.851,p=.517] B2 [F(5.721,337.568)=.456,p=.833]. For the vowels /ɛ/-/æ/-/ɑ/ B1 [(3.582,211.351)=.827, p=.498] and B2 [F(3.547, 209.254)=1.705,p=.158]. No Duncan's MRT was performed because this interaction did not reach significance in the repeated measures ANOVAs.

Q5. Do the L1 Spanish speakers produce English vowels significantly differently from those vowels' closest Spanish counterpart?

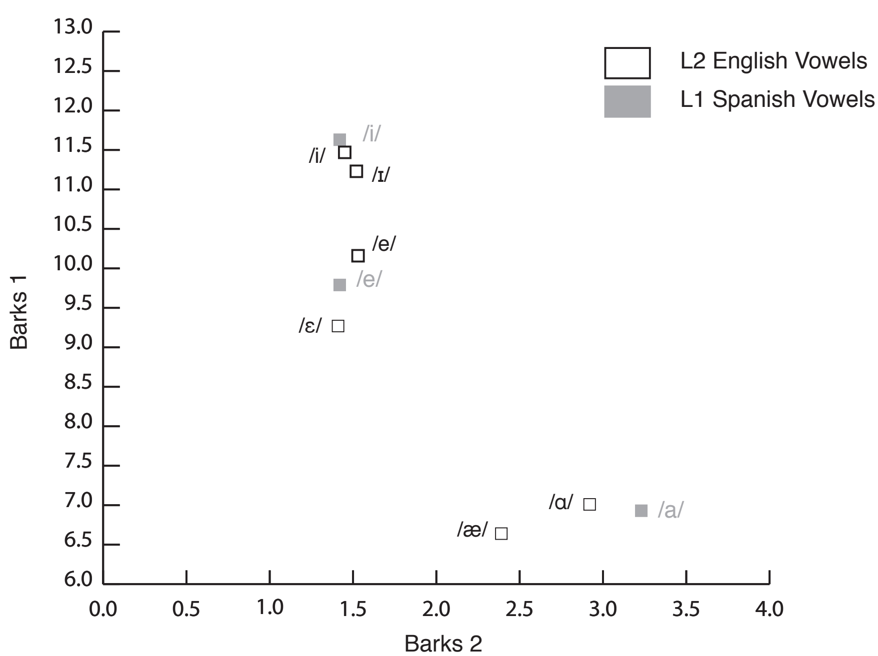


Figure 14. *Production of L1 Spanish Vowels and L2 English Vowels.* Displays the relationship between L1 Spanish vowels and L2 English vowels by all 124 participants.

Table 47. *Average B1 and B2 of English and Spanish Vowels*. Mean B1 and B2 measurements of L1 Spanish and L2 English vowels by all Spanish-speaking participants (n=124).

Vowel	L2 English				Vowel	L1 Spanish			
	Vertical Placement (B1)		Horizontal Placement (B2)			Vertical Placement (B1)		Horizontal Placement (B2)	
	Mean	SD	Mean	SD		Mean	SD	Mean	SD
/i/	11.471	.5677	1.451	.4972	/i/	11.623	.5448	1.4217	.5830
/ɪ/	11.233	.5054	1.5263	.4073					
/e/	10.162	.5050	1.510	.4858	/e/	9.788	.6349	1.4247	.5103
/ɛ/	9.271	.6178	1.4102	.4650					
/æ/	6.647	.7032	2.386	.5618					
/ɑ/	7.006	.6406	2.927	.6582	/a/	6.9322	.7973	3.2311	.6552

To determine if the L1 Spanish speakers produced Spanish and English vowels differently (Figure 14), the B1 and B2 measurements for the English vowels and their closest Spanish counterparts were analyzed in multiple repeated measure ANOVAs (Tables 48, 50, and 52). If the vowel variable was significant, paired t-tests were performed to see which vowels differed.

Table 48. *Comparison of Barks between Spanish /i/ and English /i/ and /ɪ/.* (Within-Subjects effects). Two repeated measures ANOVAs were conducted, the first compared B1 of English /i/ and /ɪ/ to Spanish /i/ and the second compared B2 measurements (****p*<.001)

	Type II Sum of Sq.	df	Mean Sq.	F	Sig.	Part. Eta Sq.	Noncent. Par.	Obs. Power
Vertical Placement (B1)								
Vowel	9.579	1.989	4.816	36.131	.000***	.277	71.860	1.0
Error	32.609	246	.133					
Horizontal Placement (B2)								
Vowel	.722	1.956	.369	2.256	.108	.018	4.413	.451
Error	39.366	246	.160					

In Table 48, the B1 and B2 values of English /i/ and /ɪ/ are compared against the B1 and B2 values of Spanish /i/ in a repeated measures ANOVA. Results from this analysis reveal that there is no difference in the horizontal placement of these three vowels, but that there is a significant difference in the B1. To locate this difference, two paired t-tests were performed comparing the English and Spanish vowels. These results are found in Table 49.

Table 49. *Comparison of Vertical Placement (B1) Measurements of English /i/-/ɪ/ and Spanish /i/.* (Paired t-test). Comparison of the B1 measurements from all 124 participants across schools and grades for English /i/ /ɪ/ and Spanish /i/ (Significance set at .025)

	Eng Vow.	Span Vow.	Mean Difference	Std. Dev.	Std. Error	t	df	Sig.
B1	/i/	/i/	-.1519	.5338	.0479	-3.169	123	.002*
B1	/ɪ/	/i/	-.3899	.5045	.0453	-8.606	123	.000*

Results from these paired t-tests (Table 49) indicate that both English /i/ and /ɪ/ are produced with significantly different B1s than Spanish /i/. In both of these cases, Spanish /i/ is produced with a larger B1, indicating that it is produced higher in the vowel space than both English /i/ and /ɪ/.

Table 50. *Comparison of Barks between Spanish /e/ and English /e/ and /ɛ/.* (Between-Subjects effects). Two repeated measures ANOVAs were conducted, the first compared the B1 measurement of English /e/ and /ɛ/ to Spanish /e/ and the second compared the B2 measurements (*p=.05, ***p<.001).

	Type II Sum of Sq.	df	Mean Sq.	F	Sig.	Part. Eta Sq.	Noncent. Par.	Obs. Power
Vertical Placement (B1)								
Vowel	49.683	1.952	25.447	99.574	.000***	.447	194.412	1.0
Error	61.372	246	.249					
Horizontal Placement (B2)								
Vowel	1.080	1.945	.555	3.656	.028*	.029	7.109	.661
Error	36.331	246	.148					

Table 51. *Comparison of Vertical Placement (B1) Measurements of English /e/ and /ɛ/ and Spanish /e/.* (Paired t-test). Comparison of the B1 and B2 measurements from all 124 participants across schools and grades (Significance set at .0125).

	Eng Vow.	Span Vow.	Mean Difference	Std. Dev.	Std. Error	t	df	Sig.
B1	/e/	/e/	.3739	.6620	.0595	6.289	123	.000*
B2	/e/	/e/	.1064	.5248	.0471	2.257	123	.026
B1	/ɛ/	/e/	-.5174	.7560	.0679	-7.621	123	.000*
B2	/ɛ/	/e/	-.0145	.5872	.0527	-.274	123	.758

Results from Table 50 indicate that there is a difference between the how the L1 Spanish speaking participants produce the B1 [$F(1.952,246)=99.574, p<.000$] and B2 [$F(1.945,246)=3.656, p<.001$] of English /e/, /ɛ/ and Spanish /e/. According to the results from

Table 51, the difference between these three vowels is found only in the B1 values. There is no difference in the horizontal placement of English /e/ (M=1.531,SD=.4857) and Spanish /e/ (M=1.4247,SD=.5103); t(123)=2.257, p=.026. There is also no difference in B2 values between English /ɛ/ (M=1.41, SD=.465) and Spanish /e/ (M=1.4247, SD=.5103); t(123)=-.274, p=.758. However, there is a significant difference in the B1 values of English /e/ (M=10.162,SD=.505) and Spanish /e/ (M=9.788,SD=.635); t(123)=6.289, p<.001. There is also a significant difference in B1 values between English /ɛ/ (M=9.271, SD=.617) and Spanish /e/ (M=9.788, SD=.635); t(123)=-7.621, p<.001.

Table 52. *Comparison of Barks between Spanish /a/ and English /æ/ and /a/.* (Within-Subjects effects). Two repeated measures ANOVAs were conducted, the first compared B1 measurements of English /æ/ and /ɑ/ to Spanish /a/ and the second compared the B2 measurements (****p*<.001).

	Type II Sum of Sq.	df	Mean Sq.	F	Sig.	Part. Eta Sq.	Noncent. Par.	Obs. Power
Vertical Placement (B1)								
Vowel	8.925	1.984	4.498	23.955	.000***	.163	47.531	1.0
Error	45.825	246	.186					
Horizontal Placement (B2)								
Vowel	45.492	1.991	22.848	157.675	.000***	.562	313.940	1.0
Error	35.488	246	.144					

Table 53. *Comparison of Vertical and Horizontal Barks Measurements of English /æ/ and /a/ and Spanish /a/.* (Paired t-test). Comparison of the B1 and B2 measurements from all 124 participants across schools and grades for English /æ/ and /ɑ/ to Spanish /a/ (Significance set at .0125)

	Eng Vow.	Span Vow.	Mean Difference	Std. Dev.	Std. Error	t	df	Sig.
B1	/æ/	/a/	-.2856	.6289	.0565	-5.057	123	.000*
B2	/æ/	/a/	-.8455	.5222	.0469	-18.030	123	.000*
B1	/ɑ/	/a/	.0735	.6180	.0555	1.324	123	.188
B2	/ɑ/	/a/	-.3039	.5355	.0481	-6.320	123	.000*

Because there is a significant effect for vowel in Table 52, a series of paired t-tests were performed (Table 53) to determine where the L1 Spanish speakers produced significantly different B1 and B2 values. The results from the paired t-test show that the L1 Spanish speaking participants produced the B1 and B2 of English /æ/ (B1: M=6.64, SD=.703; B2: M=2.38, SD=.562) significantly differently than the B1 and B2 of Spanish /a/ (M=6.93, SD=.797; B2: M=3.23, SD=.655) [B1: t(123)=-5.057, p<.001; B2: t(123)=-18.030, p<.001]. The average B1 of English /ɑ/ (M=7.006, SD=.641) was not significantly different than the B1 of Spanish /a/ (M=6.93, SD=.797), t(123)=1.324, p=.188. The B2 values of English /ɑ/ (M=2.927, SD=.658) were significantly smaller than the mean B2 values of Spanish /a/ (M=3.23, SD=.655), t(123)=-6.320, p<.001.

3.5 DISCUSSION – PRODUCTION

3.5.1 Overall vowel production

As a single group, the L1 Spanish speaking participants produced each L2 English vowel in a distinct way. They produced each high and mid front vowel (/i ɪ e ɛ/) with its own distinct height, but they did not produce the vowels with differing horizontal placement. The remaining three vowels, /ɛ æ ɑ/, were produced with significantly different horizontal and vertical placements.

Producing the /i-ɪ/ distinction based on spectral cues has been shown to be difficult for English language learners whose L1 does not make tense-lax distinctions, like Spanish (Flege, Bohn & Jang, 1997). But, because the participants from both schools perceived the within- and between-category vowel pairings containing these two phones reasonably well, it is less surprising that they produced the sounds significantly differently – and because the difference in the production was spectral (i.e. produced with significantly different heights) it is likely that they distinguished these phones in the perception task using acoustic cues as well. Findings here support results obtained by Flege (1992) who saw that a group of L1 Spanish/L2 English early learners were able to produce English /i/ and /ɪ/ differently.

As a single group, the L1 Spanish speakers were also able to produce spectral differences in their pronunciation of English /ɛ/ and /æ/. This was a bit unexpected because the participants could not perceive a difference between these two sounds in the discrimination task. One study (Fullana, 2006) found that difficulty perceiving a difference between two vowels (less than 65% accuracy) did not preclude production of those same vowels with less of a foreign-accentedness rating than the production of another vowel pairing that was discriminated with an accuracy of nearly 100%. Though no intelligibility or foreign-accentedness rating was part of the current study, results from Fullana (2006) show that difficulties in perception may not always lead to the conflation of two L2 phones in learners' interlanguage sound system.

Even though each L2 English vowel was produced in a significantly different manner from the other L2 English vowels, this does not necessarily mean that the L1 Spanish participants had native-like production of English vowels. L1 English speakers use the features of vowel height and backness to distinguish between vowels but, as the results associated with Figure 10 show, the L1 Spanish speaking participants only produced differences in height between /i ɪ e ɛ/, not backness. This seems to support a statement made by Saito (2016) that L2 learners may discriminate L2 sounds based on a different set of cues or features than are used by native speakers of the target language.

3.5.2 Vowel production by school

Because there was no difference between the schools' perception scores for the vowel pairings containing /i/ and /ɪ/ it was predicted that the schools' productions of these vowels would not differ significantly either. However, participants from the non-bilingual school produced English /i/ significantly more forward (lower B2) than the participants from the bilingual school – when compared against Spanish /i/, English /i/ has been found to be more peripheral (Bradlow, 1995), indicating that the non-bilingual production of English /i/ might be more native-like. One possible explanation for the bilingual school's production of English /i/ is the finding that as a new phonetic category gets established, an existing category may be produced inaccurately from time to time (Cebrian, 2007). However, seeing that the bilingual school's production on English /ɪ/ did not differ from the production of English /ɪ/ by the non-bilingual school, this may not be applicable. The prediction that the bilingual students' production of /ɪ/ would be slightly more native-like based on the increase of exposure, therefore, cannot be substantiated here.

In the perception task there was a significant difference between how accurately the participants from each school perceived the /ɛ-æ/ contrast as containing two distinct vowels. Participants from both schools perceived this contrast very poorly – at or below chance – and because of this inability to perceive differences between the tokens of this contrast, it was predicted that we may not see a difference in production.

For the productions of /ɛ/ and /æ/, it was predicted that the bilingual students might pronounce these L2 phones significantly differently than their non-bilingual school counterparts if they associated them less with Spanish /e/ and /a/. Results showed that there was no significant difference in the production of /ɛ/ based on school type, which could possibly be due to [ɛ] existing as an allophone of Spanish /e/ in specific consonantal environments (Dalbor, 1980;

Flege et al., 1997). It should be noted, however, that both schools' productions of English /ɛ/ were produced considerably more forward that the /ɛ/ produced by the L1 English speakers who provided the stimuli for the perception task (Figure 12).

There was a significant effect of school on the production of /æ/. The bilingual school produced this sound significantly more forward than the non-bilingual school. This also makes the bilingual school students' production of /æ/ less similar to English /æ/, as produced by the L1 English speakers (Figure 12). However, their production of /æ/ is also further away from Spanish /a/ than the non-bilinguals', which may indicate that the bilingual school students are exaggerating their production of English /æ/ in order to avoid equivalence classification with Spanish /a/ (Flege, Schirru, & MacKay, 2003).

3.5.3 Vowel production by grade

There was considerable variation in the pronunciation of English vowels produced by different grades. This was unexpected because in the perception task neither the overall perception score nor the performance on individual vowel pairings was affected significantly by the age of the participants.

One clear pattern to emerge from this effect of grade is the lack of B2 variance produced by the second graders. The second graders produced each one of the front high and front mid vowels significantly further back than either the 4th and/or 6th graders. Additionally, their productions of /æ/ and /ɑ/ were produced significantly more forward than either the 4th or 6th graders. A study by Fabra and Romero (2012) found that native Catalan speakers produced L2 English vowels less peripherally than L1 English speakers, but that the vowel space expanded as proficiency increased. The production of the second graders here seems to add support to this finding, assuming that proficiency – which wasn't measured for - improved as the students progressed through the English programs at both schools.

3.5.4 Vowel production by school*grade groups

There was no significant difference the way that individual school*grade groups produced English vowels. This was not predicted because this interaction was significant for a number of vowel pairings in the discrimination task.

3.5.5 L2 English versus L1 Spanish production

The L1 Spanish speakers produced all the L2 English vowels significantly differently than their closest Spanish counterpart. The fact that they produced English /i/ differently from Spanish /i/ was somewhat unexpected because these vowels have been found to be both perceptually and acoustically similar (Flege, 1995; Mendez, 1983; Iverson & Evans, 2009). Previous studies have found that it is often difficult to perceive differences between very similar L1-L2 phones and that these two phones could be conflated and produced in identical ways in both languages (Flege, 1987; Flege, 1995).

However, another acoustic analysis of English and Spanish vowels by Bradlow (1995) found that L1 English vowels were produced significantly more forward than their Spanish counterparts. This resulted in more peripheral front vowels and more centralized back vowels. Looking at Figures 12 and 14, we see that the participants' productions of L2 English /i/ look much more similar to Spanish /i/ than to the English /i/ produced by native speakers.

L2 English /ɪ/ was produced significantly lower than Spanish /i/, but still considerably higher than L1 English /ɪ/ is produced. Even though there is a spectral distinction made in their production between /i/ and /ɪ/ it seems to show less spectral movements between the tense-lax pair than what is produced by L1 English speakers. Flege et al. (1997) also found less spectral movement in the production of /i/ and /ɪ/ by a group of adult L1 Spanish speakers when they compared them to the productions made by L1 English speakers.

L2 English /ɛ/ was produced significantly differently than Spanish /e/, which, as explained earlier in this section, could be explained through the existence of the Spanish allophone [ɛ] (Dalbor, 1980; Flege et al., 1997). This could indicate that the L1 Spanish participants of this study perceived English /ɛ/ and Spanish [ɛ] to be equivalent or it could suggest that they simply perceived a difference between English /ɛ/ and Spanish /e/. More specific perception studies examining the differences between Spanish /e/ and its allophone [ɛ] must be done before firm conclusions can be drawn.

/æ/ is often assimilation onto Spanish /a/ and has been perceived as English /ɑ/ when produced by L1 Spanish speakers (Flege et al., 1997) – the SLM could predict, therefore, that the two phones would be produced similarly. However, the L1 Spanish participants produced L2 English /æ/ significantly more forward than the Spanish counterpart /a/. This more fronted production may suggest that the current participants perceive a difference between Spanish /a/ and English /æ/ and that their fronting of the L2 phone could be a method of maintaining English /æ/ and Spanish /a/ as two autonomous categories (Flege et al., 2003).

4.0 GENERAL DISCUSSION AND CONCLUSION

This dissertation examined how differing amounts of L2 exposure affect the perception and production of English front vowels. Results showed that the increase in exposure had a beneficial effect on the overall performance of the bilingual school students on a categorical discrimination task. Larson-Hall (2008) found that L1 Japanese speakers who were exposed to approximately 4 hours of English per week before middle school performed significantly better on a phonology task than those who were first exposed to English class in middle school. This shows that even minimal exposure, over time, has the potential to benefit one's perception of L2 sounds and supports claims by researchers that experience in an L2 makes one's perception more native-like (Bohn & Flege, 1997; Højen & Flege, 2006; Levy & Strange, 2008; Fabra, 2005). Therefore, it is still possible that the participants from the non-bilingual school have benefitted from 3-5 hours of English per week – just not to the same degree that the bilingual school students did.

Additional results from the perception task found that the 6th graders from the bilingual school did not perform significantly better than any of the other school-grade groups. This was somewhat surprising considering that Jia and Aaronson (2003) found phonological improvements in a group of immigrant children after only three years. However, perhaps the difference here lies with the amount of exposure the two groups received - 14 hours of English per week over four years may not provide comparable L2 input as to what was received by the immigrants in Jia and Aaronson's study. This difference in the amount of L2 input between students attending a bilingual school and immigrants living in the target language environment continues to uphold the claim made by White and Genesee (1996) that the amount of L2 input provided in the classroom (be it bilingual or non-bilingual) is not equal to the amount of input experienced by young immigrants living in the L2 community.

The perception task also tested the predictions of PAM-L2 (Best & Tyler, 2007) regarding the participants' ability to discriminate /i-ɪ/ and /ɛ-æ/. Using the L2-L1 assimilation patterns from a group of L1 Spanish speakers learning English (Flege, 1991), these contrasts were predicted to be relatively easy to discriminate for the current set of participants because the individual tokens of the /i-ɪ/ and /ɛ-æ/ contrasts were assimilated to the same L1 category (Spanish /i/ and /a/, respectively) at different rates, creating a *Category-Goodness Pattern*. The prediction proved correct for the /i-ɪ/ contrast, which was discriminated correctly 71.12% of the time, but it failed to capture the discrimination difficulty of the /ɛ-æ/ contrast, which was perceived at a rate lower than chance.

The discrimination task alone cannot tell us what cues the participants were using to discriminate between two sounds. In Section 2.7.2.1, it was acknowledged that the students could have been distinguishing /i/ and /ɪ/ due to their temporal differences instead of acoustic differences. Previous studies have shown that L2 learners may distinguish sounds of the target language using cues different from those used by native speakers of the L2 (Saito, 2016). Moreover, L1 Spanish speakers have been shown to use duration, not acoustic cues, in their discrimination of these two categories (Barrios et al., 2016; Hillenbrand et al., 2000). If the L1 Spanish speaking participants relied solely on duration to distinguish /i/ and /ɪ/, we would expect to see them produce these vowels with statistically similar B1 and B2 measurements – but we see that they produced English /i/ and /ɪ/ at significantly different heights. Because they produced these vowels in significantly different ways, it is probable that they were also using acoustic cues in the perception task. Because of the participants' reliable performance in the perception task (both within- and between category pairings) and their distinct productions of /i/ and /ɪ/, it is possible that students from both schools have established a new category for L2 English /ɪ/.

Results here differ from previous studies that have found that the /i-ɪ/ contrast is difficult for L1 Spanish speakers to perceive and produce based on spectral cues (Bohn, 1995; Flege, Bohn & Jang, 1997; Gulinello, 2010). This difference may be explained by the fact that they used adult participants and the current study used children - as studies have shown (e.g., Bond & Adamescu, 1979) younger language learners can be better at perceiving non-native sounds than adults, perhaps due to the less firmly-established L1 phonetic categories (Baker et al., 2008; Flege, 1995).

The participants from the bilingual school identified the vowel pairings /ɛ-ɛ/ and /æ-æ/ as consisting of the same phone significantly more reliably than participants from the non-bilingual school, though both groups averaged scores of approximately 70% or above. Their ability to discriminate between these two English phones, on the other hand, did not support PAM-L2's prediction of relative ease - neither group perceived a difference between /ɛ/ and /æ/ at a rate much better than chance. Therefore, it was a surprise to see that both groups produced /ɛ/ significantly differently from how they produced /æ/.

In addition to being produced differently from each other, the participants' production of English /ɛ/ and /æ/ were produced significantly differently from Spanish /e/ and /a/. L2 English /ɛ/ was produced significantly lower than Spanish /e/ and L2 English /æ/ was produced significantly more forward that Spanish /a/. This could suggest that these L2 English phones occupy previously uninhabited areas of the L1 Spanish vowel space between Spanish /e/ and /a/, and could provide an explanation as to why participants from neither school could discriminate the /ɛ-æ/ contrast well – if these L2 English phones were no longer strongly associated with Spanish /e/ or Spanish /a/ (as evidenced through their distinct pronunciation) perhaps the students were in the process of establishing new L2 phonetic categories through the division of an uncategorized area of L1 Spanish vowel space.

The picture-naming task revealed that, as a single group, the participants of this study produced each of the English vowels /i ɪ e ɛ æ ɑ/ in a significantly different way. When production of these vowels was examined by school, it was revealed that participants from both schools produced English /ɪ/, /e/ and /ɛ/ with non-significant differences. It is unknown, however, if the differences between groups in the productions of /i/, /æ/ and /ɑ/ have an impact on the speakers' intelligibility since L1 English speakers did not assess the L1 Spanish participants' productions for accentedness or intelligibility.

It is also important to point out that simply because each L2 English vowel was produced in a significantly different way than the other five, it doesn't mean that each vowel was produced in a native-like fashion. The participants tended to favor height in differentiating pronunciations between /i ɪ e ɛ/ and less on backness, unlike L1 English speakers who generally produce differences in both height and backness in their productions of these vowels.

In addition to the unique pronunciation between English vowels, it was revealed that the English vowels also differed significantly from their closest Spanish counterparts. Because English and Spanish /i/, /e/, and /a/-/ɑ/ have been found to be perceptually similar, the SLM predicts that language learners will produce these phones the same way in both languages - this production difference, therefore, was unexpected.

The connection between perception and production can be seen most clearly via the L2 English phones /i/ and /ɪ/. The participants' consistent performance in the categorical discrimination task indicated they were able to discern differences between the two L2 phones; the fact that they produced both of these phones differently suggests that perhaps they have established a new category for English /ɪ/. The connection between perception and production of the phones /ɛ/ and /æ/, however, is more tenuous – the students consistently disregarded insignificant within-category differences in the /ɛ-ɛ/ and /æ-æ/ pairings but were unable to perceive the relevant acoustic differences when paired together, /ɛ-æ/. They were, however, still be able to produce the /ɛ/ and /æ/ in significantly different ways that were also significantly different than the nearest L1 counterpart. These differences may show that the participants do not have clearly established L2 English categories for /ɛ/ and /æ/ - but that they are in the process of figuring out what the defining cues for each are.

Even though age effects were not observed in the perception task (i.e. no one grade perceived the vowel pairings more accurately than any other), there were significant differences between grades in production. The second graders, regardless of school, produced /i/, /ɪ/, /e/ and /ɛ/ significantly further back than either the 4th and/or 6th graders. Moreover, the 2nd graders' production of /æ/ was significantly more forward than the 6th graders' and their production of /ɑ/ was significantly more forward than both the 4th and 6th graders. This pattern seems to support the claim Fabra and Romero (2012) made after observing L1 Catalan speakers learning English – as language learners gained more experience and became more proficient, their vowel space expanded to become more like the native English speakers' productions.

With the increased emphasis on foreign language education, this dissertation aimed to examine the effects of increased L2 input on the perception and production of L1 Spanish speakers who attended either a Spanish-English bilingual school or a non-bilingual school. The increase in L2 exposure provided by the bilingual school proved to be significantly beneficial to the overall perception of English front vowels, but the effect was small. In terms of production, both groups of students were able to produce the five English front vowels in significantly different ways but, because no follow-up tasks were conducted to assess the intelligibility of their pronunciations, it cannot be said with authority whether an increase in L2 input improved the bilingual school's production of L2 English vowels. Even though the results from this dissertation exposed the benefit of increased L2 exposure in the bilingual school only for perception, it is possible that the increased exposure also benefitted other aspects of L2 acquisition that were not measured such as vocabulary growth, morphology and syntax. Before any firm conclusions can be made weighing the educational benefits against the costs, all linguistic areas should be examined.

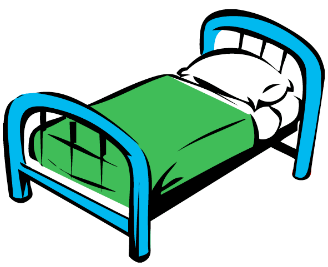
5.0 SUGGESTIONS FOR FUTURE RESEARCH

In this study it was determined that students who attend a bilingual school perceive L2 English front vowels more accurately than similar students who do not attend a Spanish-English bilingual school. The students at both schools were matched to have similar L1 backgrounds and age of first significant exposure (AOA) but differed in the amount of L2 English they received. In the future, it will be crucial to work with monolingual Spanish students of similar ages to determine if the minimal amount of exposure received by the non-bilingual students is in fact enough to perform significantly better on English perception and production tasks.

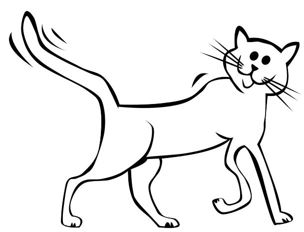
Additionally, it will be very useful to continue this cross-sectional study to include more age groups. As the students get older and progress through their respective English programs, their L1 categories will become more defined and it will be interesting to see if (and at what point) the bilingual students' perception scores continue to rise and if the non-bilingual students' perception scores continue to fall.

APPENDIX A

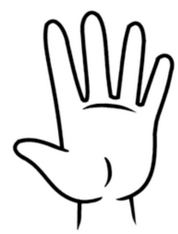
STIMULI FOR PICTURE-NAMING TASK



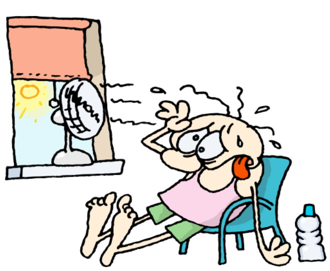
b______ ______



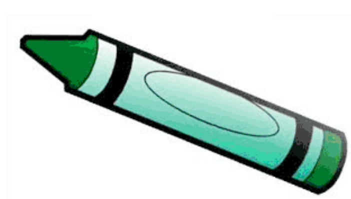
c _____ ______



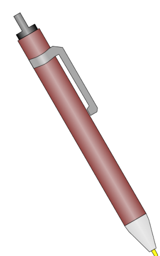
h ______ ______ ______



h ______ _______



g r ______ ______ ______

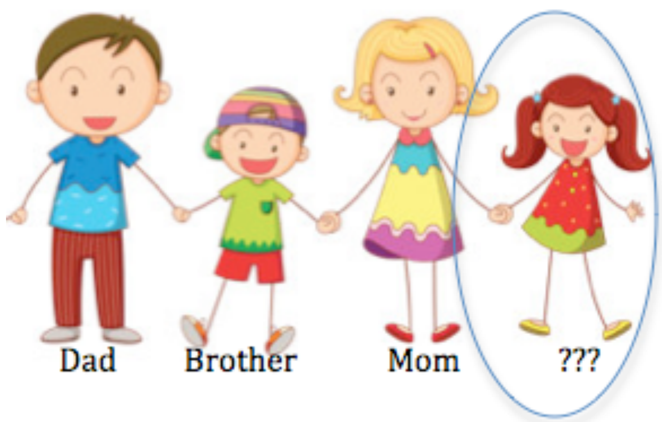


p ______ ______

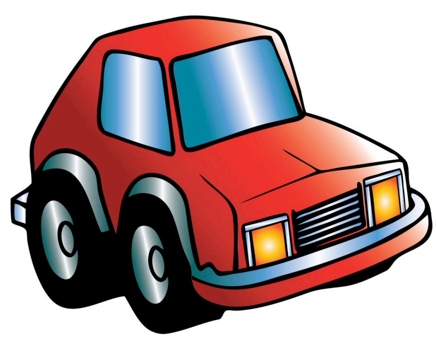
p ______ ______ ______ ______

b ______ ______ ______

b ______ ______



s ______ ______ ______ ______ ______



t ______ ______ ______ ______ ______ ______

c ______ ______

APPENDIX B

STATISTICS FOR DISTANCE BETWEEN /æ/ AND /ɑ/

Table 54. *Mean Distance by Grade*. Mean distance by grade with standard deviations.

	Mean	Std. Deviation
2^{nd}	.63839	.450994
4^{th}	.93119	.501658
6^{th}	1.05720	.597938

Table 55. *Comparing Distance between L2 English /æ/ and /a/ by Grade.* Comparing the distance between two vowels by grade with specific purpose of seeing if the 2nd graders' production were closer together than the other grades (***p*=.01).

	Type III Sum of Sq.	Df	Mean Square	F	Sig.	Partial Eta Sq.	Noncent. Param.	Obs. Power
Intercept	94.875	1	94.875	351.092	.000	.744	351.092	1.000
Grade	3.705	2	1.853	6.856	.002**	.102	13.712	.916
Error	32.698	121	.270					

Table 56. *Post-Hoc Analysis of Vowel Distance between English /æ/ and /a/.* Results show 2nd graders produce these two English sounds closer together than the 4th and 6th graders (**p*=.05, ***p*=.01).

Grade 1	Grade 2	Mean Difference	Std. Error	Sig.
2	4	-.29279	.113566	.033*
2	6	-.41881	.116238	.001**
4	6	-.12601	.113566	.808

BIBLIOGRAPHY

Abrahamsson, N. (1999). Vowel epenthesis of/sC (C)/onsets in Spanish/Swedish interphonology: A longitudinal case study. *Language learning*, *49*(3), 473-508.

Abrahamsson, N. (2003). Development and recoverability of L2 codas.*Studies in second language acquisition*, *25*(03), 313-349.

Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, *59*(2), 249-306.

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese/r/and English/l/and/r. *Journal of Phonetics*, *32*(2), 233-250.

Baker, W., Trofimovich, P., Flege, J. E., Mack, M., & Halter, R. (2008). Child—Adult Differences in Second-Language Phonological Learning: The Role of Cross-Language Similarity. *Language and Speech*, *51*(4), 317-342.

Barrios, S., Jiang, N., & Idsardi, W. J. (2016). Similarity in L2 Phonology: Evidence from L1 Spanish late-learners' perception and lexical representation of English vowel contrasts. *Second Language Research*, *1*, 29.

Bassetti, B. (2006). Orthographic input and phonological representations in learners of Chinese as a foreign language. *Written Language & Literacy*,*9*(1), 95-114.

Best, C. T. (1995). Chapter• 6 a^ A Direct Realist View of Cross-Language Speech Perception. *Speech perception and linguistic experience: Issues in cross-language research*, 171-204.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of phonetics*,*20*(3), 305-330.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. *Language experience in second language speech learning: In honor of James Emil Flege*, 13-34.

Bohn, O. S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. *Speech perception and linguistic experience: Issues in cross-language research*, 279-304.

Bohn, O. S., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, *11*(03), 303-328.

Bohn, O. S., & Flege, J. E. (1997). Perception and production of a new vowel category by adult second language learners. *Second-language speech: Structure and process*, 53-73.

Bohn, O. S., & Steinlen, A. K. (2003, August). Consonantal context affects cross-language perception of vowels. In *Proceedings of the 15th International Congress of phonetic Sciences* (pp. 2289-2292).

Bond, Z. S., & Adamescu, L. (1979). Identification of novel phonetic segments by children, adolescents and adults. *Phonetica*, *36*(3), 182-186.

Bongaerts, T. (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. *Second language acquisition and the critical period hypothesis*, 133-159.

Borden, G., Gerber, A., & Milsark, G. (1983). PRODUCTION AND PERCEPTION OF THE/r/-/l/CONTRAST IN KOREAN ADULTS LEARNING ENGLISH. *Language Learning*, *33*(4), 499-526.

Bradlow, A. R. (1995). A comparative acoustic study of English and Spanish vowels. *The Journal of the Acoustical Society of America*, *97*(3), 1916-1924.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. I. (1999). Training Japanese listeners to identify English/r/and/l: Long-term retention of learning in perception and production. *Perception & psychophysics*, *61*(5), 977-985.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. I. (1997). Training Japanese listeners to identify English/r/and/l: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*(4), 2299-2310.

Bruning, J. L., & Kintz, B. L. (1987). *Computational handbook of statistics* . Scott, Foresman & Co.

Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011). Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Applied Psycholinguistics*, *32*(01), 51-67.

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, 34, 372-387.

Cebrian, J. (2007). Old sounds in new contrasts: L2 production of the English tense-lax vowel distinction. In *Poster presented at the 16 th International Congress of Phonetic Sciences, Universitat des Saarlandes, Germany. Retrieved December* (Vol. 30, p. 2007).

Cebrian, J. (2008). The effect of perceptual factors in the acquisition of an L2 vowel contrast. *Contrast in Phonology: Theory, Perception, Acquisition*, *13*, 303-321.

Cenoz, J., & Lecumberri, L. G. (1999). The effect of training on the discrimination of English vowels. *IRAL-International Review of Applied Linguistics in Language Teaching*, *37*(4), 261-276.

Collier, V. P.. (1989). How Long? A Synthesis of Research on Academic Achievement in a Second Language. *TESOL Quarterly*, *23*(3), 509–531. http://doi.org.pitt.idm.oclc.org/10.2307/3586923

Conboy, B. T., & Kuhl, P. K. (2011). Impact of second language experience in infancy: brain measures of first and second language speech perception.*Developmental science*, *14*(2), 242-248.

Consejería de Educación, Juventud y Deporte (2016) Procedimientos para la Habilitación lingüística y Acreditación para imparter el currículo de Inglés Avanzado. Retrieved from http://www.madrid.org/cs/Satellite?blobcol=urldata&blobheader=application%2Fpdf&blobheadername1=Content-disposition&blobheadername2=cadena&blobheadervalue1=filename%3DFolleto_A4_v2.pdf&blobheadervalue2=language%3Des%26site%3DPortalEducacionRRHH&blobkey=id&blobtable=MungoBlobs&blobwhere=1352873782568&ssbinary=true

Dalbor, J. B. (1980). *Spanish pronunciation: theory and practice: an introductory manual of spanish phonology and remedial drill*. Harcourt College Pub.

Darcy, I., & Krüger, F. (2012). Vowel perception and production in Turkish children acquiring L2 German. *Journal of Phonetics*, *40*(4), 568-581.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in second language acquisition*, *22*(04), 499-533.

DeKeyser, R., & Larson-Hall, J. (2009). What does the critical period really mean. *In Handbook of bilingualism: Psycholinguistic approaches*, 88-108.

Donato, R., & Tucker, G. R. (2010). *A tale of two schools: Developing sustainable early foreign language programs* (Vol. 20). Multilingual Matters.

Ekstrand, L. H. (1976). Social and Individual Frame Factors in L2 Learning: Comparative Aspects.

Escudero, P. (2006). The phonological and phonetic development of new vowel contrasts in Spanish learners of English. *English with a Latin Beat: Studies in Portuguese/Spanish-English Interphonology*, 41-55.

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, *26*(04), 551-585.

Fabra, L. R. (2005). Predicting ease of acquisition of L2 speech sounds: A perceived dissimilarity test. *VIAL, Vigo international journal of applied linguistics*, (2), 75-92.

Fabra, L. R., & Romero, J. (2012). Native Catalan learners' perception and production of English vowels. *Journal of Phonetics*, *40*(3), 491-508.

Fabiano-Smith, L., & Goldstein, B. A. (2010). Phonological acquisition in bilingual Spanish–English speaking children. *Journal of Speech, Language, and Hearing Research*, *53*(1), 160-178.

Flege, J. E. (1987). The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of phonetics*, *15*(1), 47-65.

Flege, J. E. (1991). The interlingual identification of Spanish and English vowels: Orthographic evidence. *The Quarterly Journal of Experimental Psychology*, *43*(3), 701-731.

Flege, J. E. (1992). Speech learning in a second language. *Phonological development: Models, research, implications*, 565-604.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, 233-277.

Flege, J. E. (2009). Give input a chance. *Input matters in SLA*, 175-190.

Flege, J. E., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Tsukada, K. (2006). Degree of foreign accent in English sentences produced by Korean children and adults. *Journal of Phonetics*, *34*(2), 153-175.

Flege, J. E., & Bohn, O. S. (1989). An instrumental study of vowel reduction and stress placement in Spanish-accented English. *Studies in second language acquisition*, *11*(01), 35-62.

Flege, J. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of phonetics*, *25*(4), 437-470.

Flege, J. E., & Eefting, W. (1986). Linguistic and developmental effects on the production and perception of stop consonants. *Phonetica*, *43*(4), 155-171.

Flege, J. E., & Liu, S. (2001). THE EFFECT OF EXPERIENCE ON ADULTS'ACQUISITION OF A SECOND LANGUAGE. *Studies in second language acquisition*, *23*(04), 527-552.

Flege, J. E., & MacKay, I. R. (2004). Perceiving vowels in a second language. *Studies in second language acquisition*, *26*(01), 1-34.

Flege, J. E., MacKay, I. R., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America*, *106*(5), 2973-2987.

Flege, J. E., Munro, M. J., & MacKay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, *97*(5), 3125-3134.

Flege, J. E., Schirru, C., & MacKay, I. R. (2003). Interaction between the native and second language phonetic subsystems. *Speech communication*,*40*(4), 467-491.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of memory and language*, *41*(1), 78-104.

Foreign language learning statistics. (2016, March 17). Retrieved from http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Foreign_language_learning_statistics

Fowler, C. A., Brown, J. M., Sabadini, L., & Weihing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language*, *49*(3), 396-413.

Fullana, N. (2006). The development of English (FL) perception and production skills: Starting age and exposure effects. *Age and the rate of foreign language learning*, *19*, 41-64.

Galunov, V. I., & CHISTOVI. LA. (1966). RELATIONSHIP OF MOTOR THEORY TO GENERAL PROBLEM OF SPEECH RECOGNITION (REVIEW). *SOVIET PHYSICS ACOUSTICS-USSR*, *11*(4), 357.

Genesee, F. (1978). Is there an optimal age for starting second language instruction?. *McGill Journal of Education/Revue des sciences de l'éducation de McGill*, *13*(002).

Genesee, F. (1987). *Learning through two languages: Studies of immersion and bilingual education*. Newbury house publishers.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia*, *9*(3), 317-323.

Gottfried, T. L. (1984). Effects of consonant context on the perception of French vowels. *Journal of Phonetics*.

Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, *29*(3), 311-343.

Guion, S. G., Flege, J. E., & Loftin, J. D. (2000). The effect of L1 use on pronunciation in Quichua–Spanish bilinguals. *Journal of Phonetics*, *28*(1), 27-42.

Gulinello, F. (2010). A longitudinal investigation of vowel acquisition. In *Proceedings of the 1st Pronunciation in Second Language Learning and Teaching Conference* (pp. 90-104).

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, *24*(04), 495-522.

Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of phonetics*, *28*(4), 377-396.

Hillenbrand, J. M., Clark, M. J., & Houde, R. A. (2000). Some effects of duration on vowel recognition. *The Journal of the Acoustical Society of America*, *108*(6), 3013-3022.

Højen, A., & Flege, J. E. (2006). Early learners' discrimination of second-language vowels. *The Journal of the Acoustical Society of America*, *119*(5), 3072-3084.

Ingram, J. C., & Park, S. G. (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of phonetics*, *25*(3), 343-370.

Ioup, G., Boustagui, E., El Tigi, M., & Moselle, M. (1994). Reexamining the critical period hypothesis. *Studies in second language acquisition*, *16*(01), 73-98.

Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakersa). *The Journal of the Acoustical Society of America*,*126*(2), 866-877.

Jeske, A. (2012) The Perception of English vowels by native Spanish speakers. (Unpublished *Masters Thesis*). University of Pittsburgh.

Jia, G., & Aaronson, D. (2003) "A longitudinal study of Chinese children and adolescents learning English in the US," Appl. Psycholing. 24, 131– 161.

Jia, G., Strange, W., Wu, Y., Collado, J., & Guan, Q. (2006). Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposurea). *The Journal of the Acoustical Society of America*, *119*(2), 1118-1130.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive psychology*, *21*(1), 60-99.

Krashen, S. D., Long, M. A., & Scarcella, R. C. (1979). Age, rate and eventual attainment in second language acquisition. *Tesol Quarterly*, 573-582.

Kuhl, P. K. (1994). Learning and representation in speech and language.*Current opinion in neurobiology*, *4*(6), 812-822.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606-608.

Ladefoged, P., & Johnson, K. (2011). *A course in phonetics*. Nelson Education.

Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese.*Applied Psycholinguistics*, *26*(02), 227-247.

Larson-Hall, J. (2008). Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. *Second language research*, *24*(1), 35-63.

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. Routledge.

Leather, J. (1999). Second-Language Research: An Introduction. *Language Learning*, *49*(s1), 1-56.

Lecumberri, M. G., & Gallardo, F. (2003). English FL sounds in school learners of different ages. *Age and the acquisition of English as a foreign language*, *135*.

Lee, A. H., & Lyster, R. (2015). The effects of corrective feedback on instructed L2 speech perception. *Studies in Second Language Acquisition*, 1-30.

Lenneberg, E. H., Chomsky, N., & Marx, O. (1967). *Biological foundations of language* (Vol. 68). New York: Wiley.

Levy, E. S. (2009). Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French. *The Journal of the Acoustical Society of America*, *125*(2), 1138.

Levy, E. S., & Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics*, *36*(1), 141-157.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English/r/and/l: A first report. *The Journal of the Acoustical Society of America*, *89*(2), 874-886.

Long, M. H. (1990). Maturational constraints on language development.*Studies in second language acquisition*, *12*(03), 251-285.

Maddieson, I., & Disner, S. F. (1984). *Patterns of sounds*. Cambridge university press.

Marckwardt, A. H. (1946). Phonemic structure and aural perception.*American Speech*, *21*(2), 106-111.

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., Xu, F., & Clahsen, H. (1992). Overregularization in language acquisition. *Monographs of the society for research in child development*, i-178.

Mayo, M. D. P. G. (2003). Age, Length of Exposure and Grammaticality Judgements in the Acquisition of English as a Foreign. *Age and the acquisition of English as a foreign language*, *4*, 94.

Mayr, R., & Escudero, P. (2010). Explaining individual variation in L2 perception: Rounded vowels in English learners of German. *Bilingualism: Language and Cognition*, *13*(03), 279-297.

McAllister, R. (1997). Perceptual foreign accent: L2 users' comprehension ability. *Second-language speech: Structure and process*, 119-132.

Meisel, J. M. (2009). Second language acquisition in early childhood. *Zeitschrift für Sprachwissenschaft*, *28*(1), 5-34.

Mendez, Antonio (1982) Production of American English vowels. *Language and Speech*, 25(2), 191-197.

Ministerio de Educación, Cultura y Deporte (2016, March 11) Retrieved from http://www.mecd.gob.es/eeuu/convocatorias-programas/convocatorias-eeuu/auxiliares-conversacion-eeuu.html

Morrison, G. S. (2002). *EFFECTS OF L1 DURATION EXPERIENCE ON JAPANESE AND SPANISH LISTENERS'PERCEPTION OF ENGLISH HIGH FRONT VOWELS* (Doctoral dissertation, SIMON FRASER UNIVERSITY).

Mortenson, David. (2011) VowelFormant_Script. Available under the GNU General Public License. Accessed September 2012.

Moyer, A. (1999). Ultimate attainment in L2 phonology. *Studies in second language acquisition*, *21*(01), 81-108.

Muchisky, D. M. (1983). Relationships between speech and reading among second language learners. *Language Learning*, *33*(1), 77-102.

Muñoz, C. (Ed.). (2006). *Age and the rate of foreign language learning* (Vol. 19). Multilingual Matters.

Muñoz, C., & Singleton, D. (2007). Foreign accent in advanced learners: Two successful profiles. *Eurosla Yearbook*, *7*(1), 171-190.

Munro, M. J., Flege, J. E., & MacKay, I. R. (1996). The effects of age of second language learning on the production of English vowels. *Applied Psycholinguistics*, *17*(03), 313-334.

Neufeld, G. G. (1979). Towards a theory of language learning ability.*Language Learning*, *29*(2), 227-241.

Neufeld, G. G. (1988). Phonological Asymmetry in Second-Language Learning and Performance. *Language Learning*, *38*(4), 531-559.

Odden, D. A. (2005). *Introducing phonology*. Cambridge university press.

Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research*, *5*(3), 261-283.

Patkowski, M. S. (1980). The sensitive period for the acquisition of syntax in a second language 1. *Language learning*, *30*(2), 449-468.

Penfield, W. (1965). Conditioning the uncommitted cortex for language learning. *Brain*, *88*(4), 787-798.

Polka, L., & Bohn, O. S. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *The Journal of the Acoustical Society of America*, *100*(1), 577-592.

Polka, L., & Bohn, O. S. (2003). Asymmetries in vowel perception. *Speech Communication*, *41*(1), 221-231.

Polka, L., & Bohn, O. S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics*,*39*(4), 467-478.

Polka, L., Molnar, M., Baum, S., Menard, L., & Steinhauer, K. (2009). Asymmetries in the mismatch negativity response to vowels by French, English, and bilingual adults: Evidence for a language-universal bias. *The Journal of the Acoustical Society of America*, *125*(4), 2756-2756.

Polka, L. & J. F. Werker (1994) Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance* 20(2): 421-435.

Porter Jr, R. J., & Castellanos, F. X. (1980). Speech-production measures of speech perception: Rapid shadowing of VCV syllables. *the Journal of the Acoustical Society of America*, *67*(4), 1349-1356.

Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P. K. (2005). Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. *Developmental Science*, *8*(2), 162-172.

Rochet, B. L. (1995). Perception and production of second-language speech sounds by adults. *Speech perception and linguistic experience: Issues in cross-language research*, 379-410.

Rvachew, S. (1994). Speech perception training can facilitate sound production learning. *Journal of Speech, Language, and Hearing Research*,*37*(2), 347-357.

Rvachew, S., Nowak, M., & Cloutier, G. (2004). Effect of phonemic perception training on the speech production and phonological awareness skills of children with expressive phonological delay. *American Journal of Speech-Language Pathology*, *13*(3), 250-263.

Saito, K. (2013). Reexamining effects of form-focused instruction on L2 pronunciation development. *Studies in Second Language Acquisition*, *35*(01), 1-29.

Scovel, T. (1969). Foreign accents, language acquisition, and cerebral dominance1. *Language learning*, *19*(3-4), 245-253.

Scovel, T. (2000). A critical review of the critical period research. *Annual Review of Applied Linguistics*, *20*, 213-223.

Seliger, H. W. (1978). Implications of a multiple critical periods hypothesis for second language learning. *Second language acquisition research: Issues and implications*, 11-19.

Sheldon, A., & Strange, W. (1982). The acquisition of/r/and/l/by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, *3*(03), 243-261.

Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics*, *66*(3), 422-429.

Snow, C. E., & Hoefnagel-Höhle, M. (1977). Age differences in the pronunciation of foreign sounds. *Language and Speech*, *20*(4), 357-365.

Snow, C. E., & Hoefnagel-Höhle, M. (1978). The critical period for language acquisition: Evidence from second language learning. *Child development*, 1114-1128.

Snow, C. E., & Hoefnagel-Höhle, M. (1979). Individual differences in second-language ability: A factor-analytic study. *Language and Speech*, *22*(2), 151-162.

Strange, W. (2006). Second language speech perception: The modification of automatic selective perceptual routines. *The Journal of the Acoustical Society of America*, *120*(5), 3137-3137.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of/rl/by Japanese adults learning English. *Perception & Psychophysics*, *36*(2), 131-145.

Thomas, Erik R. and Kendall, Tyler. 2007. NORM: The vowel normalization and plotting suite. Available at http://lingtools.uoregon.edu/norm/norm1.php (Accessed January 2016)

Thomson, R. I. (2011). Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *Calico Journal*,*28*(3), 744-765.

Trinity College London (2009*) Exam Information: Graded Examinations in Spoken English (GESE)* (6th ed.) Trinity College London

Trost, J. E., & Canter, G. J. (1974). Apraxia of speech in patients with Broca's aphasia: A study of phoneme production accuracy and error patterns. *Brain and language*, *1*(1), 63-79.

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics*, *33*(3), 263-290.

U.S. Department of Education (2010, December 8) Education and the anguage Gap: Secretary Arne Duncan's Remarks at the Foreign Language Summit. Retrieved from http://www.ed.gov/news/speeches/education-and-language-gap-secretary-arne-duncans-remarks-foreign-language-summit

Valdman, A. (1984) Review article: D'accord. La prononciation du francais international et perfectionnement. Caruner, S. and M. P. Hagiwara. New York: Wiley. Methodologie de l'enseignement de la prononciation. Callamand, L. Paris, CLE [Didactique des langues étrangeresl, 1981. Pp. 191. In *studies in Second Language Acquisition* 6(2): 226-32.

Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System*, *32*(4), 539-552.

Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training.*The Journal of the Acoustical Society of America*, *113*(2), 1033-1043.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, *106*(6), 3649-3658.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant behavior and development*, *7*(1), 49-63.

White, L., & Genesee, F. (1996). How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second language research*, *12*(3), 233-265.

Yamada, R. A. (1995). Age and acquisition of second language speech sounds: Perception of American English/r/and/l/by native speakers of Japanese. *Speech perception and linguistic experience: Issues in cross-language research*, 305-320.