Spring 2023

Impact of Accent on Receptive Language Assessments for Bilingual Children

Katherine Stariknok
University of New Hampshire, Durham

Impact of Accent on Receptive Language Assessments for Bilingual Children

Katherine Stariknok

Department of Communication Sciences and Disorders

Honors Thesis

Thesis Advisor: Dr. Jill Thorson

Communication, Acquisition & Translational Studies Lab

Abstract

Language assessments are used to screen and diagnose children with language disorders. Many speech-language pathologists (SLP) practicing in Spanish in the United States do not speak Spanish as their first language, so they are administering these assessments in accented speech. This study aims to find what effect an SLP's accent will have on a bilingual child's language assessment. Initial findings show that participants were largely able to understand words in the L2 accent, with most of the errors due to several repeating consonantal features (i.e., voicing alone; place and manner of articulation). The broader impact of this work is to understand the constellation of phonetic and acoustic differences between L1 and L2 speakers to help SLPs choose which aspects of their pronunciations are most salient and likely to impact test results, which may lead to misdiagnoses.

Keywords: simultaneous bilingual, accent, language disorders, receptive vocabulary assessment

Impact of Accent on Receptive Language Assessments for Bilingual Children

Preface on Terminology

Language changes rapidly, and the Communication Sciences and Disorders field is no exception. Research articles written just a few years ago may contain language we now widely consider outdated, which can make it difficult to discuss prior research without either changing the author's words or using words we now know to have better alternatives. One example of this is in how we discuss speakers of a second language. This can be described as foreign-accented, accented, or second language (L2) speech. Similarly, the way someone produces their first language can be described as native, heritage, or first language (L1). What is now internationally known as developmental language disorder (DLD) was referred to as specific language impairment (SLI), language delay, and receptive and/or expressive language impairment until recently, so most of the existing literature uses the former terminology.

In addition to finding better terminology over time, some words also have field-specific meanings. One relevant example is the distinction between comprehensibility and intelligibility, which are defined differently across different disciplines. In some contexts, the distinction may be relevant, but for this study, they will be used interchangeably to mean speech that a listener can understand the meaning of. Accent and dialect are also words that are often used interchangeably outside the field of linguistics. In the field of linguistics, and the present study, dialects are variations of an L1 (e.g., Spain or Puerto Rican Spanish), and accents are the features of someone's L2 that differ from an L1 speaker.

Introduction

Childhood multilingualism is understudied, but will only become more important as the multilingual population in the United States grows. Due to the infinite combinations of type and

extent of exposure and social circumstances, each multilingual child is unique in their speech and language development. The current project studies how Spanish-English bilingual children understand an accented word when hearing an L2 Spanish speaker administer a language assessment. The focus is on simultaneous bilinguals, or children who learn multiple languages from birth.

Bilingual research in language acquisition has evolved rapidly. Until relatively recently, many people believed that raising a child with more than one language from birth would confuse them and delay their learning. Some bilingual children appeared to not meet their language milestones on time, putting them at greater risk of developing a language disorder like DLD. Recent research in the field has found that not only is that a myth, but bilingual children are ahead of their peers in some specific cognitive areas, including metalinguistic awareness and executive functioning (Hoff, 2014). One cause of this misconception is that most children are assessed in only one of their languages. So, even if a child knew 100 words across their two languages, they would score significantly lower than their monolingual peers on a monolingual assessment. Reliable and accurate assessments developed specifically for multilingual children are critical to assess individual multilingual speakers.

Multilingualism and Language Disorders

There are infinite ways a child can be raised bilingually. Language disorders can also present differently in every child, and not every child will show all the signs and symptoms associated with their diagnosis, making the intersection a difficult topic to study. While bilinguals and monolinguals learn words at the same rate, bilingual children often appear to know fewer words or have a smaller vocabulary due to the overlap between their two lexicons (Hoff, 2014). Similarly, one of the earliest warning signs DLD is a smaller vocabulary size

(Ellis & Thal, 2008). Therefore, bilingual children may be overdiagnosed with language delays because they are not being assessed in both languages, or underdiagnosed if the difference is assumed to be solely due to their bilingualism.

Recent research has attempted to study both these populations together. Nayeb et al. (2020) analyzed the diagnosis of DLD in bilingual children who spoke Swedish and a family language. Nurses assessed the children in either one of their two languages, both, or just Swedish with parental input in their mother tongue. Then, an SLP evaluated them to check the nurse's results. The SLP found that 29% of the participants (32 children) met the diagnostic criteria for DLD, and 25% (28 children) were suspected to have a developmental disability. These proportions are far higher than in the general population, where between 7-14% of children meet the diagnostic criteria for DLD. The disproportionately high DLD diagnosis level might have been due to risk factors like socio-economic status, or there could have potentially also been overdiagnosis by the SLP, even though that was regarded as the gold standard. The main finding was that screening in only one of the languages led to many false positives, which decreased when they were assessed in both. Screening in the culturally dominant language is current practice but yielded low sensitivity in this study. This study used a "blinded clinical examination as a gold standard for DLD diagnosis" (Nayeb et al., 2020, p. 270), meaning the SLP had the ultimate judgment of what was a true/false positive/negative. If the nurse found the child to qualify for the diagnosis and the SLP did not, it was considered a false positive.

Highlighting some of the difficulties in studying this population, Nayeb et al. (2020) discussed the strengths and limitations of their study. Among the strengths, they included the SLP diagnosis and the Reynell test. The authors mention that while translated test items are typically not recommended, this particular test had previously been evaluated for use with

bilingual children, and the SLP and bilingual staff reviewed the translations. Among their listed limitations, Nayeb et al. (2020) include the sample size, use of only six languages, and the low socioeconomic areas served by all the included health centers. Since bilingual children and children with DLD present with many individual differences, a robust sample size is even more important to obtain generalizable results. They do not specify what qualified someone to be considered a bilingual provider in this study, so it is unclear whether they were L1 or L2 speakers of the participants' second language. Since the SLPs' diagnoses were considered to be the most accurate, there is no entirely objective way to assess the number of false positives and negatives. It is possible that biases in standardized assessments, like those used in Nayeb et al. (2020) led to overdiagnosis, which would explain the discrepancy between their findings and the estimated prevalence of DLD.

In contrast, Westman et al. (2008) found that in a sample of Finnish-Swedish simultaneous bilinguals, being bilingual did not increase the risk of a child having DLD. Their study was conducted in Finland, where Swedish is a co-official language, although only 6% of the population speaks it. There are Swedish-language schools at all levels in Finland, but there has recently been a trend of more parents sending their Swedish-speaking children to Finnish-language schools. Many professionals still mistakenly believe it is better and advise parents of bilingual children with DLD to choose one language to teach their child to avoid further delays. The study compared Swedish-Finnish bilinguals with Swedish monolinguals, who were all assessed in Swedish. They were screened at five years old using a variety of receptive and expressive language tests, and the children with the lowest converted composite scores were selected to be in the language impairment (LI) risk group. The control group was selected by random sampling of the remaining children. Those children underwent neuropsychological

assessments one year later, and the scores for selected subtests were used to discriminate between children with language disorders and typically developing. Westman et al. (2008) found that there were proportionally more bilingual children that fell into the LI risk group, which they said may be due to the tests used in language screening. However, they generally concluded that a child's bilingualism did not impact the language development of the LI risk group more than the control group. The language profiles of the bilingual and monolingual children with LI were comparable.

Westman et al. (2008) only assessed in one language but still found that bilingual children were not at a higher risk of language impairment than monolinguals. These findings contradict Nayeb et al. (2020), with the differences in methods potentially explaining these discrepancies. Nayeb et al. (2020) looked at children who were disproportionately from lower socioeconomic statuses. Westman et al. (2008), on the other hand, included bilingual children that had, on average, higher educated parents than the monolingual children. In addition to the impact of SES and parental education levels, the differences in cultural implications of bilingualism, and the languages that were being looked at, may have contributed to the different findings. This highlights the importance of examining for bias in the assessments as well as the sample.

Rezzonico et al. (2015) measured the lexical diversity of 40 preschoolers in a story-telling setting and had similar findings to Westman et al. (2008). They examined the data from 40 children at two test points, six months apart from each other. Narrative retell was selected as it has been found to be a less biased assessment method. By measuring Information Scores and Sentence Length Scores, they were able to identify the ideas the children were able to convey as well as the complexity of the language they were using to do so. The children in the study all had

home languages that were not English, but “heard and spoke predominantly English in the home and only English at school” (Rezzonico et al., 2015, p. 835). They found that children with DLD scored lower than typically developing (TD) children on Information Score and microstructural measures (lexical diversity, sentence length, first mentions, and verb accuracy) with bilingualism showing no significant effect. The one area that differentiated the four groups (bilingual/monolingual typically developing/with DLD) was verb accuracy. This reflects the difficulty with English verbs that bilinguals demonstrated in previous studies (Paradis, 2005; Paradis et al., 2008) as well as morphological impairments that distinguish children with and without DLD (Paradis et al., 2013). Rezzonico et al. (2015) found few differences between bilinguals and monolinguals with DLD. The differences they did note can be attributed to previously identified distinctions of a very specific skill (verb accuracy). In summary, both Rezzonico et al. (2015) and Westman et al. (2008) found that monolingual and bilingual children with DLD consistently score lower on a variety of lexical tasks than typically developing children, but their bilingualism does not seem to additionally impair most of their test results, while Nayeb et al. (2020) found higher proportions of the bilingual population qualified for DLD diagnoses.

Standardized Bilingual Assessments

Standardized tests have long been used to confirm societal biases against certain populations. According to Hoff (2014), the origin of the misbelief that bilingual children are at a disadvantage due to their multiple languages was early research that compared bilingual immigrants to monolingual citizens using a standardized intelligence assessment. Using a standardized test normed on one population and comparing the results of people from very different backgrounds, researchers claimed that “immigrants were genetically inferior” and “their

bilingualism was to blame" without considering other factors (Hoff, 2014, p. 280). Recent research in the field attempts to overcome some of these biases by including the population that will be assessed in the norming process for standardized assessments. Laing and Kamhi (2003) outline types of potential bias present in norm-referenced assessments being used with diverse groups of children. The first, content bias, is caused by the assumption that the children will have the same background knowledge and experiences. For example, a very common test procedure requires the child to point or label, but research has found that Hispanic children may not be familiar with that procedure, and consequently score lower even if they know the content being tested (Anderson, 2002; Peña et al., 1992). Linguistic bias is caused by a mismatch between the examiner's language, the child's language, and/or the language/dialect expected in the child's responses. This type of bias can lead to overdiagnosis when the examiner labels dialectical differences as errors. It can also lead to underdiagnosis if the examiner assumes any error is just a dialectical difference. Linguistic and content biases may be present in any assessment, while a bias specific to norm-referenced tests is disproportionate representation in the normative samples. Historically, norms have reflected the dominant identity, but recently researchers have made efforts to use normative samples that resemble the larger population. Even with more diverse samples, norming biases are still present. One solution proposed for this type of bias by Laing and Kamhi (2003) is the development of population-specific assessments.

The assessment used in the current study is the Receptive One-Word Picture Vocabulary Test 4th Edition Spanish-Bilingual Edition (ROWPVT-4: SBE) (Martin, 2013). The test manual addresses bias in language assessments. Bias is defined as "systematic errors in testing are found among groups of the same ability level but having different characteristics, such as ethnicity, race, or gender. Bias is not evidenced solely by overall differences in group mean scores"

(Martin, 2013, p. 42). Martin (2013) states that collected data from individually administered tests is more expensive than group-administered tests and claims that since the ROWPVT-4 Spanish-Bilingual Edition uses items from the same pool as the English one, for which there was no evidence of bias, it was "not necessary to conduct further item bias analyses" (Martin, 2013, p. 42). The norms for the bilingual version were obtained using a sample of 1,260 children that reflected the overall US Hispanic population. It is impossible for any assessment, especially standardized ones, to be completely free from bias, and the ROWPVT-4: SBE is no exception. For the present study, the assessment is not being used for diagnostic purposes, but solely as a task to examine accent, while also keeping in mind that assessments like this are often used as part of the diagnostic process and the implications have real-world effects.

Conceptual Scoring

To attempt to overcome some of the biases present in the current practice of SLP diagnoses, Gross et al. (2014) studied the scoring of receptive vocabulary measures for English-Spanish bilingual children. They found that conceptual scoring decreased the difference in scores of monolingual and bilingual children. Conceptual scoring is taking the sum of the words a child knows, regardless of what language. If a child can correctly identify *dog*, *perro, agua*, and *pencil,* they would score three points (not four), because *dog* and *perro* are the same conceptually. They would not lose two points for not knowing *water* or *lapiz*, the translations of *agua* and *pencil*, because they do know the concept in at least one of their languages. Most of the research on conceptual scoring has been on expressive, not receptive measures. Additionally, the majority of this research did not control for the accent of the person giving the assessment. The research assistants administering the test in Gross et al. (2014) were all bilingual, but their first languages and accents were not addressed in the research article. In their study, conceptual

scoring helped simultaneous bilinguals more significantly in comparison to sequential bilinguals, with 90% scoring within the average range for all measures. Gross et al. (2014) attempted to show that monolingual norms could be used for scoring bilingual language assessments, even though the American Speech-Language-Hearing Association (ASHA) advises against it, as it has been found to lead to biased outcomes (Pearson, 1998). One hurdle to conceptual scoring is that a translation of a word may be at a different level in the translated language, so the age norms may not align anymore. Taking all this into consideration, Gross et al. (2014) found that on receptive vocabulary tests, conceptual scoring equalized simultaneous bilinguals and monolinguals, but not sequential bilinguals.

Accent/dialect

Although accents can often be identified by listening to them, what constitutes an accent is difficult to define. ASHA defines an accent as "a phonetic trait from a person's original language (L1) that is carried over [to] a second language (L2)" (ASHA, n.d.). While this narrow definition only includes phonetic differences, Munro (1998) uses a broader definition of accent which includes any non-pathological, partially systemic speech characteristic of an L2 learner that is different from an L1 speaker's production. Specifically, this definition includes phonemic production, vocal quality, and intonation as contributors to the perception of an accent. When a speech-language pathologist is conducting an assessment in a language other than their L1, they will inevitably be administering the assessment in an accent.

One way in which L2 speech commonly differs from L1 is the phonemic inventory, or the collection of sounds that exist in a language. Phonemic inventories vary by language, and within a language by region and dialect. The phonetic differences between the L1 and L2 speakers will be analyzed in the present study as a contributing factor to the perception of an

accent. ASHA publishes a collection of resources for clinicians practicing in a language other than English, including phonemic inventories of 20 languages. In their Spanish section, ASHA includes a "standard" Spanish Phonemic Inventory chart, as well as more specific resources for Cuban, Puerto Rican, and Mexican Spanish. Like English, there is no universally accepted "standard" dialect of Spanish. Goldstein, the author who ASHA cited in their Standard Spanish Phonemic Inventory chart, acknowledged this limitation but argued in favor of publishing a standard Latin American phonology, saying a standard would aid in discussing dialectical differences (Goldstein, 2000). Comparing the so-called "standard" Spanish phonemic inventory to that of English, there are several key differences. English has 12 vowels (also called monophthongs) and 5 diphthongs, while Spanish has five vowels and no diphthongs (Sundara, 2022). English contains 12 consonants that Spanish does not, while Spanish has five that English does not. The differences in the consonant inventory of the two languages are shown in the following Table 1.

Table 1. Comparative phonemic inventories of Spanish and English.

	Bilabial	Labiodental	Interdental	Alveolar	Palatal	Velar	Glottal
Stop	**p** **b**			**t** **d**		**k** **g**	ʔ
Nasal	**m**			**n**	ɲ	ŋ	
Tap/flap				**ɾ**			
Trill				r			
Fricative	β	**f** v	θ ð	**s** z	ʃ ʒ	x ɣ	h
Affricate					**ʧ** ʤ		
Glide	w				**j**		
Liquid				**l**	ɻ		

Note: Phonemes present in: **Both Languages (bold)**, Spanish Only (red), English Only (blue).
Chart Prepared by Stariknok, K., based on Shriberg (2019); Gildersleeve-Neumann & Goldstein (2022).

Infants undergo a process within the first year of life, starting before birth, where they become more attuned to the phonology of the language(s) they are exposed to and less attuned to others. Newborn infants can discriminate, or hear the difference between, vowels and consonant contrasts that are present in other languages but not their own. As they develop, they lose the ability to discriminate between phonemes that are not present in their language(s) (Kuhl, 2009). Sequential bilinguals who learn one language after they have narrowed their perceptual discrimination to only their first language have trouble hearing the difference between target morphemes, so their productions are likely to be more accented than those who were exposed to multiple languages within the first year.

One proposed possibility for how L2 learners overcome phonetic mispronunciations in one of their languages is that they encounter many cognates, and by hearing words in both languages they learn to be flexible when they hear a vowel pronounced differently than they expect (Sundara, 2022). Exposure to multiple accents in both/all the languages they are learning may help young simultaneous multilinguals accommodate accents, as infants perceptually narrow to the language/dialect/accent they hear most, which often is in their home.

Voice Onset Time

Voice onset time (VOT), an acoustic feature, is another difference between first and second-language speakers that impacts the perception of accent. VOT is the measure of time between the release of a burst and the onset of voicing. Average VOT measures for the same phonemes vary between languages. For the voiceless stops that are present in both phonemic inventories (/p, t, k/), the average VOT in English is significantly longer than in Spanish (Flege,

1991). English productions of voiceless stops are often aspirated, which means there is a frication sound similar to /h/ between the stop release and voicing.

Research has been conducted with second language learners to see if L1 and L2 speakers use similar VOTs. In other words, the aim was to find whether VOT was a pronunciation aspect that a language learner was able to change when speaking their L2. Flege (1991) found that when speaking English, Spanish-English bilinguals with Spanish as their L1 produced stops with VOTs that were shorter than L1 English speakers. The earlier they started learning English, the closer their VOTs were to an English-speaking monolingual. This evidence suggests that acquiring an L2 earlier on may help reduce the perception of an accent with respect to this acoustic feature. VOT is a phonemic quality that would be included in ASHA's more limited definition of accent as well as Munro's (1998) and could contribute to a listener's perception of whether speech was native/heritage (L1) or accented (L2). However, VOT differences often only differ by a matter of milliseconds, so it is unlikely that a listener would be able to explicitly point to VOT as a contributing factor, whereas phonetic substitutions are often more noticeable.

Lexical Choice

Another dialectical variation is the choice of which word is used when there are several synonyms. Many test items on the ROWPVT-4: SBE require the clinician to choose between two to four synonyms. Ideally, the clinician would be familiar with the most frequently used word in the region each of their clients is from. If the clinician selected a word that is uncommon in the client's dialect, they may not be able to choose the correct option despite knowing the word. For standardized assessments that use pre-selected words like the ROWPVT-4: SBE, there are several solutions. The best practice would be to ask the guardian which word the child would be most familiar with out of the options and use that option during the assessment. When that is not

possible, the clinician would have to rely on their experience with that region's dialects or research in advance of the session. For this study, the guardians were asked to select which lexical item their child would prefer, and those were used in the assessment.

Understanding versus Replicating Accented Speech

Schmale et al. (2011) found that English-speaking toddlers aged 2;6 were able to generalize novel words learned from a Spanish-accented speaker, but toddlers aged 2;0 were less successful. The novel word stimuli used specifically did not include any stop consonants, to avoid influence from VOT differences. The L1 English speaker spoke in the toddler's dialect, while the L2 English speaker had Dominican Spanish as her L1, and the two speakers' voices were judged to be highly similar acoustically. The experiment included one salience, three training, and two test blocks. In the salience block, the toddlers were shown two test objects. In training, the novel "word" was paired with the object. In the test block, the participants were presented with trained and novel pairings. Different groups heard L1 or L2 speech during training, and the other during the test phase. At age 2;0, the toddlers succeeded when trained by the accented speaker, but not the L1 speaker. Schmale et. al (2011) ruled out individual differences, as the L1 English speaker was used in a control experiment and the toddlers were able to learn the words. They concluded that the most likely explanation was that accented speech includes more variability than L1 speech, which may lead to more opportunities for the toddlers to parse which phonemic variability was necessary to encode in their lexicons. The 2;6 participants were able to successfully learn the novel words regardless of which speaker they heard in the training and test phases.

In a comparative study of four- and seven-year-olds' abilities to recognize a word spoken in an unfamiliar regional dialect (London-Glaswegian), Nathan et al. (1998) studied participants'

abilities to repeat and define a single word. The four-year-olds' productions varied, with some more phonetically similar to the Glaswegian pronunciation they heard. However, many were just repeating the sounds with no lexical representation being accessed. The seven-year-olds were able to understand the words more often than the four-year-olds, but their repetitions were phonetically dissimilar to the stimuli. Synthesizing Schmale et. al (2011) and Nathan et al. (1998), typically developing children can better generalize accents as they get older. During the school years, children's abilities to phonetically imitate dissimilar dialects decrease, but they gain the ability to map a word in an unfamiliar accent to an item in their lexicon. For the present study, children will only be hearing words in one of their L1 accents, but the accent will not match the language in which the word is being read.

Comprehensibility of L2 Speech

To investigate the comprehensibility of accented speech, Gibson (2019) assessed 15 bilingual 5-year-olds using recorded L1 and L2 speech. Participants were assessed using the ROWPVT-4: SBE as well as the Test de Vocabulario en Imágenes Peabody: Adaptación Hispanoamericana (TVIP; Dunn, Lugo, Padilla, & Dunn, 1986). The ROWPVT-4: SBE was administered following standard procedure, and the TVIP was administered using the recorded stimuli. The participants were sequential bilinguals, and the L2 Spanish speaker had only been exposed to Spanish in high school classes but listened to the L1 recording and repeated it as closely as possible. She did this three times with each word, and more for some words until it was intelligible, as judged by adult L1 Spanish speakers. Participants were tested two months apart using the TVIP, one time in L1 and one in L2 speech. A practice effect was ruled out, and the participants scored significantly higher when hearing the L1 speaker. Gibson (2019) used their results to argue that further research should be conducted using recorded stimuli, to

determine whether monolingual SLPs may be able to administer standardized assessments to children whose language they do not speak.

There have been multiple studies that attempt to create "error gravity hierarchies," or errors that influence a listener's ability to comprehend L2 speech. This research is complicated by opposing findings between different languages and study designs. Some research has argued that the search for error gravity hierarchies is pointless, and the frequency is far more significant than the type of errors (Albrechtsen et al., 1980). Munro and Derwing (1995) define intelligibility as "the extent to which a speaker's message is actually understood by a listener" but they state that "there is no universally accepted way of assessing it" (p. 289). They examined the impact of an L2 accent by asking participants to transcribe sentences they heard and then comparing the deviations between the participant transcriptions with the intended utterances, as well as a comprehensibility judgment the participants made. L1 English and Mandarin speakers recorded themselves telling a spontaneous story based on a page of cartoons. Researchers then segmented the stories into utterances and selected 36 samples to be used as stimuli. The listeners were L1 English speakers who were studying linguistics or ESL teaching at a Canadian university. They transcribed what they heard using standard orthography, and then rated each sample on a comprehensibility scale of 1 (extremely easy to understand) to 9 (impossible to understand). Listeners then participated in a second session, where they heard the same stimuli, but this time were asked to rate the degree of foreign accent on a 9-point scale (1 = no foreign accent, 9 = very strong foreign accent). The L1 English stimuli all received the lowest mean accent scores, but interestingly one sample scored as less comprehensible than 11 of the Mandarin-influenced samples while still scoring low on the accent scale. The authors cited nonpathological native speech factors (e.g., speaking rate and clarity, word choice) as potential

reasons. Their found that the L2 stimuli were highly intelligible, even with a perception of a heavy accent. They also found large individual differences in accent perception ratings. The implications of this research are that accent and comprehensibility are not directly linked. In other words, an L2 speaker can have a very heavy accent but still have their messages be perfectly understood. This study was limited in that it only studied the perception of one accent variety (Mandarin) in a specific population (university students with a higher-than-average understanding of linguistics), so it may not be generalizable to other varieties/languages and populations.

Weil (2001) studied the ability of a listener to generalize L2 speakers of the same linguistic background. There were three speakers, two L1 Marathi speakers (who also spoke Hindi) and one L1 Russian speaker. Noted differences between the languages' consonant inventories include more pronounced aspiration and trilled alveolars in Marathi and more palatalized consonants in Russian. Six listener groups were split between training and non-training groups. The training groups received a pre-test on Day 1 and three days of training with one of the Marathi speakers, and on Day 5 received a post-test with either the other Marathi or the Russian speaker. Non-training groups only did the Day 5 test. Words were not repeated on different days to avoid memorization of specific tokens. Listeners completed five tasks on test days: writing phonetically balanced (PB) word lists, filling in Haskins Sentences, answering true/false questions about prose passages, transcribing Harvard Sentences, and completing a Modified Rhyme Test (MRT). Weil (2001) predicted that training groups would perform better regardless of speaker (practice effects), and groups hearing the same speaker would perform better (talker effect). If accents were generalizable, the groups hearing the other Marathi speaker would perform better than groups switching to the Russian speaker (accent effect). Weil (2001)

expected an additive effect between the three, with each independently impacting intelligibility. In the PB and MRT tasks, there was no accent effect, only talker. In the Haskins and Harvard tasks, there was no talker effect, only accent. The accent effect appeared to only occur for sentence contexts, not for word-level tasks (PB, MRT), and the reverse for the talker effect. One proposed explanation for the different findings in different tasks is that prosodic elements may be more salient at a sentence-level than in isolated words. Similar to Munro and Derwing (1995), the listeners in Weil (2001) were all college students, implying a similar age and education level, making it difficult to generalize the results to a more diverse population.

Salient Features of L2 Speech

Since research has consistently shown that perceived accent does not directly correlate to comprehensibility, a more pertinent research question is which features lead to less comprehensible L2 productions. Trofimovich and Isaacs (2012) studied which linguistic measures are most tied to comprehensibility ratings of novice raters and experienced ESL teachers, as well as which measures distinguish between low, intermediate, and high L2 comprehensibility. There were 40 French L1/English L2 speakers and 60 L1 English listeners. The speakers were recorded telling a story based on provided images, similar to Weil (2001). The recorded speech samples were transcribed and analyzed for 19 linguistic measures. The listeners provided qualitative reports which were categorically coded. The categories that the measures fell into were phonology (e.g., segmental error ratio, word stress error ratio), fluency (e.g., pause error ratio), linguistic resources (e.g., grammatical accuracy, lexical error ratio), and discourse (story cohesion, breadth, and depth). They found that the categories that distinguished accent from comprehensibility in the teacher ratings were all phonological, while the categories that distinguished comprehensibility were discursive. Factors from each of these categories

impacted both accent perception and comprehension. Trofimovich and Isaacs (2012) concluded that accent and comprehensibility are "overlapping yet distinct constructs" (p. 913). Comprehensibility, which is what the present study is concerned with, was found to be mostly related to grammatical or lexical errors, which would not be present in isolated-word tasks like in the ROWPVT-4.

Santamaría Busto (2015) looked at differences in experts' and non-experts' ratings of L2 speech in five categories, including intelligibility, comprehensibility, and accent. Similar to Trofimovich and Isaacs (2012), they found that intelligibility and comprehensibility were very similar, but accent ratings differed. They found that pronunciation was a far more salient factor in non-experts' comprehensibility ratings, as shown in Figure 1.

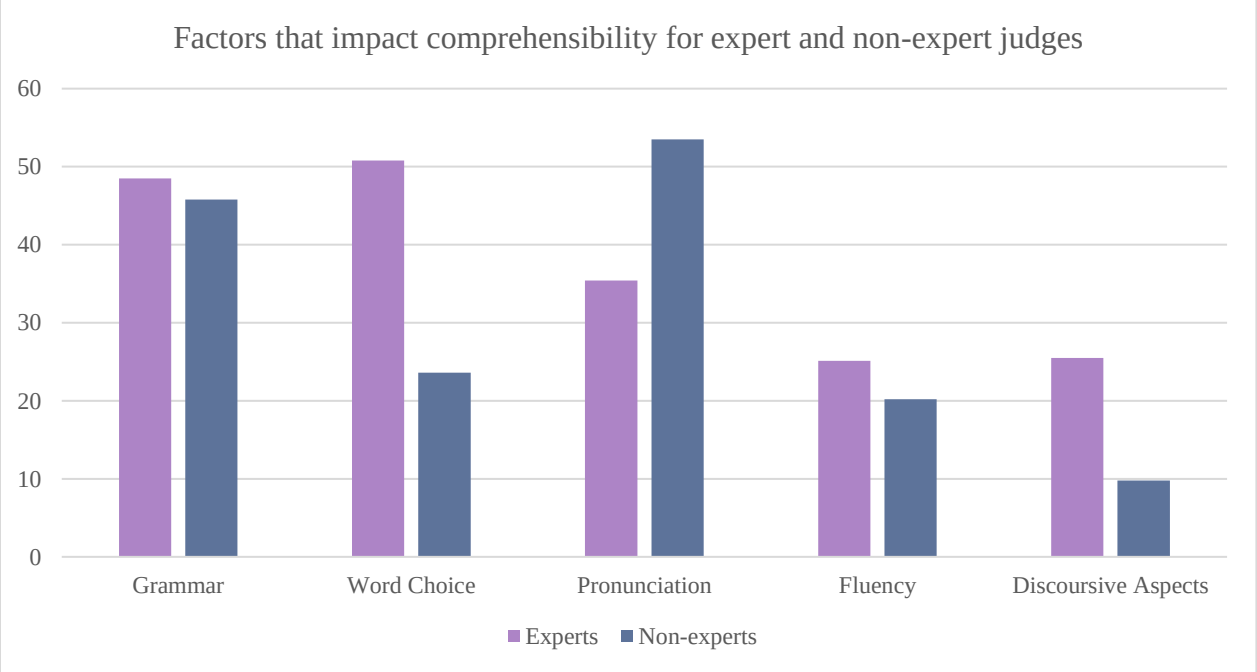

Figure 1. Factors that impact comprehensibility for expert and non-expert judges
Note. Adapted from Santamaría Busto, E. (2015). Percepción y evaluación de la pronunciación del español como L2. In *Revista Española de Lingüística* (1st ed., Vol. 45, p. 195). Sociedad Española de Lingüística.

Looking closer at pronunciation, the expert judges rated consonant and vowel production as being the top two subcategories (out of six) as contributing factors for decreased comprehensibility.

Bilingual Speech-Language Pathologists

As the governing body for the field of Speech-Language Pathology, ASHA published guidelines that bilingual service providers (BSPs) must meet. Among the requirements are that clinicians who call themselves bilingual:

> must be able to speak their primary language and to speak (or sign) at least one other language with native or near-native proficiency in lexicon (vocabulary), semantics (meaning), phonology (pronunciation), morphology/syntax (grammar), and pragmatics (uses) during clinical management. In addition to linguistic proficiency, the audiologist or SLP must have the specific knowledge and skill sets necessary for the services to be delivered. (ASHA, n.d.)

Although multiple graduate SLP programs advertise as being bilingual, ASHA does not accredit programs in this way. Each program's requirements are different and bilingual SLPs can choose to self-designate as such. The variations of programs and people who decide they have met the criteria lead to a lot of individual differences between bilingual service providers.

In ASHA's 2021 demographic profile of multilingual service providers, 17,373 out of 213,115 people were self-indicated multilingual service providers, 15,728 of those being SLPs and the rest audiologists. 66.5% of the multilingual providers were Spanish-language service providers (SSP), with the remainder being comprised of 83 other spoken languages and various manual languages. 46% of the survey respondents indicated their ethnicity as Hispanic/Latino. In their report, ASHA also published data on the number of SSPs in each state compared to the

2019 American Community Survey numbers of families who speak English less than very well and speak Spanish at home. In New Hampshire, there are 11 SSPs, all of whom are SLPs, which comprises 1.0% of service providers in the state. 2.5% of families speak English less than very well, and 2.8% use Spanish at home (American Speech-Language-Hearing Association, 2022). These numbers show a large gap between people who may benefit from receiving services in Spanish and clinicians who are qualified to provide them.

The Present Study

This study aims to find what effect, if any, the L2 accent of a service provider (i.e., SLP) will have on a bilingual child's language assessment. Based on previous research it is theorized that the typically-developing participants will be able to achieve age-appropriate conceptual scores when assessed in both languages due to the matched interlanguage and their age (Schmale et al., 2011; Wang & van Heuven, 2015). However, counterevidence may show that the current participants are too young to overcome the pronunciation differences of the L2 speaker and match the recorded stimuli to an item in their lexicon in an isolated word task (Nathan et al., 1998; Weil, 2001). Secondary aims of the study include 1) understanding which features of L2 productions potentially contribute to the inability to correctly respond on a receptive vocabulary assessment and 2) examining the evaluative process of a child's bilingual speech-language assessment. Specifically, it is predicted that phonetic differences will cause the most difficulty for non-expert listeners, with specific errors likely to repeat due to the phonemic inventories of the two languages (Santamaría Busto, 2015). The broader impact of this work is that it will inform the administration of language assessments by bilingual SLPs as well as their career preparation and training.

Methods

Ethics

The current study was approved by the Institutional Review Board at the University of New Hampshire, UNH IRB-FY2022-393.

Participants

2 children participated in the study (2 females; ages 5;8.26, 4;8.0). All met the inclusion criteria of being between the ages of 3 and 7 years old, simultaneous bilingual speakers of English and Spanish, and having normal or corrected-to-normal vision. Although normal or correct-to-normal hearing was an inclusion criterion, Participant 001 did not pass the hearing screening at 500 Hz in either ear, and Participant 002 needed 30dB for 500 and 1000 Hz in her right ear. See Table 2 for participant details.

Table 2. Participant information.

Participant ID	Age	Hearing L1 or L2 Stimuli	Parent L1
001	5;8.26	L2	One Spanish, one English
002	4;8.0	L2	Both Spanish

Materials

To determine that the participants were truly simultaneous bilinguals with neither English nor Spanish being their more dominant language, the children's guardians filled out two linguistic environment questionnaires. The first was the Alberta Language and Development Questionnaire, which is a parent-report language screening that includes early developmental milestones such as the child's age when they first walked, first said a word, etc. (Paradis et al., 2010). It also asks for the child's current abilities in the first language, behavior patterns and activity preferences, and family history. The second was the Bilingual Input-Output Survey, a parent-report language assessment that can be used with children who are four to six years old

(Peña et al., 2014). This questionnaire asks the parents to indicate how much the child hears (input) and speaks (output) in each language across multiple contexts, including at school/preschool/daycare and home. The parent also estimates how much the child was exposed to each language by year, which will help determine if both languages were acquired as L1 or if one came later than the other.

To assess the child, the Receptive One-Word Picture Vocabulary Test- Fourth Edition Spanish Bilingual Edition was used, which is a norm-referenced, standardized assessment of word recognition that can be used on people aged 2 to over 80 years old (Martin, 2013). The Spanish-Bilingual version of the test was co-normed with the English version and provides percentile scores, standard scores, and age equivalents. The ROWPVT-4: SBE contains 180 items. The assessment is stopped after the child cannot correctly point to four out of six consecutive words.

Stimuli

Two female speakers recorded the stimuli for the study in careful, clear speech. The verbal stimuli (180 test words, 4 sample words) were recorded in a soundproof booth. Each speaker recorded both the Spanish and English test items, so the procedure could be followed as closely as possible, and the participants would only hear one speaker. The speaker who recorded the accented stimuli was a 21-year-old female with Eastern New England U.S. English as L1 and Spanish as L2. She started learning Spanish at 12 years old. The speaker who recorded L1 stimuli was a 40-year-old female with Liman Peruvian Spanish as L1 and English as L2. She started learning English at 5 years old. Speakers were instructed to speak in a clear, non-creaky voice and use falling intonation. The L2 Spanish speaker included phonetic and lexical stress variations that differed from the L1 speaker. Both were told to emphasize phonetic differences

that contribute to the perception of accent in their L2 recordings, especially in their vowel productions. They also included prosodic differences by placing the lexical stress on syllables that L1 speakers would not (e.g., TAzon/taZON). Recordings were made using a Shure SM11 microphone and a Zoom H4n Pro recorder. They were then segmented and scaled to equal mean intensities using Praat (Boersma & Weenink, 2023). A second listener reviewed the files for naturalness before they were used in the study.

Procedure

Before the study began, informed consent was obtained in English or Spanish from the caregiver (depending on language preference). After consent was obtained from the legal guardian, assent was obtained by the child (in either Spanish or English). The child had their hearing screened in both ears at 500, 1000, 2000, and 4000 Hz. The caregiver then completed the Alberta Language and Development Questionnaire and the Bilingual Input-Output Survey. After determining that the participants had normal or correct to normal hearing and were simultaneously acquiring English and Spanish, the clinician administered the ROWPVT-4: SBE, adhering to the protocol with two exceptions. Since the assessment is designed to have the stimuli read to the client, the instructions were altered slightly to reflect the use of audio recordings. Instead of saying “Quiero que señales a la imagen o que me digas el numero de la imagen que muestra la palabra que *yo digo* [emphasis added]” [I want you to point to the picture or tell me the number of the picture that shows the word *I say*], the clinician said “Quiero que señales a la imagen o que me digas el numero de la imagen que muestra la palabra que *tu oyes de la computadora*” [I want you to point to the picture or tell me the number of the picture that shows the word *you hear from the computer*] (Martin, 2013, p. 2). The assessment was also started at test item 001, not the base for the participants’ ages, to allow more data to be collected.

The clinician administering the assessment was a second-year graduate student at the University of New Hampshire, studying to become a bilingual SLP. She was supervised by a certified bilingual SLP. The stimuli were played on the clinician's laptop with a uniform volume setting used for each participant and word.

The participants listened to recordings in Spanish, spoken by a female L1 English speaker. After a word was played, the child was asked to select the image that matched the vocabulary word. If the child did not know the word from the Spanish stimuli, the clinician played the same speaker saying the word in English, as the protocol directs (Martin, 2013).

After completion of the study, the caregiver was provided a monetary incentive in the form of a $15 gift card and the child received a book/small gift of their choosing.

Analysis

If the participants were not able to respond to the prompt when the stimuli were presented with L1 English accented speech in Spanish, then the participants would hear the L1 English version. If these were answered correctly, those items were added to a list of "incomprehensible" words. This showed that they had a lexical representation of the word but were unable to answer correctly when they heard the accented speech. It is possible that a participant would know only the English word, and that it was not the pronunciation of the Spanish that prevented them from correctly answering. For example, if the word is more academic in nature and would be taught in school, and the participant attends a school that exclusively is English-speaking (which both do), they may not know words like "octágono/octagon" or "cuartos/quarters" in Spanish, but they would in English. For this study, the words that they only answered correctly in English were not the "academic" items or words that they would likely only have heard in environments where English was spoken.

Phonetics/Prosody

The words from the incomprehensible list were phonetically transcribed in the International Phonetic Alphabet (IPA) for both speakers. Phonetic and lexical stress differences between the two speakers were recorded as contributors of accent that made the speech incomprehensible. The varying consonants were categorized by which distinctive features (place, manner, and voicing) differed between the L1 and L2 productions.

Acoustics

Two items with each stop consonant in word-initial position were selected, one from the list of lexical items where the participant did not correctly answer when hearing the Spanish stimuli but then got it correct with the English, and one that was not on that list, meaning they were able to understand the word in Spanish. The VOT of those 12 words from each speaker (24 total sound files) was measured in Praat (Boersma & Weenink, 2023). Paired sample t-tests were performed on voiced/voiceless stops from the L1/L2 speakers to determine if there were significant differences between the words on the correct and incorrect lists and between the L1 and L2 speakers.

Results

Phonetic/Lexical Stress Analysis

When scoring the ROWPVT according to the standard procedure, which considers responses in both languages, the participants' conceptual scores were both above average, as shown in Table 3. However, only counting the Spanish scores led to decreases in both participants' scores, illustrated in Table 4.

Table 3. Conceptual Scores for the ROWPVT (Spanish/English responses included)

Participant	Raw Score	Standard Score	Percentile Rank
001	74	112	79%

002	74	116	86%

Table 4. Spanish-only Scores for the ROWPVT

Participant	Raw Score	Standard Score	Percentile Rank
001	63	105	63%
002	42	97	42%

Table 5 shows the words each or both participants were not able to understand in the accented speech, as well as the IPA transcriptions for each speaker. Differences in primary lexical stress are indicated, as well as the number of differing vowels and consonants. The consonants were then analyzed by differing feature(s), either place (P), manner (M), voicing (V), or a combination. Tables 5-7 show the differences between the L1 and L2 speaker's pronunciations for words that the participants were unable to understand in the L2 Spanish but responded correctly with the English stimuli.

Table 5. Incomprehensible list (both participants)

Item (English)	Spanish	Accented/L2 Transcription	L1 Transcription	Primary Stress Differs?	Differing Vowels	Differing Consonants	Features Changed		
004 chair	silla	/ˈsɪlə/	/ˈsije/		2/2	1/2	2: /l/, /j/ P, M		
019 cookie	galleta	/gəˈlɛtə/	/gaˈjeta/		3/3	1/3	2: /l/, /j/ P, M		
026 happy	feliz	/ˈfɛlɪz/	/feˈlis/	Y	2/2	1/3	1: /z/, /s/ V		
028 mailman	cartero	/ˈkɑɻtəɻo/	/karˈteɾo/	Y	2/3	2/4	2: /ɻ/, /r/ P, M	2: /ɻ/, /ɾ/ P, M	
037 broken	roto	/ˈɻɑto/	/ˈroto/		1/2	1/2	2: /ɻ/, /r/ P, M		
049 bowl	tazón	/ˈtæzɪn/	/taˈson/	Y	2/2	1/3	1: /z/, /s/ V		
058 examination	examen	/ɪgˈzæmɪn/	/ekˈsamen/		3/3	2/4	1: /g/, /k/ V	1: /z/, /s/ V	
071 protect	proteger	/ˈpɻɑdədʒɚ/	/pɾoteˈxeɾ/	Y	3/3	3/4 (/5)	2: /ɻ/, /ɾ/ P, M	1: /d/, /t/ V	2: /dʒ/, /x/ P, M

Table 6. Incomprehensible list (participant 1)

Item (English)	Spanish	Accented/L2 Transcription	L1 Transcription	Primary Stress Differs?	Differing Vowels	Differing Consonants	Features Changed	
047 barking	ladrido	/ˈlædɻɪdo/	/ladˈɾido/	Y	2/3	1/4	2: /ɻ/, /ɾ/ P, M	
060 sailboat	velero	/vəˈliɻo/	/beˈleɾo/		2/3	2/3	2: /v/, /b/ P, M	2: /ɻ/, /ɾ/ P, M
068 vine	vid	/ˈvɪd/	/ˈvid/		1/1	0		

Table 7. Incomprehensible list (participant 2)

Item (English)	Spanish	Accented/L2 Transcription	L1 Transcription	Primary Stress Differs?	Differing Vowels	Differing Consonants	Features Changed
003 spoon	cuchara	/kəˈʧɑɻə/	/kuˈʧarə/		2/3	1/3	2: /ɻ/, /ɾ/ P, M
007 hand	mano	/ˈmæno/	/ˈmano/		1/1	0	
016 clock	reloj	/ɻəˈlɑ/	/reˈlox/		2/2	1(2)/2	2: /ɻ/, /r/ P, M
024 people	gente	/ˈhɛnti/	/ˈhɛnte/		1/2	0	
025 cutting	corta	/ˈkoʊɻtə/	/ˈkortə/		1/2	1/3	2: /ɻ/, /r/ P, M
030 purse	bolsa	/ˈbɑlsə/	/ˈbolsa/		2/2	0	
031 jump	saltar	/sɑlˈtɑɻ/	/salˈtar/		2/2	1/4	2: /ɻ/, /r/ P, M
034 jungle	selva	/ˈsɛlvə/	/ˈsɛlβə/		1/2	1/3	1: /v/, /β/ P
036 spilling	derramar	/diˈɻæmə˞	/derəˈmar/	Y	3/3	1/3	2: /ɻ/, /r/ P, M
039 noise	ruido	/ɻuˈido/	/ruˈido/		0	1/2	2: /ɻ/, /r/ P, M
040 coat	abrigo	/ˈæbɻigo/	/aˈbɾigo/	Y	2/3	1/3	2: /ɻ/, /ɾ/ P, M
041 snake	serpiente	/səɻpiˈɛnte/	/serpiˈɛnte/		1/3	1/4	2: /ɻ/, /ɾ/ P, M
046 juggler	malabarista	/ˈmæləbəɻɪstə/	/maləbaɾˈistə/	Y	2/5	1/6	2: /ɻ/, /ɾ/ P, M
048 puddle	charco	/ˈʧɑɻko/	/ˈʧarko/		1/2	1/3	2: /ɻ/, /r/ P, M
052 melted	derretido	/diɻəˈtido/	/dereˈtido/		2/4	1/4	2: /ɻ/, /r/ P, M
057 hatching	incubar	/ɪnˈkʌbə˞	/inkuˈbaɾ/	Y	2/3	0	
062 distress	angustia	/ænˈgʌstiə/	/anˈgustia/		3/4	0	

Out of the 27 consonant substitutions, there were observable patterns, with certain phoneme substitutions accounting for most of the errors. The “r” sound in English differed significantly from the closest Spanish equivalents. The L1 English speaker produced her “r”s as a retroflex approximate, represented by the IPA /ɻ/. English speakers typically produce the “r” sound as either retroflex or alveolar approximates **(/ɹ/)**, but one speaker usually uses the same articulation consistently. In Spanish, there are also two typical “r” productions, but unlike English, speakers use both depending on the word position and surrounding phonemes (/ɾ/ is the alveolar tap, and /r/ is the alveolar trill). Eight out of the 27 consonant errors were /ɻ/ for /ɾ/ substitutions, and 9 were /ɻ/ for /r/, meaning 17 out of 27 consonant errors were due to the L1 English speaker pronouncing either an alveolar tap or trill as the retroflex approximate.

Six of the remaining 10 errors were voicing alone, with the place and manner being the same as the L1 Spanish production. Voiced /g/ and /z/ (phonemes that do not exist in L1 Spanish) were substituted for unvoiced /k/ and /s/. Only four consonant substitutions were between two phonemes that both languages shared (/l/, /j/; /t/, /d/; /k/, /g/). Figure 2 shows the differing features of the L2 productions that the participants were not able to comprehend.

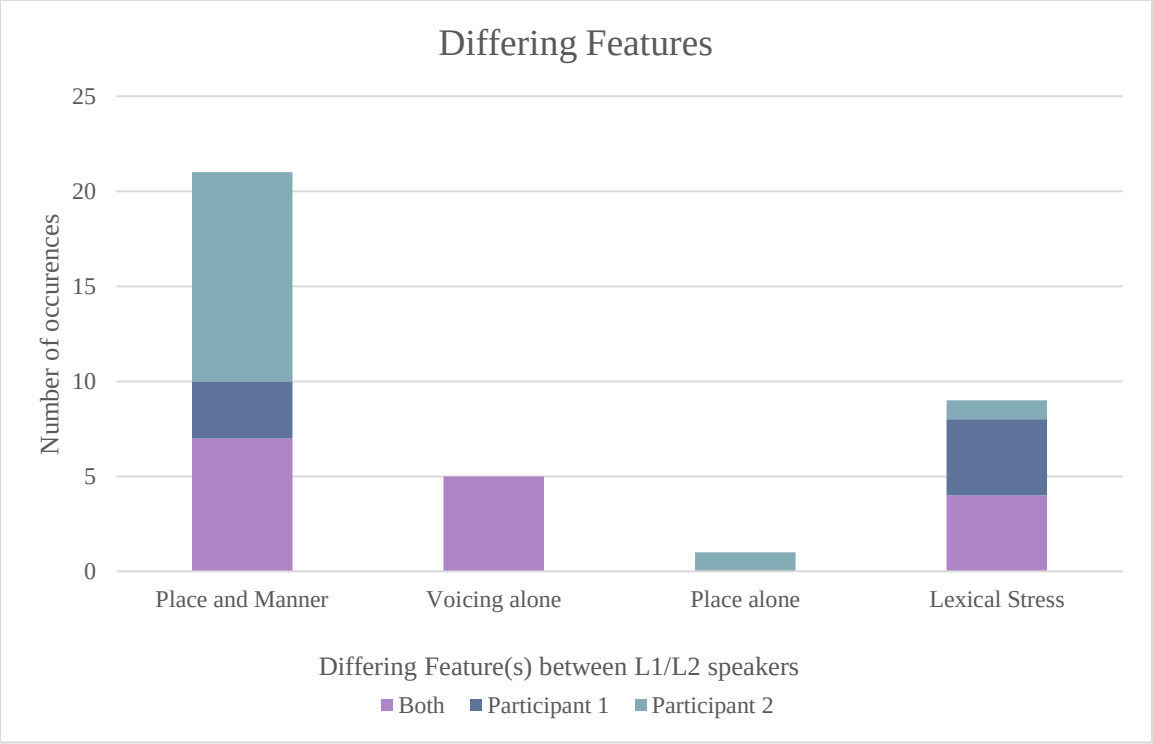

Figure 2. Accent distinctive features for consonants on incomprehensible list

VOT Analysis

Some vocal qualities cannot be captured using IPA transcription, including VOT. VOT was analyzed separately using Praat. Twelve items from each speaker were acoustically analyzed. T-tests between the correct/incorrect test items were found to be insignificant within each speaker, showing that VOT alone did not lead to a word being incomprehensible to the participants. Insignificant differences were found between the comprehensible and incomprehensible words for L1 voiced consonants [comprehensible] (M = 16.16, SD = [6.88]) and [incomprehensible] (M = 27.84, SD = [8.80]); $t(3)$ =4.30, p = 0.13, L1 unvoiced [comprehensible] (M = 25.89, SD = [9.66]) and [incomprehensible] (M = 32.41, SD = [30.48]); $t(3)$ =4.30, p = 0.65, L2 voiced [comprehensible] (M = 26.43, SD = [13.39]) and [incomprehensible] (M = 23.64, SD = [8.91]); $t(3)$ =4.30, p = 0.40, and L2 unvoiced [comprehensible] (M = 86.30, SD = [8.38]) and

[incomprehensible] (*M* = 85.17, *SD* = [9.45]); *t*(3) =4.30, *p* = 0.95. The comprehensible and incomprehensible values were collapsed and then analyzed for differences between speakers. T-tests comparing speakers were significant between L1 and L2 voiceless stops [L1] (*M* = 29.15, *SD* = [8.38]) and [L2] (*M* = 85.73, *SD* = [9.45]); *t*(5) =2.57, *p* = 0.017, and insignificant for voiced stops [L1] (*M* = 22.00, *SD* = [3.89]) and [L2] (*M* = 25.04, *SD* = [4.20]); *t*(5) =2.57, *p* = 0.50. Figures 3 and 4 show the differences of VOT by speaker.

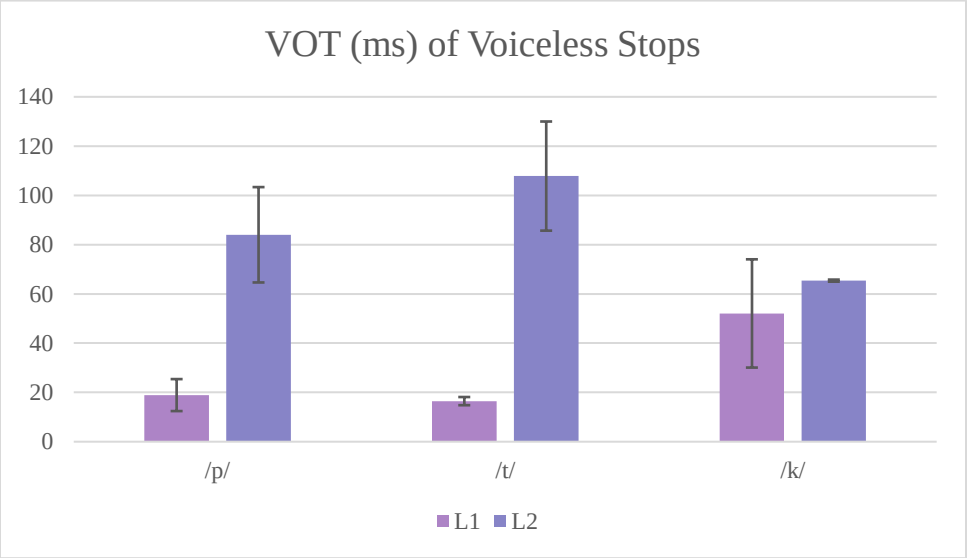

Figure 3. VOT of voiceless stops /p/, /t/, /k/ by speaker

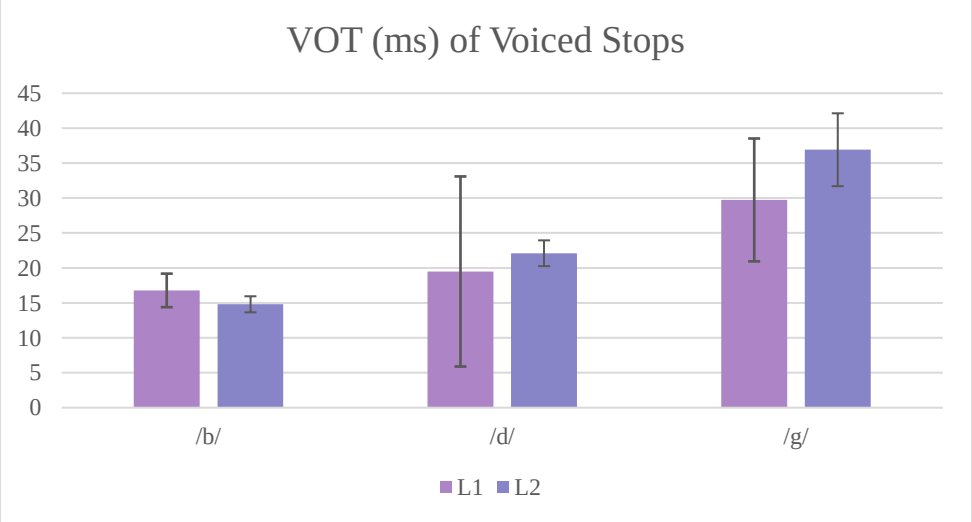

Figure 4. VOT of voiced stops /b/, /d/, /g/ by speaker

Discussion:

The first hypothesis, that the participants would be able to attain age-appropriate conceptual scores due to the matched interlanguage and their age, was confirmed by the data collected in the present study. While the participants were still able to score within the normal range, they would likely have scored higher if they had heard an L1 speaker. If they had only been evaluated in one language, as many bilingual children unfortunately still are, their results would have been drastically different, with one of the participant's Spanish-only percentile being less than half of their conceptually scored percentile.

Language exposure also seemed to affect how capable the participants were of accepting L2 variations. Participant 1 is an only child with one Spanish and one English-speaking parent. At home, she hears both English and Spanish, as well as accented Spanish and English when the parents speak their L2s to each other. Participant 2 has an older sibling who speaks both languages at home, but both parents are heritage Spanish speakers, and her home language exposure is almost entirely Spanish, spoken by L1 speakers. Participant 1 was possibly able to "ignore" or inhibit the accent more easily because she hears accented Spanish more often than Participant 2. Participant 2 also used more external processing when she heard the stimuli. For multiple items, she recast the word to hear an L1 pronunciation, then answered correctly in Spanish. Several times, she tried different pronunciations aloud, running through the options in her lexicon with the same word-initial sound or syllable pattern. She also giggled when the accented word was particularly far from the target (e.g., substituting /l/ for /j/ in "silla"), showing that she knew the speaker was making errors.

Due to findings on the distinction between perceived accent and actual comprehensibility, it may be more productive to address the salient features that most frequently lead to errors,

instead of a general goal of “accent reduction”. This study attempts to show which phonetic features, specifically in consonants, were more likely to lead to incomprehensibility. Spanish-language service providers may use these findings to choose which aspects of their L2 speech is more likely to lead to flawed assessment results. For example, they may pay special attention to learning how and when to roll their “r”s.

The current study may serve as a pilot study for future research with larger samples. It may also encourage future research to compare results obtained using recorded stimuli for standardized language assessments. If future participants score lower hearing L2 speech compared to L1 but score similarly with recorded and in-person stimuli, one tool used to bridge the gap between qualified bilingual service providers and children needing speech services could be the use of recorded, L1 stimuli. The participant pool could expand to include non-typically developing children, as that is the population that it is critical for language assessments to accurately test.

Conclusion

Spanish-English simultaneous bilingual children were able to score within the normal range on a norm-referenced standardized receptive language assessment (the ROWVT-4) when hearing recorded stimuli from an L2 speaker emphasizing factors that contribute to the perception of an accent. Several phonemic substitutions comprised the majority of lexical items the participants were not able to correctly answer in Spanish, providing possible guidance for SLPs practicing in Spanish as their L2.

References

Albrechtsen, D., Henriksen, B., & Færch, C. (1980). Native speaker reactions to learners' spoken interlanguage. *Language Learning, 30*, 365-396. https://doi.org/10.1111/j.1467-1770.1980.tb00324.x.

American Speech-Language-Hearing Association. (2022). 2021 demographic profile of ASHA members providing multilingual services. Retrieved Mar 21 2023, from https://www.asha.org/siteassets/surveys/demographic-profile-bilingual-spanish-service-members.pdf.

American Speech-Language-Hearing Association (n.d.). Bilingual service delivery (Practice Portal). Retrieved Nov 10, 2022, from www.asha.org/Practice-Portal/Professional-Issues/Bilingual-Service-Delivery/.

American Speech-Language-Hearing Association (n.d.). Phonemic inventories and cultural and linguistic information across languages. Retrieved Feb 7, 2023, from https://www.asha.org/practice/multicultural/phono/.

American Speech-Language-Hearing Association. (1998). Students and professionals who speak English with accents and nonstandard dialects: Issues and recommendations [Position statement]. Retrieved Dec 14, 2022, from https://www.asha.org/policy/tr1998-00154/.

Anderson, R. (2002). Practical assessment strategies with Hispanic students. In A. E. Brice (Ed.). *The Hispanic Child* (pp. 143-184). Boston, MA: Allyn & Bacon.

Barstein, J., Martin, G. E., Lee, M., & Losh, M. (2018). A duck wearing boots?! Pragmatic language strategies for repairing communication breakdowns across genetically based neurodevelopmental disabilities. *Journal of Speech, Language, and Hearing Research, 61*(6), 1440-1454. https://doi.org/10.1044/2018_JSLHR-L-17-0064.

Boersma, P. & Weenink, D. (2023). Praat: doing phonetics by computer [Computer program]. Version 6.3.09. retrieved 2 March 2023 from http://www.praat.org/.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science, 208*(4448), 1174–1176. https://doi.org/10.1126/science.7375928.

Dunn, L. M., Padilla, E. R., Lugo, D. E., & Dunn, L. M. (1986). *TVIP: Test de vocabulario en imágenes Peabody: adaptación Hispanoamericana = Peabody picture vocabulary test [revised]: Hispanic American adaptation*. Circle Pines, Minn: American Guidance Service.

Ellis, E. M., & Thal, D. J. (2008). Early language delay and risk for language impairment. *Perspectives on Language Learning and Education*, *15*(3), 93–100. https://doi.org/10.1044/lle15.3.93.

Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America, 89*, 395-411. https://doi.org/10.1121/1.400473.

Gibson, T. A. (2019). The influence of native- versus foreign-accented speech on Spanish–English bilingual children's Spanish receptive vocabulary performance: A pilot study. *Language, Speech, and Hearing Services in Schools*, *50*(4), 710–716. https://doi.org/10.1044/2019_LSHSS-18-0136.

Gildersleeve-Neumann C., & Goldstein, B. (2022). Speech sound development and disorders. In Goldstein, B. (Ed.). *Bilingual Language Development & Disorders in Spanish–English Speakers* (3rd ed., pp. 307-347). Paul H. Brookes Publishing.

Goldstein, B. (2001). Transcription of Spanish and Spanish-Influenced English. *Communication Disorders Quarterly*, *23*(1), 54–60. https://doi.org/10.1177/152574010102300108.

Gross, M., Buac, M., & Kaushanskaya, M. (2014). Conceptual scoring of receptive and expressive vocabulary measures in simultaneous and sequential bilingual children. *American Journal of Speech-Language Pathology*, *23*(4), 574–586. https://doi.org/10.1044/2014_AJSLP-13-0026.

Hoff, E. (2014). *Language development*. Belmont, CA: Wadsworth Cengage Learning.

Korkman, M., Kemp, S., & Kirk, U. (1998). *NEPSY: A developmental neuropsychological assessment; manual*. PsychCorp.

Kohnert, K. (2013). *Language disorders in bilingual children and adults*, 2nd ed. (2013). San Diego, CA: Plural Publishing Inc.

Kuhl, P. K. (2009). Early language acquisition: Phonetic and word learning, neural substrates, and a theoretical model. In B. Moore, L. Tyler & W. Marslen-Wilson (Eds.), *The Perception of Speech: From Sound to Meaning (*pp 103-131). Oxford, UK: Oxford University Press.

Laing, S. P., & Kamhi, A. (2003). Alternative assessment of language and literacy in culturally and linguistically diverse populations. *Language, speech, and hearing services in schools*, *34*(1), 44–55. https://doi.org/10.1044/0161-1461(2003/005).

Martin, A. (2013). *Receptive one-word picture vocabulary test-4 ROWPVT-4: SBE: Spanish-bilingual edition* (Spanish-bilingual edition). Los Angeles, CA: Academic Therapy Publications.

Morton, E. S., Brundage, S. B., & Hancock, A. B. (2010). Validity of the proficiency in oral English communication screening. *Contemporary Issues in Communication Science and Disorders*, *37*(Fall), 153–166. https://doi.org/10.1044/cicsd_36_F_153.

Munro, M.J. and Derwing, T.M. (1995), Foreign accent, comprehensibility, and intelligibility in

the speech of second language learners. *Language Learning*, 45: 73-97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x.

Munro, M. J. (1998). The effects of noise on the intelligibility of foreign-accented speech. *Studies in Second Language Acquisition*, *20*(2), 139–154. https://doi.org/10.1017/S0272263198002022.

Nathan, L., Wells, B., & Donlan, C. (1998). Children's comprehension of unfamiliar regional accents: A preliminary investigation. *Journal of Child Language*, 25, 343–365. https://doi.org/10.1017/S0305000998003444.

Nayeb, L., Lagerberg, D., Sarkadi, A., Salameh, E., & Eriksson, M. (2021). Identifying language disorder in bilingual children aged 2.5 years requires screening in both languages. *Acta Paediatrica*, *110*(1), 265–272. https://doi.org/10.1111/apa.15343.

Paradis, J. (2005). Grammatical morphology in children learning English as a second language: implications of similarities with specific language impairment. *Language, Speech, and Hearing Services in Schools*, 36, 172–187. https://doi.org/10.1044/0161-1461(2005/019).

Paradis, J., Rice, M.L., Crago, M., and Marquis, J. (2008). The acquisition of tense in English: distinguishing child second language from first language and specific language impairment. *Applied Psycholinguistics*, 29, 689–722. https://doi.org/10.1017/S0142716408080296.

Paradis, J, Schneider, P., and Duncan, T. S. (2008). Discriminating children with language impairment among English-language learners from diverse first-language backgrounds. *Journal of Speech, Language, and Hearing Research*, 56, 971–981. https://doi.org/10.1044/1092-4388(2012/12-0050).

Paradis, J., Emmerzael, K., & Sorenson Duncan, T. (2010). Assessment of English language learners: Using parent report on first language development. *Journal of Communication Disorders*, 43, 474-497.

Park, J., Miller, C. A., Sanjeevan, T., van Hell, J. G., Weiss, D. J., & Mainela-Arnold, E. (2019). Bilingualism and attention in typically developing children and children with developmental language disorder. *Journal of Speech, Language, and Hearing Research*, *62*(11), 4105–4118. https://doi.org/10.1044/2019_JSLHR-L-18-0341.

Pearson, B. Z. (1998). Assessing lexical development in bilingual babies and toddlers. *International Journal of Bilingualism*, *2*(3), 347–372. https://doi.org/10.1177/136700699800200305.

Peña, E., Quinn, R., & Iglesias, A. (1992). The application of dynamic methods to language assessment: A nonbiased procedure. *The Journal of Special Education*, *26*(3), 269–280. https://doi.org/10.1177/002246699202600304.

Peña, E. D., Gutiérrez-Clellen, V. F., Iglesias, A., Goldstein, B. A., & Bedore, L. M. (2014). Bilingual English Spanish Assessment (BESA). San Rafael, CA: AR Clinical Publications.

Rezzonico, S., Chen, X., Cleave, P. L., Greenberg, J., Hipfner-Boucher, K., Johnson, C. J., Milburn, T., Pelletier, J., Weitzman, E., & Girolametto, L. (2015). Oral narratives in monolingual and bilingual preschoolers with SLI: Narratives in bilingual children with SLI. *International Journal of Language & Communication Disorders*, *50*(6), 830–841. https://doi.org/10.1111/1460-6984.12179.

Santamaría Busto, E. (2015). Percepción y evaluación de la pronunciación del español como L2. In *Revista Española de Lingüística* (1st ed., Vol. 45, pp. 175–207). Sociedad Española de Lingüística.

Schmale, R., Hollich, G., & Seidl, A. (2011). Contending with foreign accent in early word learning. *Journal of Child Language*, *38*(5), 1096–1108. https://doi.org/10.1017/S0305000910000619.

Shriberg, L. D. (2019). *Clinical phonetics*. Boston, MA: Pearson.

Sundara, M. (2022). Development of speech perception and production in bilingual preverbal infants. In Goldstein, B. (Ed.). *Bilingual Language Development & Disorders in Spanish–English Speakers* (3rd ed., pp. 67-93). Paul H. Brookes Publishing.

Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, *15*(4), 905–916. https://doi.org/10.1017/S1366728912000168.

Wang, H., & van Heuven, V. J. (2015). The interlanguage speech intelligibility benefit as bias toward native-language phonology. *I-Perception*, *6*(6), 204166951561366. https://doi.org/10.1177/2041669515613661.

Weil, S. (2001). Foreign accented speech: Encoding and generalization. *The Journal of the Acoustical Society of America*, *109*(5), 2473–2473. https://doi.org/10.1121/1.4744779.

Westman, M., Korkman, M., Mickos, A., & Byring, R. (2008). Language profiles of monolingual and bilingual Finnish preschool children at risk for language impairment. *International Journal of Language & Communication Disorders*, *43*(6), 699–711. https://doi.org/10.1080/13682820701839200.