2016-03-01

Psychometrically Equivalent Trisyllabic Words for Testing Spanish Pediatric Speech Recognition Thresholds

Jessica Lee Graham
Brigham Young University - Provo

Psychometrically Equivalent Trisyllabic Words for Testing

Spanish Pediatric Speech Recognition Thresholds

Jessica Lee Graham

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Richard W. Harris, Chair
David L. McPherson
Christopher D. Dromey

Department of Communication Disorders

Brigham Young University

March 2016

ABSTRACT

Psychometrically Equivalent Trisyllabic Words for Testing
Spanish Pediatric Speech Recognition Thresholds

Jessica Lee Graham
Department of Communication Disorders, BYU
Master of Science

The purpose of this study was to use previously recorded Spanish trisyllabic words tested on adults in the measurement of the speech recognition threshold of Spanish-speaking children in order to (a) determine the words' appropriateness when testing children and (b) compare psychometric functions between adults and children. A selection of 28 frequently used trisyllabic words was chosen from previously recorded samples of male and female adult native speakers of Spanish. These words were then presented to 20 native Spanish-speaking children with normal hearing between the ages of 4 and 8 years. The words were presented starting at -5 dB HL and ascended in 5 dB increments until the presentation level reached 15 dB HL. Using logistic regression, psychometric functions were calculated for each word. Resulting pediatric thresholds were found to be 8.7 dB higher for male talkers and 11.0 dB higher for female talkers than previously reported adult thresholds. These results indicate a clinically significant threshold difference between pediatric and adult populations. Future research should be conducted to measure the speech recognition threshold (SRT) in children of varying ages to determine the age at which the SRT approximates with adult performance.

Keywords: speech audiometry, speech recognition threshold, Spanish, pediatric, SRT

ACKNOWLEDGMENTS

I wish to thank my thesis advisor, Dr. Harris, for his guidance and support throughout this project. I extend my gratitude to all of the parents and children who gave their time and efforts to this research. Many thanks are due to my parents for always placing an emphasis on the pursuit of education and motivating me to earn a master's degree. I am grateful to all of the members of my cohort for making graduate school an enjoyable experience with unforgettable friends and memories. Most of all, I wish to thank my husband for his continual support and motivation throughout these two busy years.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

LIST OF APPENDICES

DESCRIPTION OF CONTENT

The body of this thesis is written as a manuscript suitable for submission to a peer-reviewed journal in speech-language pathology and audiology. This thesis will be suitable for submission to a peer-reviewed journal when combined with additional research concluding the age at which a child's speech thresholds approximate those of an adult. An annotated bibliography is presented in Appendix A.

Introduction

Comprehensive Audiological Testing

Audiological testing was developed to determine the presence, type, degree, and configuration of a patient's hearing loss. A typical evaluation includes various components: otoscopy, tympanometry, acoustic reflex testing, pure-tone air and bone conduction audiometry, and speech audiometry, including speech recognition thresholds (SRT) and word recognition scores (WRS). Otoscopy presents a view of the ear canal and tympanic membrane; tympanometry tests the mobility of the tympanic membrane to detect fluid and overall health within the middle ear; acoustic reflex testing uses a high-intensity sound to stimulate an involuntary muscle contraction within the middle ear; pure-tone audiometry identifies hearing threshold levels; speech audiometry provides information on word recognition abilities. Audiologists perform these tests to determine the etiology of an individual's hearing difficulties (Egan, 1979).

Pure-tone audiometry uses simple tones to determine the lowest detectable intensities a listener can perceive at given frequencies, thus identifying an individual's hearing thresholds. Typically, both octave and mid-octave frequencies between 125 to 8000 Hz are presented to the listener; however, often only octave frequencies are presented when listener fatigue is a problem. Testing can be performed through air or bone conduction. Pure-tone audiometry does not provide a complete understanding of a person's hearing abilities. Everyday stimuli, including speech, consist of more complex sounds than the simple tones used in pure-tone testing. To further understand an individual's hearing abilities, speech must be tested (Egan, 1979; Martin & Clark, 2009).

Speech Audiometry

Since its introduction, speech audiometry has been used to supplement the results of pure-tone audiometry. The development of speech audiometry came from the need to clinically quantify an individual's ability to hear speech (American Speech-Language & Hearing Association, 1988). Speech audiometry testing establishes a patient's threshold for speech stimuli. This testing reveals a patient's speech comprehension (Berardino et al., 2010). Speech audiometry testing enables the diagnosis of hearing losses and conditions and informs treatment plans for patients with significant hearing loss. This testing can inform the audiologist of whether or not the person's hearing loss requires the use of a hearing aid. Common conditions such as phonemic regression, a decrease in intelligibility of speech disproportionate to the pure-tone hearing loss associated with aging, can only be revealed when speech audiometry testing is completed. The audiologist takes what they have learned from pure-tone audiometry and uses speech audiometry to further clarify and interpret the data (Carhart, 1951).

Because of the difficulties of standardizing pronunciation, timing, and voice-presentation levels among presenters, recorded stimuli have become the preferred method (ASHA, 1988). However, speech stimuli may be presented live or as recordings. Using compact discs (CDs) or other computer media to store digitally recorded materials adds reliability to test scores (Studebaker, Sherbecoe, McDaniel, & Gwaltney, 1999). Speech audiometry consists of testing both SRT and WRS. Words used in the measurement of SRT are the focus of the present study.

Speech Recognition Threshold

The SRT is defined as the lowest intensity level at which a patient can correctly recognize 50% of words. This means the patient is able to correctly recognize the words presented to them—not merely detect the presence of speech. Results from the SRT enhance the audiologist's ability to properly prescribe hearing amplification devices and provide aural rehabilitation. Combined pure-tone and speech testing results show whether the patient would benefit from the use of a hearing aid based upon how much amplification is required for them to correctly recognize speech (ASHA, 1977; Carhart, 1951; Egan, 1979; Martin & Clark, 2009). A patient's SRT is used to confirm the validity of the pure-tone audiogram. Discrepancies found between the two thresholds can indicate an auditory disorder, malfunctioning equipment, or pseudohypacusis (ASHA, 1988).

Testing a patient's SRT in English typically uses spondaic words or words with equal stress on each syllable. This is because spondaic words have steeper psychometric function slopes than monosyllabic words. A steeper psychometric function slope allows clinicians to more efficiently determine the patient's SRT (Hudgins, Hawkins, Karlin, & Stevens, 1947; Martin & Clark, 2009; Ramkissoon, Proctor, Lansing, & Bilger, 2002). However, testing the SRT in other languages frequently requires the use of trisyllabic words to create equivalent steepness of psychometric function slopes (Nissen et al., 2011; Nissen, Harris, Jennings, Eggett, & Buck, 2005).

Words chosen for SRT testing should be representative of the most common words used in a language. A corpus is frequently used for word selection because it provides commonly used words that adequately represent the test population's vocabulary. Epstein (1978) and Rudmin (1987) found through their research that the most important characteristic of a word list is word familiarity to avoid testing the patient's vocabulary rather than hearing ability, which would confound the audiometric result. Age and native language are two factors that can greatly affect vocabulary familiarity within a word list. If common words in the patient's language are selected the testing avoids measuring the patient's vocabulary or intelligence rather than hearing

(ASHA, 1988; Ramkissoon, 2001). Another important aspect of word lists is homogeneity of audibility. Homogeneity of audibility requires that the words must be homogenous in their loudness presentation level across both individual phonemes and entire words (James, Bowsher, & Simpson, 1991). This means that within the word list each word must be equal in the listener's ability to recognize the word (Hudgins et al., 1947; Ramkissoon, 2001). ASHA (1988) detailed proper SRT procedure, including the requirement to use a closed-set of words in which the patient is familiarized with the word list beforehand to enhance recognition.

Word lists for testing Spanish-speaking patients have been developed to assess SRT for adults (Berruecos, 1967; Cancel, 1965; Spitzer, 1980). Word lists have also been created to assess the SRT in children (Comstock & Martin, 1984; Martin & Hart, 1978; Schneider, 1992; Spitzer, 1980). Ramkissoon and colleagues (2002) produced a speech audiometry test designed to test a multilingual population, accommodating both native-English speaking audiologists and English second language (ESL) patients from all language backgrounds using English monosyllabic digits rather than spondaic words. Ramkissoon and colleague's materials were not developed within the past decade, used analog recording rather than today's digital recording devices, and failed to address the pediatric Spanish-speaking population. Limited materials exist that test a child's SRT with normative data (Thibodeau, 2007). Without these materials, a major portion of the U.S. population would be assessed using non-ideal techniques. Therefore, there is a clear demand to develop digitally recorded speech audiometry materials with standardized audibility and steep psychometric function slopes.

The Spanish Language

Spanish, a romance language, is known as a macrolanguage because of its pervasive distribution. Spanish is the official language in 20 countries, including Argentina, Costa Rica, Honduras, Mexico, and Spain. Latin, French, Italian, Basque, and Arabic languages heavily influenced Spanish vocabulary. Spanish is the second most commonly spoken language in the United States with over 33.1 million people speaking it as their native language. Of the 55 million Hispanic United States residents age 4 and older, 38 million primarily speak Spanish in the home. In 2010, of the 308.7 million residents of the U.S., 50.5 million (16%) people were of Hispanic origin (U.S. Census Bureau, 2011).

Dialects. There are 12 dialects of Spanish separated by geographical distribution and varying in their phonology, grammar, and lexicon. Of the 12, Mexican Spanish has by far the most speakers. This dialect alone is spoken by over 20% of the world's Spanish speakers, or approximately 112 million people. Mexican Spanish is the primary Spanish dialect in the United States and Mexico (Penny, 2000). Because the Mexican Spanish dialect is the most widespread, it was chosen as the focus of this study.

Prosody. Spanish is a syllable-timed language classified by rhythm, meaning each syllable has equal duration regardless of stress. The general pattern is a falling tone on declarative sentences and wh- questions with a rising tone for yes/no questions. Stress most frequently occurs on the last syllable of a word, but can be changed by presence of an accent mark (Cressey, 1978).

Alphabet. The Spanish alphabet contains 29 letters: a, b, c, ch, d, e, f, g, h, i, j, k, l, ll, m, n, ñ, o, p, q, r, s, t, u, v, w, x, y, z. The core letters are the same as the English language with the added digraphs ch and ll. An accent mark is placed above vowels to distinguish between homophones (el or él), interrogative pronouns (que or qué), and demonstratives (este or esté) (Cressey, 1978).

Structure. Spanish uses a two-gender noun system, fifty-conjugated verb forms, and nouns, adjectives, and determiners distinguished by both number and gender. Syntax places modifiers after the noun they are describing; therefore, the main language structure is subject-verb-object (Penny, 2000).

With the expanding cultural melting pot in the United States, more children are immigrating with Spanish as their primary language and limited fluency in English. The purpose of this study was to develop and evaluate a set of Mexican Spanish trisyllabic word lists for use in the measurement of pediatric speech recognition using current digital technology and methods.

Spanish Speech Audiometry

Cancel (1965) was the first to develop Spanish speech audiometry materials in bisyllabic word lists. This test consisted of two forms of multiple-choice and was regarded as clinically acceptable. Since these first advancements in Spanish audiometry, tests were developed to specifically test discrimination in patients with Mexican and Chilean dialects in the 1960s (Berruecos, 1967). The age of these tests as well as the outdated analog recording technology used renders them linguistically and culturally inappropriate for testing patients now (Harris et al., 2007). Keller (2009) developed the most recent digitally recorded Spanish speech audiometric materials. Her research provided 28 homogeneous words that were shown to be effective in measuring the SRT of native Spanish-speaking adults.

Studies have shown that audiological testing done in a speaker's non-native language yields inaccurate results because of the speaker's unfamiliarity with the language's structure and vocabulary categorization (Berruecos, 1967; Bradlow & Pisoni, 1999). Carhart's (1951) idea that speech audiometry materials should recognize linguistic differences between languages and be developed in each person's native tongue is still widely believed today.

The Individuals with Disabilities Education Improvement Act of 2004 mandates all children be tested in their native language. The American Speech-Language-Hearing Association (ASHA) is committed to upholding linguistic diversity within audiological practice; however, speech audiometry material has not been developed for the pediatric Spanish-speaking population within the United States in 30 years (Comstock & Martin, 1984).

Pediatric Speech Audiometry

Children represent one group where altering aspects of audiometric testing is necessary. These aspects include accommodating vocabulary and response method in accordance with their age and linguistic repertoire (Comstock & Martin, 1984). The early and accurate assessment of a child's hearing abilities is essential to determine whether hearing impairment may be affecting the child's linguistic abilities (Rudmin, 1987).

Many attempts have been made to create a test to assess the speech recognition abilities of all children. However, creating a test that is relevant to all children is difficult due to the wide range of receptive language abilities. The Pediatric Speech Intelligibility Test (PSI) (Jerger, Jerger, & Abrams, 1983) is one example where picture stimulus cards representing both monosyllabic words and sentences are used to test children 3-6 years of age. The PSI mainly tests for peripheral and central disorders in children, administered using a sound field (Jerger et al., 1983). Cramer and Erber (1974) developed a spondaic recognition test for use with hearing-impaired children between the ages of 5 and 9 years. Comstock & Martin (1984) developed a picture-pointing speech recognition test for Spanish pediatric patients. This study included four lists of 25 bisyllabic words. However, Cokely & Yager (1993) assessed the four two-syllable

Spanish word lists in 1989 and found the lists were not equivalent to one another and contained incomparable performance-intensity function slopes.

Given the lack of recently developed material that tests Spanish-speaking children and the importance of testing in a child's native language, the creation of such materials proves to be a necessary step in the field of audiology. The purpose of the present investigation is to (a) construct a word list, (b) collect normative data on the words, and (e) select words that are homogeneous with respect to the psychometric function slope. These materials can be used and distributed within the United States and Central and South American countries to audiologists testing Spanish-speaking children.

Method

The methodology of this study was based on Keller's (2009) study of psychometrically equivalent trisyllabic words for speech reception testing in Spanish. Keller standardized the word list through testing native Spanish-speaking adults. Results revealed a mean 50% intelligibility point of 4.7 dB HL for male talkers and 3.0 dB HL for female talkers with a mean pure-tone average (PTA) of 5.83 dB HL. The words were successfully adjusted to equate the 50% intelligibility point with the mean PTA.

Participants

The participants in this study were native Spanish-speakers of Mexican heritage. A total of 20 subjects (12 female, 8 male) ranging in age from 4 to 8 years (M = 6.3) participated in the evaluation of the Spanish trisyllabic words. Six of the 20 participants were born in Mexico, while the other 14 were born within the United States to at least one parent native to Mexico. The participants and their parents indicated they speak Spanish on a daily basis. All participants had normal pure-tone air conduction thresholds of ≤ 15 dB HL at all octave frequencies ranging from 250 to 8000 Hz in the test ear. Additionally, each had static acoustic admittance between 0.3 and 1.4 mmhos with peak pressure between -10 and +50 daPa (ASHA, 1990). The mean pure-tone average for the 20 participants was 5.2 dB HL. Summary statistics describing the pure-tone thresholds of participants are listed in Table 1. Participants were recruited by word of mouth. A copy of the informed consent form can be found in Appendix B.

Materials

Words. Trisyllabic words were chosen as stimuli for the SRT materials. The words were selected from a corpus of most commonly used words in Spanish and tested on Spanish-speaking adults in Keller's thesis (Keller, 2009). Trisyllabic words in Spanish are equivalent to bisyllabic words in English, and therefore were used instead (Comstock & Martin, 1984). The total of 90 words were rated by six native judges on a scale of 1 to 5 based on how familiar the words would be to a native speaker of Spanish and whether the word was appropriate to use (1=extremely, 2=very, 3=average, 4=seldom used, 5=rarely used). The words ultimately selected for recording received an average score of 1 or 2, with any word scoring a 5 from any judge being eliminated from the word list. Of the 90 trisyllabic words considered, 62 were eliminated for (a) being culturally insensitive, (b) being unfamiliar, (c) representing inappropriate content, or (d) having multiple meanings. Fifteen of the twenty-eight remaining trisyllabic words were found in one or more of the following pediatric Spanish dictionaries: *Mi Primer Diccionario Larousse* (Larousse, 2011), *Mi Primer Diccionario* (Everest, 2013), and *Mi Primer Diccionario* (Parragon, 2007).

Talkers. The talker was asked to pronounce each trisyllabic word at least four times with a slight pause between each production during the recording sessions. Talkers were asked to speak at a natural rate with normal intonation patterns.

Table 1

Pure Tone Threshold (dB HL) Descriptive Statistics for 20 Normally Hearing Spanish Speaking Pediatric Participants

Frequency (Hz)	*M*	*Minimum*	*Maximum*	*SD*
125	7.7	0.0	15.0	5.2
250	7.5	0.0	15.0	5.3
500	6.8	0.0	15.0	4.7
750	5.9	0.0	15.0	4.9
1000	5.3	0.0	10.0	4.7
1500	4.5	-5.0	10.0	4.7
2000	3.5	-5.0	15.0	5.6
3000	3.6	-5.0	15.0	6.7
4000	2.3	-10.0	15.0	6.0
6000	4.5	-5.0	15.0	7.2
8000	4.8	-5.0	15.0	7.0
PTA	5.2	-1.7	11.7	4.4

Note. PTA = arithmetic average of thresholds at 500, 1000, & 2000 Hz

The first and last repetition of each word was excluded from the study to avoid possible list effects. In addition, one native judge rated the medial repetitions of each word for perceived quality of production. The best production of each word was then selected for inclusion in the selection of trisyllabic words used for testing. Any word that was judged to be a poor recording (peak clipping, extraneous noise, etc.), mispronounced, or produced with an unnatural intonation pattern was rerecorded or eliminated from the study prior to listener evaluation.

After the word selection process, the intensity of each trisyllabic word included in the test materials was edited as a single utterance using SADiE disk editor software (Version 5.2.2) to yield the same average root mean square (RMS) power as that of a 1000 Hz calibration tone in an initial attempt to equate test-word threshold audibility (Harris, Nielson, McPherson, Skarzynski, & Eggett, 2004). Each of the individually recorded and edited words was then saved as 24-bit wav file. Finally, the words were edited using Adobe Audition (Version 1.5) to remove any extraneous noise and provide consistent presentation timing.

Recordings. All recordings were made in a large anechoic chamber located in the Eyring Science Center on Brigham Young University campus in Provo, Utah, USA. A Larson-Davis model 2541 microphone (Larson Davis, Provo, UT) was positioned approximately 15 cm from the talker at a 0° azimuth and was covered by a 7.62 cm windscreen. The microphone was connected to a Larson-Davis model 900B microphone preamp, which was coupled to a Larson-Davis model 2200C preamp power supply. The signal was then digitized by an Apogee AD-8000 24-bit analog-digital converter (Apogee, Santa Monica, CA) and subsequently stored on a hard drive for later editing. A 44.1 kHz sampling rate with 24-bit quantization was used for all recordings, and every effort was made to utilize the full range of the 24-bit analog-to-digital converter. Ambient noise levels in the anechoic chamber were approximately 0 dB SPL, which allowed a signal-to-noise ratio of at least 65 dB during recording. The signal-to-noise ratio was verified by measurement of ambient noise and speech levels on the recording.

Procedures

A screener from the Contextual Probe of Articulation Competence—Spanish (CPAC-S) developed by Goldstein & Iglesias (2000) was administered to each participant prior to hearing testing. The CPAC-S eliminated the possibility of a phonological or articulation disorder, which would prevent accurate judgment of word repetition.

Randomization and timing of the word presentation was controlled using custom software from the 24-bit wav files to the external input of a Grason Stadler model 1761 audiometer (Grason Stadler, Eden Prairie, MN). The stimuli were routed from the audiometer to the subject via a single TDH-50P headphone. All testing was carried out in a double-walled sound suite that met American National Standards Institute (ANSI) S3.1 standards for maximum permissible ambient noise levels for the not covered ears condition, using one-third octave-bands (ANSI, 1999).

Prior to testing each subject, the external inputs to the audiometer were calibrated to 0 VU using a 1000 Hz calibration tone. The audiometer was calibrated prior to and at the conclusion of data collection. Audiometric calibration was performed in accordance with ANSI S3.6 specifications (ANSI, 2004). No changes in calibration were necessary throughout the course of data collection.

After passing articulation, tympanometry, acoustic reflex, and pure-tone hearing screeners, each subject participated in one test session. Due to parent preference, testing was done within 1 hour-long session. In order to minimize listener fatigue, several rest periods were given during each test session with sticker and toy incentives. The 28 trisyllabic words were presented to each of the participants beginning at -5 dB HL and ascended in 5 dB increments until the presentation level reached 15 dB HL. Keller's (2009) thesis presented words in increments of 2 dB; these data were reanalyzed to determine differences in slope between using increments of 2, 4, 6, and 8 dB. When comparing increments of 2 dB with 6 dB, no statistically significant differences were revealed within the male recordings and only one aberrant word was found within the female recordings. These statistically insignificant results confirmed the use of 5 dB increments in order to reduce listener fatigue throughout the study. The sequence of the 28 words was randomized prior to presentation at each intensity level. Each subject listened to both the male and female talker recordings of all 28 trisyllabic words, in a sequence determined randomly. Subjects repeated words verbally, which were scored as being correct (1) or incorrect (0) by a native Spanish judge. Interstimulus intervals were variable with the interval dependent on child response time and judge rating time. Once the word was scored, the next randomized word played immediately. Prior to the evaluation of the trisyllabic words, each individual was given the following instructions:

> You will hear some Spanish words, which may become louder or softer. At the very soft levels it may be difficult to hear the words. Please listen carefully and repeat the words you hear. If you are unsure of a word, you are encouraged to guess. If you have no guess, please be quiet and listen for the next word. Do you have any questions?

Results

After raw data were collected, logistic regression was used to obtain both the regression slope and the regression intercept for each of the 28 male and female trisyllabic words. These values were then inserted into a modified logistic regression equation that was designed to calculate the percent correct at each intensity level. The original logistic regression equation follows:

$$\log \frac{p}{1-p} = a + b \times i \tag{1}$$

In Equation 1, p is the proportion correct at any given intensity level, a is the regression intercept, b is the regression slope, and i is the presentation level in dB HL. When Equation 1 is solved for p and multiplied by 100, Equation 2 is obtained where P is percent correct recognition:

$$P = \left(1 - \frac{\exp(a+b\times i)}{1+\exp(a+b\times i)}\right) \times 100 \tag{2}$$

In Equation 2, P is percentage of correct recognition, a is the regression intercept, b is the regression slope, and i is the presentation intensity in dB HL. By inserting the regression slope, regression intercept, and presentation level into Equation 2, it is possible to predict the percentage correct at any specified intensity level. Percentage of correct recognition was calculated for each of the trisyllabic words for a range of -5 to 15 dB HL in 5 dB increments.

In order to calculate the intensity level required for a given proportion, Equation 1 was solved for i (see Equation 2). By inserting the desired proportions into Equation 2, it is possible to calculate the threshold (intensity required for 50% intelligibility), the slope (%/dB at threshold), and the slope (%/dB) from 20 to 80% for each psychometric function (Harris et al., 2007; Nissen et al., 2005). When solving for the threshold, Equation 2 can be simplified to Equation 3.

$$\text{Intensity required for 50\% intelligibility } (dB) = \text{——} \tag{3}$$

Calculations of threshold (intensity required for 50% intelligibility), slope at 50%, and slope from 20% to 80% were made for each trisyllabic word using the logistic regression slopes and intercepts. As the function approaches extremes, the resulting line becomes more curved; the function appears most linear at the 50% intelligibility mark. Previous endeavors to create speech audiometry materials have included psychometric calculations for both 50% and 20% to 80% intelligibility (Harris et al., 2007; Nissen et al., 2005); the inclusion of both calculations for this study ensures capability for future comparisons from this to other works. Table 2 (male talker) and Table 3 (female talker) present the participants' mean performance on each of the 28 trisyllabic words.

Thresholds for the 28 trisyllabic words ranged from 11.0 to 15.8 dB HL (M = 13.4 dB HL) for the male talker words, and from 11.8 to 18.5 dB HL (M = 14.0 dB HL) for the female talker words. Psychometric functions for each trisyllabic word were calculated with Equation 2 using the logistic regression intercept and slope values. The slopes at 50% ranged from 5.9%/dB HL to 14.4%/dB HL (M = 9.4) for the male talker, and from 3.5%/dB HL to 13.5%/dB HL (M = 7.7) for the female talker. Figure 1 (male talker) and Figure 2 (female talker) contain the predicted psychometric functions for each of the 28 words with the logistic regression slopes and intercepts being used to fit these data.

From the 28 words, a selection of 12 words with the steepest psychometric function slopes for both the male and female talker recordings were selected for inclusion in a final list of 12 selected trisyllabic words. The selected male talker words had a mean slope at 50% of 9.8%/dB HL with a mean threshold of 12.3 dB HL. In comparison, the total 28 words were 4%/dB less steep with a higher threshold by 1.7 dB. The female talker selected words had a mean slope at 50% of 8.3%/dB HL, with a mean threshold of 12.6 dB HL. In comparison,

Table 2

Mean Performance for 28 Spanish Male Talker Trisyllabic SRT Words for 20 Normally Hearing Spanish Pediatric Participants

#	Word	a[a]	b[b]	Slope at 50% [c]	Slope 20-80%[d]	Threshold[e]
1	apenas	5.03070	-0.36380	9.1	7.9	13.8
2	apoyo	6.42490	-0.53230	13.3	11.5	12.1
3	comienza	5.37450	-0.43130	10.8	9.3	12.5
4	comprender	6.07930	-0.39780	9.9	8.6	15.3
5	derecho	3.48190	-0.28730	7.2	6.2	12.1
6	ejemplo	4.85920	-0.38680	9.7	8.4	12.6
7	entonces	4.11050	-0.32050	8.0	6.9	12.8
8	figura	3.71400	-0.27580	6.9	6.0	13.5
9	general	4.86230	-0.31230	7.8	6.8	15.6
10	levantar	4.49060	-0.33640	8.4	7.3	13.3
11	manera	6.40850	-0.47480	11.9	10.3	13.5
12	minuto	6.15560	-0.47570	11.9	10.3	12.9
13	momento	3.20240	-0.23550	5.9	5.1	13.6
14	ninguno	5.40380	-0.38470	9.6	8.3	14.0
15	nosotros	4.85920	-0.38680	9.7	8.4	12.6
16	obtener	4.29010	-0.28350	7.1	6.1	15.1
17	pequeño	3.84670	-0.33130	8.3	7.2	11.6
18	perfecto	4.48510	-0.39190	9.8	8.5	11.4
19	persona	3.90170	-0.35530	8.9	7.7	11.0
20	pregunta	5.23060	-0.38970	9.7	8.4	13.4
21	primero	4.85920	-0.38680	9.7	8.4	12.6
22	programa	8.83350	-0.57630	14.4	12.5	15.3
23	propone	6.13510	-0.44290	11.1	9.6	13.9
24	proyecto	5.12870	-0.40040	10.0	8.7	12.8
25	recoger	6.56770	-0.45580	11.4	9.9	14.4
26	recordar	3.49070	-0.25500	6.4	5.5	13.7
27	tamaño	4.16700	-0.26380	6.6	5.7	15.8
28	tarea	4.73420	-0.34820	8.7	7.5	13.6
	M	5.00456	-0.37438	9.4	8.1	13.4
	Min	3.20240	-0.57630	5.9	5.1	11.0
	Max	8.83350	-0.23550	14.4	12.5	15.8
	Range	5.63110	0.34080	8.5	7.4	4.8
	SD	1.21711	0.08289	2.1	1.8	1.3

[a]a = regression intercept. [b]b = regression slope. [c]Psychometric function slope (%/dB) at 50% was calculated from 49.999 to 50.001%. [d]Psychometric function slope (%/dB) from 20-80%. [e]Intensity required for 50% intelligibility.

Table 3

Mean Performance for 28 Spanish Female Talker Trisyllabic SRT Words for 20 Normally Hearing Spanish Pediatric Participants

#	Word	a[a]	b[b]	Slope at 50% [c]	Slope 20-80%[d]	Threshold[e]
1	apenas	3.20410	-0.26820	6.7	5.8	11.9
2	apoyo	4.85290	-0.33980	8.5	7.4	14.3
3	comienza	3.40170	-0.23840	6.0	5.2	14.3
4	comprender	8.50400	-0.54090	13.5	11.7	15.7
5	derecho	3.20410	-0.26820	6.7	5.8	11.9
6	ejemplo	5.45740	-0.41820	10.5	9.1	13.0
7	entonces	5.12870	-0.40040	10.0	8.7	12.8
8	figura	4.49060	-0.33640	8.4	7.3	13.3
9	general	3.51780	-0.24100	6.0	5.2	14.6
10	levantar	3.71400	-0.27580	6.9	6.0	13.5
11	manera	4.11050	-0.32050	8.0	6.9	12.8
12	minuto	5.23060	-0.38970	9.7	8.4	13.4
13	momento	2.58740	-0.13970	3.5	3.0	18.5
14	ninguno	5.66990	-0.38640	9.7	8.4	14.7
15	nosotros	3.61100	-0.25790	6.4	5.6	14.0
16	obtener	4.16700	-0.26380	6.6	5.7	15.8
17	pequeño	3.37770	-0.26890	6.7	5.8	12.6
18	perfecto	4.27940	-0.36300	9.1	7.9	11.8
19	persona	3.74650	-0.26190	6.5	5.7	14.3
20	pregunta	3.95840	-0.31530	7.9	6.8	12.6
21	primero	4.28580	-0.32730	8.2	7.1	13.1
22	programa	3.51780	-0.24100	6.0	5.2	14.6
23	propone	3.61040	-0.21460	5.4	4.6	16.8
24	proyecto	3.69970	-0.29130	7.3	6.3	12.7
25	recoger	5.45740	-0.41820	10.5	9.1	13.0
26	recordar	3.35410	-0.20870	5.2	4.5	16.1
27	tamaño	4.16700	-0.26380	6.6	5.7	15.8
28	tarea	4.33890	-0.31440	7.9	6.8	13.8
	M	4.23731	-0.30620	7.7	6.6	14.0
	Min	2.58740	-0.54090	3.5	3.0	11.8
	Max	8.50400	-0.13970	13.5	11.7	18.5
	Range	5.91660	0.40120	10.0	8.7	6.7
	SD	1.14302	0.08177	2.0	1.8	1.6

[a]a = regression intercept. [b]b = regression slope. [c]Psychometric function slope (%/dB) at 50% was calculated from 49.999 to 50.001%. [d]Psychometric function slope (%/dB) from 20-80%. [e]Intensity required for 50% intelligibility.

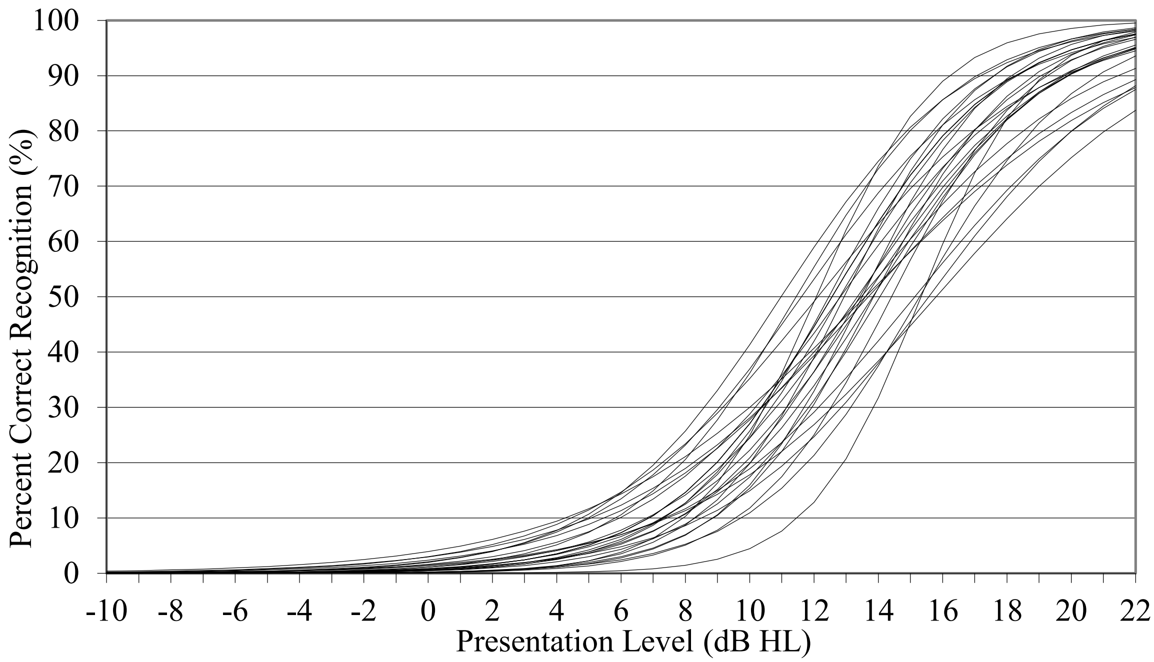

Figure 1. Mean psychometric functions for Spanish pediatric trisyllabic words for male talker recordings.

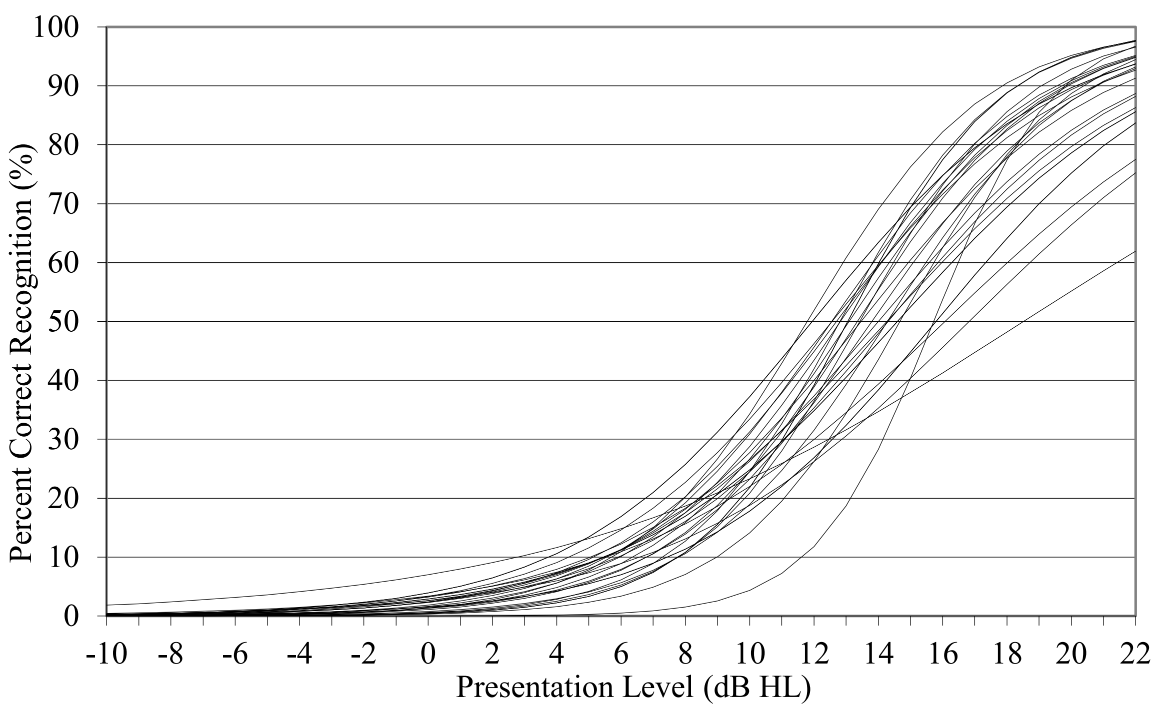

Figure 2. Mean psychometric functions for Spanish pediatric trisyllabic words for female talker recordings.

the total 28 words were .6%/dB less steep with a higher threshold by 1.4 dB. These selected word lists provide lower thresholds and steeper psychometric function slopes than the mean performances for all 28 words. The use of this selected word list would lead to a reduction in overall test time as well as improved testing reliability (Wilson & Strouse, 1999). The selected word list data are shown in Table 4 (male talker) and Table 5 (female talker).

Discussion

The first purpose of this study was to create a set of psychometrically homogeneous pediatric Spanish trisyllabic words for use in measuring the SRT of pediatric Spanish listeners. These results are not equivalent with the results from Keller's (2009) thesis testing Spanish-speaking adults. For Spanish adults, the threshold for trisyllabic SRT words was 4.7 dB HL for male talkers and 3.0 dB HL for female talkers with a mean PTA of 5.83 dB HL. The thresholds and mean PTA were equivalent and provided results indicative of valid SRT measurement. These results are displayed in Table 6 and Table 7. Conversely, results for Spanish-speaking children in this study yielded a 13.4 dB HL threshold for male talkers and 14.0 dB HL for female talkers with a mean PTA of 5.2 dB HL. This is a difference of 8.7 dB for male talkers and 11.0 dB for female talkers. Therefore, the 28 words selected for use in Keller's thesis and the current study result in several dB over the mean threshold expected for adults. Keller's resulting thresholds enabled adjustment of the words to meet the mean PTA of 5.83 dB HL. Pediatric thresholds do not leave enough available headroom on each word to permit raising their intensity to meet the mean PTA of 5.2 dB HL. Therefore, the words were not digitally adjusted to equate thresholds with the mean PTA.

Speech audiometry testing results are expected to show a mean threshold equivalent to the participant's mean PTA (Carhart, 1946). A difference of 6 dB or less between the SRT and

Table 4

Mean Performance for 12 Selected Spanish Male Talker Trisyllabic SRT Words

#	Word	a[a]	b[b]	Slope at 50% [c]	Slope 20-80%[d]	Threshold[e]
1	apoyo	6.42490	-0.53230	13.3	11.5	12.1
2	comienza	5.37450	-0.43130	10.8	9.3	12.5
3	derecho	3.48190	-0.28730	7.2	6.2	12.1
4	ejemplo	4.85920	-0.38680	9.7	8.4	12.6
5	entonces	4.11050	-0.32050	8.0	6.9	12.8
6	minuto	6.15560	-0.47570	11.9	10.3	12.9
7	nosotros	4.85920	-0.38680	9.7	8.4	12.6
8	pequeño	3.84670	-0.33130	8.3	7.2	11.6
9	perfecto	4.48510	-0.39190	9.8	8.5	11.4
10	persona	3.90170	-0.35530	8.9	7.7	11.0
11	primero	4.85920	-0.38680	9.7	8.4	12.6
12	proyecto	5.12870	-0.40040	10.0	8.7	12.8
	M	4.79060	-0.39053	9.8	8.5	12.3
	Min	3.48190	-0.53230	7.2	6.2	11.0
	Max	6.42490	-0.28730	13.3	11.5	12.9
	Range	2.94300	0.24500	6.1	5.3	1.9
	SD	0.90217	0.06705	1.7	1.4	0.6

[a]a = regression intercept. [b]b = regression slope. [c]Psychometric function slope (%/dB) at 50% was calculated from 49.999 to 50.001%. [d]Psychometric function slope (%/dB) from 20-80%. [e]Intensity required for 50% intelligibility.

Table 5

Mean Performance for 12 Selected Spanish Female Talker Trisyllabic SRT Words

#	Word	a[a]	b[b]	Slope at 50% [c]	Slope 20-80%[d]	Threshold[e]
1	apenas	3.20410	-0.26820	6.7	5.8	11.9
2	derecho	3.20410	-0.26820	6.7	5.8	11.9
3	ejemplo	5.45740	-0.41820	10.5	9.1	13.0
4	entonces	5.12870	-0.40040	10.0	8.7	12.8
5	figura	4.49060	-0.33640	8.4	7.3	13.3
6	manera	4.11050	-0.32050	8.0	6.9	12.8
7	pequeño	3.37770	-0.26890	6.7	5.8	12.6
8	persona	3.74650	-0.26190	6.5	5.7	14.3
9	pregunta	3.95840	-0.31530	7.9	6.8	12.6
10	primero	4.28580	-0.32730	8.2	7.1	13.1
11	proyecto	3.69970	-0.29130	7.3	6.3	12.7
12	recoger	5.45740	-0.41820	10.5	9.1	13.0
	M	4.26516	-0.32969	8.3	7.1	12.6
	Min	3.20410	-0.41820	6.7	5.7	11.8
	Max	5.45740	-0.26190	10.5	9.1	13.3
	Range	2.25330	0.15630	3.8	4.4	1.5
	SD	0.79120	0.05878	1.4	1.3	0.5

[a]a = regression intercept. [b]b = regression slope. [c]Psychometric function slope (%/dB) at 50% was calculated from 49.999 to 50.001%. [d]Psychometric function slope (%/dB) from 20-80%. [e]Intensity required for 50% intelligibility.

Table 6

Mean Performance for 28 Selected Adult Spanish Male Trisyllabic SRT Words

#	Word	a[a]	b[b]	Slope at 50% [c]	Slope 20-80%[d]	Threshold[e]	ΔdB[f]
	M	1.90041	-0.40238	10.1	8.7	4.7	-1.1
	Min	0.04949	-0.50602	7.3	6.3	0.1	-5.7
	Max	4.24838	-0.29004	12.7	11.0	10.0	4.2
	Range	4.19889	0.21598	5.4	4.7	9.9	9.9
	SD	0.92554	0.06320	1.6	1.4	2.2	2.2

Note. Adapted from *Psychometrically equivalent trisyllabic words for speech reception threshold testing in Spanish,* by L. Keller, 2009, Brigham Young University.
[a]a = regression intercept. [b]b = regression slope. [c]Psychometric function slope (%/dB) at 50% was calculated from 49.999 to 50.001%. [d]Psychometric function slope (%/dB) from 20-80%. [e]Intensity required for 50% intelligibility. [f]Change in intensity required to adjust the threshold of a word to the mean PTA of the subjects (5.83 dB HL).

Table 7

Mean Performance for 28 Selected Adult Spanish Female Trisyllabic SRT Words

#	Word	a[a]	b[b]	Slope at 50% [c]	Slope 20-80%[d]	Threshold[e]	ΔdB[f]
	M	1.03123	-0.34682	8.7	7.5	3.0	-2.8
	Min	-0.31338	-0.51285	7.1	6.1	-0.7	-6.6
	Max	2.92879	-0.28334	12.8	11.1	8.6	2.8
	Range	3.24217	0.22951	5.7	5.0	9.4	9.4
	SD	0.79477	0.05452	1.4	1.2	2.4	2.4

Note. Adapted from *Psychometrically equivalent trisyllabic words for speech reception threshold testing in Spanish,* by L. Keller, 2009, Brigham Young University.
[a]a = regression intercept. [b]b = regression slope. [c]Psychometric function slope (%/dB) at 50% was calculated from 49.999 to 50.001%. [d]Psychometric function slope (%/dB) from 20-80%. [e]Intensity required for 50% intelligibility. [f]Change in intensity required to adjust the threshold of a word to the mean PTA of the subjects (5.83 dB HL).

PTA is generally accepted as a clinically valid result (Carhart, 1971; Olsen & Matkin, 1991; Preece & Fowler, 1992), although ASHA indicates a difference within the range of 0.3-3.1 dB to be clinically acceptable (ASHA, 1988). According to these ranges, the present study resulted in a threshold difference that is clinically unacceptable.

Few studies have researched speech recognition thresholds for children or made the comparison between adults and children. Ramkissoon and colleagues' 2014 study of speech recognition thresholds tested spondee words compared with digits. Thirty typically developing children ages 5;0 to 8;11 years were tested using a 5-dB step size. Testing with spondee stimuli resulted in a mean PTA of 7.8 dB HL and a mean SRT of 7.7 dB HL. These results show a clinically acceptable difference between the mean PTA and SRT. The CID W-1 word list (Hirsh et al., 1952) used in Ramkissoon's study resulted in clinically valid thresholds. This study contrasts with the current study in that the monolingual English speakers had acceptable thresholds when tested using spondee words.

A French study conducted in 2006 adapted the Hearing in Noise Test for children. Sentence speech reception test results were compared between children with a mean age of 7 and children with a mean age of 11. This study was a preliminary examination of how sentence speech recognition threshold data is affected by age. The researchers found a 7 dB difference between the 7- and 11-year olds (Laroche et al., 2006). These results are consistent with the present studies finding of large threshold differences between children and adults.

A 2009 study compared adult and child speech recognition performance using a speech-in-noise test. Sixty-three children between the ages of 6 and 12 years were tested. Results revealed that adults required an 8 dB signal-to-noise ratio (SNR), children 9-12 years old required a 15 dB SNR, and 6-8 year olds required a 15+ dB SNR to to obtain a speech recognition score of 95% (Neuman, Wroblewski, Hajicek, & Rubinstein, 2009). The current results agree with this previous study's findings in that the children required significantly higher thresholds than adults in order to hear the signal.

The adults were originally tested on the full list of 90 words. Had the adults been tested with the 28-word list, their thresholds may have been closer to the children's. Performance on each stimulus item may have resulted in different thresholds and slopes because of testing using 28 word stimuli rather than 90. The adults may have had lower performance when given the smaller set of stimuli. A more limited set of stimuli may have been a reason for such high thresholds for the children.

Given that the pediatric thresholds were higher than adult results by 8.7 dB for male talkers and 11.0 dB for female talkers, task-related performance errors within the current study's methods are implicated. Performance errors may have occurred through word familiarity, experimental instructions, allowed response time, or inattention and fatigue.

Children's higher thresholds may have been due to unfamiliarity with the words, which could potentially be explained by the receptive-expressive gap theory (Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991). This theory applies to primary Spanish learners and secondary English speakers. Gibson, Oller, Jarmulowicz, and Ethington's 2012 study looked at bilingual children 5-6 years of age and the results showed significant vocabulary recognition differences among the children of the same ages. This study concluded vocabulary differences were heavily influenced by the children's attendance in English schooling. Exposure to English schooling appears to create a small receptive-expressive gap in English and a large receptive-expressive gap in the child's primary language—Spanish. Gibson and colleagues concluded that this is a natural phenomenon that primary Spanish speaking children entering English school experience.

They also believe this gap closes as the child ages. Similarly, the 20 participants in the present study are of similar ages and speak Spanish in the home while attending an English school. Their more limited Spanish expressive lexicon according to this theory may have hindered their ability to recognize and repeat the words, thus necessitating higher intensity levels than their PTA for speech recognition to occur. This research presents a possible explanation for children in the present study to have higher thresholds than adults (Gibson et al., 2012).

The instructions given during the experiment may have been too limited for the children to accurately respond in the intended way. The instructions were given verbally and matched the directions that were given to the adults in Keller's (2009) study. The children may have required customized instructions and practice sessions to clearly understand the expectations for the study. It is possible that many children did not respond at lower thresholds because they were not certain if they should guess or repeat the word when it was quiet. When the children guessed a word they usually did not guess words from the previously enclosed list. This gave the experimenters the impression that the children were unaware of the fact that they had been familiarized with the set of words they should be guessing.

Furthermore, the children may have required a longer response time to process the words they heard and verbally repeat them. Researchers have recognized that cognitive processing speeds drastically change between children and adults and also vary among young children of different ages. Processing speed, or reaction time, varies according to the cognitive load put on the child in the task and upon the developmental period the child is in (Espy, Kaufmann, McDiarmid, & Glisky, 1999; Kail, 1991; Kiselev, Espy, & Sheffield, 2009). Research should be done to assess response time required at different ages during speech audiometry and other audiological testing.

Test-retest reliability was not assessed in this study, but should be to determine the reliability of the present results. If the children were retested using the same stimulus items, their thresholds may have decreased due to familiarization with the test items and with the response demands. For test-retest reliability, the acceptable margin of error for a word list is +/- 5 dB (Jacobs, 2012).

It could be speculated that attention and fatigue may explain why the children required higher intensity stimuli for word recognition. Throughout the study, the participants appeared inattentive or fatigued while sitting in the soundbooth. Slight inattention could have greatly decreased their ability to recognize and repeat the words as they were presented.

Resulting adult thresholds were 1.13 dB below the mean PTA of 5.83 dB HL (male talkers) and 2.83 dB below the mean PTA of 5.83 dB HL (female talkers); comparatively, pediatric thresholds were 8.2 dB above the mean PTA of 5.2 dB HL (male talkers) and 8.8 dB above the mean PTA of 5.2 dB HL (female talkers). Test-retest reliability, word familiarity, experimental instructions, allowed response time, or inattention and fatigue may have contributed to the statistically significant differences between the mean thresholds and PTA of adults and children.

Conclusions and Future Research

This project represents one attempt at developing speech audiometry materials for Spanish-speaking children. Further research efforts are necessary. This study included testing conducted in quiet with a single presentation of materials on children between 4 to 8 years with normal hearing. The results show a discrepancy of 8.2 dB for male talkers and 8.8 dB for female talkers at the 50% intelligibility point in comparison to the mean PTA of 5.2 dB HL. This compares to adult results with discrepancies of 1.13 dB for male and 2.83 dB for female talkers at the 50% intelligibility point in comparison to the mean PTA of 5.83 dB HL. The significant difference in results between these two studies is most likely due to the use of a pediatric test population.

Future studies should use similar methodology to test children at increasing age increments beginning with our target population. Testing in incremental pediatric ages would determine the age at which the 50% intelligibility point for the words agrees with the mean PTA. This determination would add predictability to the audiological testing performance of children compared with adults or children of other ages. Reaction time in children during SRT testing should be researched in order to compare adult and pediatric population testing times and to define the amount of time required to validly assess a child's speech recognition abilities at specific age increments. The efforts of this study are not applicable to children of other ages, children with different Spanish dialectical backgrounds, or children with hearing loss or cochlear implants.

The purpose of this study was to use previously recorded Spanish trisyllabic words tested on adults in the measurement of the speech recognition threshold of Spanish-speaking children in order to (a) determine the words appropriateness for testing children and (b) compare psychometric functions between adults and children. Twenty-eight Spanish trisyllabic words are available on CD. This electronic source allows clinicians to select test stimuli individually for each patient. These materials serve as a baseline for future research into pediatric Spanish speech audiometry, as well as benefit all clinicians working with Spanish-speaking children across the world.

References

American National Standards Institute. (1999). *Maximum permissible ambient noise levels for audiometric test rooms*. ANSI S3.1-1999. New York, NY: ANSI.

American National Standards Institute. (2004). *Specification for audiometers*. ANSI S3.6-2004. New York, NY: ANSI.

ASHA. (1977). Guidelines for determining threshold level for speech. *ASHA, 19*, 241-243. doi:10.1044/policy.GL1988-00008

ASHA. (1988). Guidelines for determining threshold level for speech. *ASHA, 30*, 85-89.

ASHA. (1990). Guidelines for screening for hearing impairments and middle-ear disorders. *ASHA, 32*, 17-24.

Berardino, F. D., Tognola, G., Paglialonga, A., Alpini, D., Grandori, F., & Cesarani, A. (2010). Influence of compact disc recording protocols on reliability and comparability of speech audiometry outcomes: Acoustic analysis. *Journal of Laryngology & Otology, 124*, 859-863. doi:10.1017/S0022215110000782

Berruecos, P. R. J. (1967). Determination of the phonetic percent in the Spanish language spoken in Mexico City, and formation of P.B. lists of trochaic words. *International Journal of Audiology, 6*, 211-216.

Bradlow, A. R., & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *Journal of the Acoustical Society of America, 106*, 2074-2085.

Cancel, C. (1965). Multiple-choice intelligibility lists for Spanish speech audiometry. *International Journal of Audiology, 4*, 91-93.

Carhart, R. (1946). Monitored live-voice as a test of auditory acuity. *Journal of the Acoustical Society of America, 17*, 339-349.

Carhart, R. (1951). Basic principles of speech audiometry. *Acta Oto-Laryngologica, 40*, 62-71.

Carhart, R. (1971). Observations on relations between threshold for pure tones and for speech. *Journal of Speech and Hearing Disorders*, *36,* 476-483.

Cokely, J. A., & Yager, C. R. (1993). Scoring Spanish word-recognition measures. *Ear & Hearing, 14*, 395-400.

Comstock, C. L., & Martin, F. N. (1984). A children's Spanish word discrimination test for non-Spanish-speaking clinicians. *Ear & Hearing, 5*, 166-170.

Cramer, K. D., & Erber, N. P. (1974). A spondee recognition test for young hearing-impaired children. *Journal of Speech and Hearing Disorders, 39*, 304-311.

Cressey, W. W. (1978). *Spanish phonology and morphology: A generative view*. Washington, DC: Georgetown University Press.

Egan, J. P. (1979). Basic aspects of speech audiometry. *Ear, Nose, and Throat Journal, 58*, 190-193.

Epstein, A. (1978). Speech audiometry. *Otolaryngologic Clinics of North America, 11*, 667-676.

Espy, K. A., Kaufmann, P. M., McDiarmid, M., & Glisky, M. L. (1999). Executive functioning in preschool children: A-not-B and other delayed response format task performance. *Brain and Cognition, 41,* 178-199.

Everest. (2013). *Mi primer diccionario* [My First Dictionary] (1st ed.). León, Spain: Gutiérrez, A. C. L.

Gibson, T. A., Oller, D. K., Jarmulowicz, L., & Ethington, C. A. (2012). The receptive-expressive gap in the vocabulary of young second-language learners: Robustness and possible mechanisms. *Bilingualism: Language and Cognition, 15,* 102-116. doi:10.1017/SI3667289100000490

Goldstein, B. & Iglesias, A. (2000). *Contextual Probes of Articulation Competence-Spanish, First Edition*. Greenville, SC: Super Duper Publications.

Harris, R. W., Nielson, W. S., McPherson, D. L., Skarzynski, H., & Eggett, D. L. (2004). Psychometrically equivalent Polish bisyllabic words spoken by male and female talkers. *Audiofonologia, 25*, 1-15.

Harris, R. W., Nissen, S. L., Pola, M. G., McPherson, D. L., Tavartkiladze, G. A., & Eggett, D. L. (2007). Psychometrically equivalent Russian speech audiometry materials by male and female talkers. *International Journal of Audiology, 46*, 47-66. doi:10.1080/14992020601058117

Hirsh, L., Davis, H., Silverman, S., Reynolds, E., Elder, E., & Benson, R. (1952). Development of materials for speech audiometry. *Journal of Speech and Hearing Disorders, 17,* 321-337.

Hudgins, C. V., Hawkins, J. E., Karlin, J. E., & Stevens, S. S. (1947). The development of recorded auditory tests for measuring hearing loss for speech. *Laryngoscope, 57*, 57-89.

Huttenlocher, J. W., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27,* 236-248.

Individuals With Disabilities Education Improvement Act of 2004, Pub. L. No. 108-446 (2004).

Jacobs, A. M. (2012). *Test-retest reliability in the determination of the speech recognition threshold.* (Unpublished master's thesis). Brigham Young University, Provo, UT.

James, C. J., Bowsher, J. M., & Simpson, P. J. (1991). Speech audiometry: Digitization effects and the non-equivalence of isophonemic word lists. *British Journal of Audiology, 25*, 111-121. doi:10.3109/03005369109079841

Jerger, S., Jerger, J., & Abrams, S. (1983). Speech audiometry in the young child. *Ear & Hearing, 4*, 56-66.

Kail, R. (1991). Developmental changes in speed of processing during childhood and adolescence. *Psychological Bulletin, 109,* 490-501.

Keller, L. (2009). *Psychometrically equivalent trisyllabic words for speech reception threshold testing in spanish* (Unpublished master's thesis). Brigham Young University, Provo, UT.

Kiselev, S., Espy, K., & Sheffield, T. (2009). Age-related differences in reaction time task performance in young children. *Developmental Cognitive Neuroscience Laboratory - Faculty and Staff Publications, 44,* 150-166.

Laroche, C., Vaillancourt, V., Melanson, C., Renault, M. E., Theriault, C., Soli, S. D., & Giguere, C. (2006). Adapting the HINT (hearing in noise test) for canadian francophone children, and preliminary data on the effects of age. *Journal of Speech-Language Pathology and Audiology, 30,* 95-109.

Larousse. (2011). *Mi primer diccionario Larousse* [My First Larousse Dictionary] (1st ed.). Barcelona, Spain: Girard, F.

Martin, F., & Hart, D. B., (1978). Measurement of speech thresholds of Spanish-speaking children by non-Spanish speaking clinicians. *Speech Hearing Disorders, 43*, 255-262.

Martin, F. N., & Clark, J. G. (2009). *Introduction to audiology.* Boston, MA: Allyn and Bacon.

Neuman, A.C., Wroblewski, M., Hajicek, J., & Rubinstein, A. (2009). Combined effects of noise and reverberation on speech recognition performance of normal-hearing children and adults. *Ear & Hearing, 31,* 336-344.

Nissen, S. L., Harris, R. W., Channell, R. W., Conklin, B., Kim, M., & Wong, L. (2011). The development of psychometrically equivalent Cantonese speech audiometry materials. *International Journal of Audiology, 50*, 191-201. doi: 10.3109/14992027.2010.542491

Nissen, S. L., Harris, R. W., Jennings, L., Eggett, D. L., & Buck, H. (2005). Psychometrically equivalent trisyllabic words for speech reception threshold testing in Mandarin. *International Journal of Audiology, 44*, 391-399.

Olsen, W. O. & Matkin, N. D. (1991). Speech audiometry. In Rintelmann (Ed.), *Hearing Assessment* (pp. 39-140). Austin, TX: Pro-Ed.

Parragon. (2007). *Mi primer diccionario* [My First Dictionary] (1st ed.). Beijing, China: Root, B.

Penny, R. J. (2000). *Variation and change in Spanish*. New York, NY: Cambridge University Press.

Preece, J. P. & Fowler, C. G. (1992). Relationship of pure-tone averages to speech reception threshold for male and female speakers. *Journal of American Academy of Audiology, 3,* 221-224.

Ramkissoon, I. (2001). Speech recognition thresholds for multilingual populations. *Communication Disorders Quarterly, 22*, 158-162.

Ramkissoon, I., Proctor, A., Lansing, C. R., & Bilger, R. C. (2002). Digit speech recognition thresholds (SRT) for non-native speakers of English. *American Journal of Audiology, 11*, 23-28. doi:10.1044/1059-0889(2002/005)

Ramkissoon, I., Estis, J. M., & Flagge, A. G. (2014). Digit speech recognition threshold (SRT) in children with normal hearing ages 5-8 years. *American Journal of Audiology, 23,* 182-189. doi:10.1044/2014_AJA-13-0039

Rudmin, F. (1987). Speech reception thresholds for digits. *Journal of Auditory Research, 27*, 15-21.

Schneider, B. S. (1992). Effect of dialect on the determination of speech-reception thresholds in Spanish-speaking children. *Language, Speech, and Hearing Services in Schools, 23*, 159-162.

Spitzer, J. B. (1980). The development of a picture speech reception threshold test in Spanish for use with urban U.S. residents of Hispanic background. *Journal of Communication Disorders, 13*, 147-151. doi:10.1016/0021-9924(80)90031-3

Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., & Gwaltney, C. A. (1999). Monosyllabic word recognition at higher-than-normal speech and noise levels. *Journal of the Acoustical Society of America, 105*, 2431-2444.

Thibodeau, L. M. (2007). Speech audiometry. In R. J. Roeser, M. Valente, & H. Hosford-Dunn (Eds.), *Audiology diagnosis* (2nd ed., pp. 288-313). New York, NY: Thieme.

U.S. Census Bureau. (2011). *The Hispanic population: 2010.* Retrieved from: http://www.census.gov/prod/cen2010/briefs/c2010br-04.pdf

Wilson, R. H., & Strouse, A. (1999). Psychometrically equivalent spondaic words spoken by a female speaker. *Journal of Speech, Language, and Hearing Research, 42*, 1336-1346.

APPENDIX A

Annotated Bibliography

American National Standards Institute. (1999). *Maximum permissible ambient noise levels for audiometric test rooms*. ANSI S3.1-1999. New York, NY: ANSI.

Summary: This standard specifies the maximum permissible ambient noise levels allowed within an audiometric test room. Maximum permissible ambient noise levels are specified from 125 to 8000 Hz in octave and one-third octave band intervals for both testing conditions (soundbooth and earphone testing) and across three test frequency ranges (125 to 8000 Hz, 250 to 8000 Hz, and 500 to 8000 Hz).
Relevance to Current Work: This standard applies to all persons who test hearing as well as all persons who install, design, or manufacture audiometric test booths. The sound booth used for this study exceeded the requirements set by this standard.

American National Standards Institute. (2004). *Specification for audiometers*. ANSI S3.6-2004. New York, NY: ANSI.

Summary: This standard ensures consistent reliability among individual test results across audiometers that comply with the standard. Compliance with the standard includes requirements for safety, equipment setup, and calibration. Audiometer output levels for testing speech audiometry are specified to present stimuli at known and calibrated levels.
Relevance to Current Work: The equipment used in this study complied with this ANSI standard to ensure reliable and reproducible results for other clinicians.

ASHA. (1977). Guidelines for determining threshold level for speech. *ASHA, 19*, 241-243. doi:10.1044/policy.GL1988-00008

Summary: This standard presents a standardized procedure that is clinically fast, reliable, and flexible for a variety of testing situations. The standard helps improve interclinician and interclinic data reliability. These guidelines give a brief history of speech audiometry, including that recorded presentation is preferred over live voice. The guidelines discuss proper calibration of equipment, test environment, testing materials, and recording results with references to other ASHA standards. To determine the threshold, specific instructions and familiarization of the test materials must be given; then, each spondaic word is to be presented in ascending 10 dB increments until a correct response is given. This becomes the starting level and all words are presented in 5 dB descending increments.
Relevance to Current Work: The guidelines presented were followed and implemented in the current study to enhance clinical reliability.

ASHA. (1988). Guidelines for determining threshold level for speech. *ASHA, 30*, 85-89.

Summary: This article provides guidelines for determining speech threshold levels. The article also defines common terminology and recommends procedures for determining the threshold in order to improve clinical data reliability. Speech audiometry was originally designed to validate pure-tone test results due to a high correlation between speech and pure-tone thresholds. Speech audiometry uses an audiometer to transduce speech according to established standards for calibration. Speech audiometry is conducted in a specific testing environment with set criteria for background noise. Standard test materials include homogeneous spondaic words. Age and language of the test population should change selection of stimulus words. Patients are tested through repeating the stimulus word from a closed set of choices. Recorded presentation of test stimuli is preferred over live voice to minimize variability among patients and audiologists. The patient should be given set instructions prior to completing the task, including instructions to listen carefully and guess. The word list should be disclosed to the patient so that the vocabulary is familiar. Testing begins with one word presented at 30-40 dB above the estimated speech recognition threshold. A correct response leads to descending in 10 dB steps until the client responds incorrectly. At an incorrect response, a second word is presented at the same level. When two consecutive words are missed at the same hearing level, increase the level by 10 dB—this is the starting level. Present two words at the starting level and at each successive 2 dB decrement. The test concludes when the client responds incorrectly to five of the last six words presented. The threshold is calculated by subracting the total number of responses from the starting level and adding one.
Relevance to the Current Work: This standardized procedure for determining SRT served as a basis for the procedure of the current study.

ASHA. (1990). Guidelines for screening for hearing impairments and middle-ear disorders. *ASHA, 32*, 17-24.

Summary: ASHA gives guidelines for identification of hearing impairments or middle-ear disorders. The guidelines are most suitable for both children and young adults. The steps include obtaining a case history, inspecting the ear, and getting acoustic immittance measurements. Screening individuals include measurement of static admittance, volume of the ear canal, and tympanometric width. To obtain static admittance, subtract the ear-canal volume from the peak admittance. Normal range of ear-canal volume for adults is between 0.6-1.5 cm^3. Poor static admittance and abnormal ear-canal volume can indicate tympanic membrane perforation. Performing the acoustic reflex must be completed to help screen for middle-ear disorders. Absence of an acoustic reflex is not wholly linked to middle-ear disorders alone.
Relevance to the Current Work: The current study screened participants using acoustic immittance measurements.

Berardino, F. D., Tognola, G., Paglialonga, A., Alpini, D., Grandori, F., & Cesarani, A. (2010). Influence of compact disc recording protocols on reliability and comparability of speech audiometry outcomes: Acoustic analysis. *The Journal of Laryngology & Otology, 124*, 859-863. doi:10.1017/S0022215110000782

Objective: The purpose of this article was to compare various compact disc recording protocols used to prepare speech test material to assess how they affect the reliability of speech audiometry testing.
Design: Acoustic analysis was performed on all compact discs. Normal hearing participants were tested using the differently prepared compact disks and the psychometric curves were compared.
Study Sample: Twelve (six women and six men) normal hearing participants were used. They had a mean age of 23 years. All subjects had normal results for otological evaluation, pure tone audiometry, and immittance audiometry. They all had a mean hearing level of 10 dB HL or less.
Results: The authors found that there were major differences between the levels at which the speech material and the calibration signal had been recorded. Overall SRT and intelligibility thresholds had statistically significant differences between compact disks ($p < 0.05$).
Conclusions: The authors concluded that to realize reliable data, recording gains used to prepare compact discs for recording should be checked and compensated for.
Relevance to Current Work: The use of compact discs and evaluation of psychometric functions are similar in scope to the current study.

Berruecos, P. R. J. (1967). Determination of the phonetic percent in the Spanish language spoken in Mexico City, and formation of P.B. lists of trochaic words. *International Journal of Audiology, 6*, 211-216.

Objective: The purpose of this study was to develop word lists for testing speech audiometry in both Mexican and Chilean dialects.
Design: A list of words was selected using common use dictionaries for the Spanish language. Two recordings of the words were made, one using a native Mexican-Spanish speaker and the other using a native Chilean-Spanish speaker. The words chosen had a trochaic stress pattern to enhance audibility.
Study Sample: Twenty participants were used in this study with normal hearing of 15 dB HL or less across all octave and mid-octave frequencies (125-8000 Hz). Ten of the participants spoke with a Chilean dialect and the other ten spoke with a Mexican dialect.
Results: The dialectical differences between the word recordings did not cause a statistically or clinically significant difference in thresholds across all participants.
Conclusions: This list of words is of viable use to clinicians among all Chilean and Mexican dialects.
Relevance to Current Work: This study included development of a word list to be given to native Mexican-Spanish speakers.

Bradlow, A. R., & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *Journal of the Acoustical Society of America, 106*, 2074-2085.

Objective: The purpose of this study was to understand the cooccuring effects of talker, listener, and item-related factors influencing speech perception testing through development of digitally recorded spoken words.
Design: Two lists of words were produced at three varying speaking rates. One list contained words with few phonetically similar sounding neighbors and the other contained harder words with many phonetically similar sounding neighbors. These lists were tested on 10 participants and the intelligibility data was obtained across two different experiments. Experiment 1 received the data from native speakers of English. Experiment 2 gathered intelligibility data from non-native speakers of English.
Study Sample: Experiment 1 included 20 (6 male, 14 female) native English listeners recruited from Indiana University between the ages of 20 and 42 years. Experiement 2 included 20 (8 males, 12 females) non-native english speakers recruited from Indiana University ranging in age from 21 to 33 years, having studied English for at least 2 years. Their ethnic backgrounds included 6 Koreans, 4 Chinese, 3 Russians, 2 Japanese, 2 Spanish, 1 Bengali, 1 Nepali, and 1 Dani. All participants reported no known speech or hearing impairments.
Results: Experiment 1 results indicated the following: lexically easier words scored much higher levels of intelligibility among native English speakers, slow and medium rate words had higher intelligibility scores than fast rate words, and that familiarity with the speaker's voice led to higher intelligibility scores. Experiment 2 results indicated the following: the ability to recognize surface phonetic information was a skill non-native English speakers used throughout testing, lexically hard words were difficult no matter the rate.
Conclusions: While native English speakers receive higher levels of intelligibility due to vocabulary familiarity, non-native English speakers still perform well because of the skill of comprehending surface phonetic information; however, due to less familiarty with the vocabulary, no amount of speaker voice familiarization or change of rate could make up for the unfamiliarity with the words contained in the recorded samples.
Relevance to Current Work: The use of non-native speakers of English and recorded word lists are similar in scope to the current study.

Cancel, C. (1965). Multiple-choice intelligibility lists for Spanish speech audiometry. *International Journal of Audiology, 4*, 91-93.

Objective: The purpose of this study was to develop a word list consisting of familiar words in the Spanish language and to add more difficult vocabulary to create a multiple-choice response form to test speech discrimination.
Design: Newspapers were used to find familiar vocabulary. From an original sample of 2,962 words, 1,000 words were chosen for the final list. The 1,000 words were broken up into 20 lists of 50 words each. Ten native Spanish speakers recorded the words onto CD. The lists were played to the participants using the signal-to-noise ratio of 10 dB then in a condition of quiet.

Study Sample: Sixty-five Spanish subjects living in America participated.
Results: The correlation coefficient between the values for listening in noise and in quiet was .53. These results correlate with a previous study testing English words. An analysis of variance showed no statistically significant differences among the individual lists.
Conclusions: Reasons for errors consisted of difficulty with the accented syllable, the presence of four choices, terminal vowels rather than terminal consonants in a word, the occurrence of vowel-consonant rather than consonant-vowel word order, and the presence of /f/ in a word.
Relevance to Current Work: This study tested Spanish-speaking participants living in the United States.

Carhart, R. (1946). Monitored live-voice as a test of auditory acuity. *Journal of the Acoustical Society of America, 17*, 339-349.

Objective: The purpose of this study was to determine the correlation between loss for a pure tone and speech audiometry, determine the numerical correspondence between these two measures, and to what degree the hearing loss can be predicted between the two measures.
Design: The SRT was measured for each participant in both their ear with significant hearing loss and their better ear. Coefficients of correlation were measured for each ear across the frequency range. The relationship between the pure-tone and speech testing were measured and compared between both of the participants ears.
Study Sample:
Results: Participants included 150 patients with significant hearing loss.
Conclusions: Test-retest reliability for speech reception was as good as the corresponding reliability for pure tones across the participants.
Relevance to Current Work: The present study performed SRT and PTA for all participants.

Carhart, R. (1951). Basic principles of speech audiometry. *Acta Oto-Laryngologica, 40*(1-6), 62-329.

Summary: The author explores the reasons why speech audiometry is such an important development, including that it must be appropriately standardized. He states that speech audiometry helps share the responsibility for educational and rehabilitational programs for those with permanent auditory impairments. Language and auditory capacities differ from person to person, and the tests need to be standardized to meet the correct language. Carhart describes the foundation for speech audiometry at the Bell Telephone Laboratories headed by Fletcher where clinical applications of speech audiometry were investigated in detail. The author emphasizes that speech audiometry includes standardized samples of a language prsented through a calibrated system to measure some aspect of hearing ability. Audiometric curves are noted as the way to report the hearing results from testing, using a method called articulation function. A normal curve should be steep. The author further defines the testing by stating the materials are designed to measure the patient's shift in threshold for speech to tell us the patient's unique presentation level required for them to audibly hear speech. This testing is

especially valid among elderly patients with hearing loss because of phonemic regression. Two criterion for a good discrimination test are using the phonetic range of the language being used and having a distribution of phonetic elements to duplicate the occurrence of these elements in the spoken version of the language being used.
Relevance to Current Work: This chapter gives a thorough overview of the methods and purpose behind speech audiometry testing.

Carhart, R. (1971). Observations on relations between threshold for pure tones and for speech. *Journal of Speech and Hearing Disorders*, *36,* 476-483.

Objective: The purpose of this study was to determine the relationship between PTA and SRT during pure-tone audiometry and speech audiometry.
Design: Participants were tested to ensure normal hearing. The PTA and SRT for each participant was tested. The SRT was tested using spondees. Testing was completed for the 500-2000 Hz frequency range. Two different groups were tested with a differnet number of subjects. Their correlation coefficients were determined as well as their SRT.
Study Sample: Two groups were tested. Group 1 had 341 Ears and Group 2 had 200 Ears. All participants were adults with normal pure-tone thresholds.
Results: The 500 and 1000 Hz thresholds were the most intimately related to the threshold. The beta coefficients are as follows: 0.45 and 0.54 for the first group and 0.46 and 0.49 for the second group.
Conclusions: It appears that the SRT can be reasonably concluded and predicted from PTA data on a subject.
Relevance to Current Work: The present study compared subjects' PTA with their SRT.

Cokely, J. A., & Yager, C. R. (1993). Scoring Spanish word-recognition measures. *Ear & Hearing, 14*, 395-400.

Objective: The purpose of this study was to measure whether or not English-speaking audiologists with no knowledge of Spanish could accurately judge listeners' oral responses to Spanish word recognition measures.
Design: The word recognition scores were determined for 10 native Spanish-speaking listeners including their oral responses to Auditec lists A-D including bisyllabic Spanish words. An analysis of variance was performed on the data. The oral responses were scored by English-speaking judges with a knowledge of Spanish and native English-speaking judges without a knowledge of Spanish. The absolute values of the numeric differences between the word recognition scores were determined.
Study Sample: Phase 1 (word recognition scores) used 10 Spanish-speaking adults (6 men, 4 women) aged 27 to 45 years (M = 32.3). Eight of the ten subjects had pure-tone thresholds of 10 dB L or less at octave frequencies from 250 to 8000 Hz. Two of the ten had thresholds of 25 dB HL at 4000 and 8000 Hz. Phase 2 included 30 subjects divided into two groups of 15 judges. The first group of judges included English-speaking listeners with no knowledge of Spanish. The second group of judges included English-speaking listeners with 2-3 years of college-level Spanish courses. All judges exhibited pure-tone thresholds of 10 dB HL or less at octave frequencies (250-8000 Hz).

Results: The absolute values from the judges results were measured and revealed a statistically significant difference ($p < .01$) between the judges unfamiliar with Spanish and the judges familiar with Spanish. The statistic significane is not linked to clinical significance. The WRS differed on average by two percent scored by judges with knowledge of Spanish and by three ercent when scored by judges with no knowledge of Spanish.
Conclusions: The data indicate that judges with no knowledge of Spanish can adequately and accurately score oral responses to Spanish word recognition measures for collecting WRS data.
Relevance to Current Work: This study also used Spanish-speaking participants to conduct speech audiometry testing.

Comstock, C. L., & Martin, F. N. (1984). A children's Spanish word discrimination test for non-Spanish-speaking clinicians. *Ear & Hearing, 5*, 166-170.

Objective: The purpose of this study was to develop a speech discrimination test to be adminstered to Spanish-speaking children.
Design: A picture-pointing speech discirmination test was developed. The test consisted of four lists of 25 Spanish bisyllabic words. Bisyllabic words were chosen because of the limited number of monosyllabic words in Spanish. The words were recorded into Spanish on the first channel of an audiotape and in English on the second channel. The dual recording was used to present the Spanish stimuli to the subjects at the same time the clinicians monitored the responses in English. Words were recorded by a native Spanish speaker from Texas. The illustrations for the picture pointing included black and white line drawings.
Study Sample: For recording the words, fifteen adults were used who passed a pure-tone screening test at 10 dB HL between 250 through 4000 Hz and were also native Spanish speakers raised in Texas. Twenty children who were native Spanish speakers and Texas natives between the ages of 3 and 8 years participated.
Results: The four word lists were found to be equivalent in sensation level used for clinical evaluation. The testing revaled increases in discrimination scores with increased age. Missed words appeared to be the result of limited vocabularies. The children were found able to listen to the recorded stimuli and point at the appropriate pictures.
Conclusions: This test was found to be a useful way to test Spanish-speaking children's speech discimnination ability. It also proved generally successful to have an English-speaking clinician perform the test.
Relevance to Current Work: The current study tested the speech discrimination of children.

Cramer, K. D., & Erber, N. P. (1974). A spondee recognition test for young hearing-impaired children. *Journal of Speech and Hearing Disorders, 39*, 304-311.

Objective: The purpose of this study was to create materials and evaluate hearing-impaired childrens' ability to recognize familiar speech material.
Design: A stimulus set was used consisting of 10 bisyllabic spondaic words. The list was not phonetically balanced. The words were selected based on their familiarity to

children. Picture cards were used to represent the test items. The words were recorded onto tape. Children were tested individually in the child's better ear and then in their poorer ear. The children were tested on consecutive days until their performance showed no improvements. They defined stable performance as four successive recognition scores within a range of 20% of each other.
Study Sample: Fifty eight hearing-impaired children (30 girls, 28 boys) ranging in age from five to nine with average pure tone thresholds ranging from 52 to 123 dB were tested.
Results: The recognition scores are bimodally distributed. One set of scores clusters around 0-65% and the other around 66-100%. The spondee recognition scores were predictable based on the children's pure-tone thresholds. For a few of the children, spondee scores could not be predicted by the pure-tone thresholds for both ears. Stable performances across testing sessions were notable in participants with pure-tone averages better than 93 or poorer than 103 dB. Inconsistent performance across test sessions was noted for children with a range of 93-103 dB.
Conclusions: The results indicate that this spondee recognition test can be used to test young hearing-impaired children ranged in age from five to nine years.
Relevance to Current Work: This study represents a work that tested young children using pre-recorded materials based on familiar vocabulary for children.

Cressey, W. W. (1978). *Spanish phonology and morphology: A generative view*. Washington, DC: Georgetown University Press.

Summary: Cressey gives a summary of the phonology and morphology of the Spanish language with detailed examples. Information includes principles of phonological analysis, the systematic phonemes of Spanish, wordlevel phonology, consonants, inflectional morphology, and feature classificaitons of the sounds of Spanish. Tables and graphs as well as detailed images are used to further knowledge of inflectional phonology.
Relevance to Current Work: This book helped guide both word selection and judging the correctness of participant responses.

Egan, J. J. (1979). Basic aspects of speech audiometry. *Ear, Nose, and Throat Journal, 58*, 190-193.

Summary: This article details the importance of speech audiometry testing and further explains the nature and methods of testing both speech recognition and speech discrimination thresholds. The method begins with the individual having the list of words disclosed to them. Next, the list of spondee words is presented beginning 25 dB above the patient's pure-tone average and decreased by 5 dB as subsequent words are presented. This progression continues until the patient responds incorrectly or gives no response; at this point, the level is increased by 5 dB. This process attempts to find the point at which the patient is able to correctly identify the words 50 percent of the time. The SRT and word discrimination score aid in determining a patient's degree and type of hearing loss and can determine a patient's candidacy for aural rehabilitation or hearing aids. Additionally, the SRT helps determine the reliability of pure-tone thresholds.

Relevance to the Current Work: This article provides justification for obtaining the SRT when conducting audiological exams.

Epstein, A. (1978). Speech audiometry. *Otolaryngologic Clinics of North America, 11*, 667-676.

Summary: The author discusses the reasons speech audiometry was groundbreaking for the field of audiology. Speech audiometry directly addresses the extent and nature of a hearing loss. The speech reception threshold can help direct the audiologist in how to utilize their residual hearing. The speech recognition threshold should be in agreement with the pure-tone thresholds. The author lists that the words included in the lists should be homogeneous, familiar, phonetically dissimilar, and encompasss a normal sample of English speech sounds. A sample list of bisyllabic words to be used for testing are given. Epstein points out that a speech reception threshold will not be consistent when a person has an extremely low discrimination score of less than 50 percent. The results from the SRT will give an indication to how well a person can follow conversational speech. Epstein states that he prefers the W-22 lists recorded by Hirsh for testing the SRT.
Relevance to Current Work: This chapter focuses on the importance of SRT and speech audiometry.

Espy, K.A., Kaufmann, P.M., McDiarmid, M., & Glisky, M.L. (1999). Executive functioning in preschool children: A-not-B and other delayed response format task performance. *Brain and Cognition, 41,* 178-199.

Objective: The purpose of this study was to measure the executive/frontal lobe functioning of preschool children in order to conclude age-related performance of working memory and inhibition processes.
Design: The A-not-B (AB) task, delayed alternation, spatial reversal, color reversal, and self-control tasks were used. Two factors were loaded on with measures from the self-controla nd delayed alternation tasks. Infants observed and retrieved a reward hidden in location A, then the reward is hidden in location B. Children were tested individually in a private room. Rewards used were cheerios, rice krispies, raisins, stickers, and pennies. Ten trials were administered in the order of presentation of AABBAABBAA. A linear regression design measured the age-related performance on each dependent measure.
Study Sample: One hundred and seventeen children aged 23 to 66 months with a mean age of 40 months participated. Fifty-five children were female and sixty-two children were male. The majority of the children were caucasian. All participants had achieved typical milestones.
Results: Age significantly predicted performance on the cognitive tasks. ABCORR was found to be .66 trials, meaning that children are expected to score an additional .66 trials per age year. Age significantly predicted DACORR and SRCORR.
Conclusions: These findings suggests that tasks adapted from neuroscientology are helpful in assessing age-related factors and brain-behavior links in children.
Relevance to Current Work: Children of different ages were tested in their ability to switch between different cognitive tasks.

Everest. (2013). *Mi primer diccionario* [My First Dictionary] (1st ed.). León, Spain: Gutiérrez, A. C. L.

Summary: This is a Spanish dictionary intended for children ages 6 and up. The dictionary contains colorful pictures next to over 1200 spanish vocabulary words. The entries give the part of speech and an example sentence.
Relevance to Current Work: This pediatric Spanish dictionary was used to cross-reference the word list to decide how familiar with the words we could expect the children to be.

Gibson, T.A., Oller, D.K., Jarmulowicz, L., & Ethington, C.A. (2012). The receptive-expressive gap in the vocabulary of young second-language learners: Robustness and possible mechanisms. *Bilingualism: Language and Cognition, 15,* 102-116. doi:10.1017/SI3667289100000490

Objective: The purpose of this study was to explore differences in a child's expressive and recpetive langauge in both their first language (L1) Spanish and second language (L2) English.
Design: Each child was given standardized vocabulary testing (WLPB-R, PPVT-III, TVIP), expressive vocabulary testing (WLPB-R), and receptive vocabulary testing (PPVT-R).
Study Sample: Kindergarten children including 222 Hispanic bilingual children, 133 monolingual English-speaking children.
Results: Bilingual children had an expressive vocabulary L1 score of 67.21 with a receptive average of 88.23—this is an expressive-receptive gap of 21.02. Bilingual children on L2 testing had an expressive-receptive gap of only 7.33 points. This difference is statistically significant. The monolingual English-speaking children had an expressive-recepive gap of 6 points with their expressive score being higher than the receptive score.
Conclusions: Bilingual children have a significant gap in their L1 between expressive and receptive language that is not shown in their L2. This is a natural phenomenon experienced by children immersed in a L2 school.
Relevance to Current Work: The current study tested bilingual children with L1 of Spanish and L2 of English. The child's expressive and receptive langauge were indirectly tested in the study.

Goldstein, B. & Iglesias, A. (2000). *Contextual Probes of Articulation Competence-Spanish, First Edition*. Greenville, SC: Super Duper Publications.

Summary: This is a norm-referenced articulation test intended to be used on Spanish-speaking children beginning at the age of 3.
Relevance to Current Work: This articulation screener was used to screen each child prior to speech recognition testing.

Harris, R. W., Nielson, W. S., McPherson, D. L., Skarzynski, H., & Eggett, D. L. (2004). Psychometrically equivalent Polish bisyllabic words spoken by male and female talkers. *Audiofonologia, 25*, 1-15.

Objective: The purpose of thist study was to develop, digitally record, and psychometrically equate Polish bisyllabic words to measure the SRT.
Design: A total of 260 bisyllabic Polish words were selected from frequency usage dictionaries. Of the 260 words considered, 70 were selected for recording and evaluation. A male and female native Polish speaker recorded the words. The words were played to each subject at 15 different intensity levels.
Study Sample: Twenty-six participants (7 male & 19 female) were used in this study. All participants were natives of Poland and ranged in age from 20 to 29 years (M = 23.7 years). Each participant had pure-tone air-conduction thresholds of 15 dB HL at octave and midoctave frequencies (125-8000 Hz) with static acoustic admittance between 0.3 and 1.4 mmhos with peak pressure between -100 and +50 daPa.
Results: Thresholds for the words ranged from 4.3 to 15.6 dB HL (M = 2.1 dB HL) for the male words and from 3.7 to 14.0 dB HL (M = 5.3 dB HL) for the female words. The slope from 20 to 80% for each psychometric function had a range of 2.7 to 12.0 %/dB for the male words and a range of 4.7 to 11.9 %/dB for the female words.
Conclusions: This study resulted in the development of 25 homogeneous bisyllabic Polish words for measuring the SRT. The words are available on CD.
Relevance to Current Work: The methods are similar in scope to the current study.

Harris, R. W., Nissen, S. L., Pola, M. G., McPherson, D. L., Tavartkiladze, G. A., & Eggett, D. L. (2007). Psychometrically equivalent Russian speech audiometry materials by male and female talkers. *International Journal of Audiology, 46*, 47-66.

Objective: The purpose of this study was to develop and evaluate Russian speech audiometry materials for testing word recognition and SRT.
Design: Native Russian talkers (male and female) digitally recorded common monosyllabic and bisyllabic Russian words. These recordings were then evaluated by 20 native Russian listeners.
Study Sample: Twenty native Russian speakers were recruited to create materials appropriate for a wide range of Russian speakers. Participants ranged in age from 16 to 50 years. Participants passed an otoscopic evaluation and ipsilateral acoustic reflex of 95 dB HL or better in the test ear at 1000 Hz. All participants had normal pure-tone air-conduction thresholds of ≤15 dB HL between 125 to 8000 Hz. Participants had typical static acoustic admittance between 0.3 and 1.4 mmhos with peak pressure between -10 and +50daPa.
Results: Psychometric functions were calculated for all words using logistic regression. Twenty-five bisyllabic words had steep psychometric function slopes and were selected. These weords were then digitally equated in intensity to match the PTA of the native listeners, resulting in speech reception thershold materials.
Conclusions: The chi-square analysis showed that the word lists were relatively homogenous for use on typically hearing listeners. The analysis also found that these words are similar in psychometric function slope to speech audiometry materials created

in English. Digital recordings of these speech audiometry materials are available on compact disc.
Relevance to the Current Work: This study is similar in use of digital recordings by native speakers, selection and evluations of materials, and development of psychometrically equivalent audiologial materials.

Hirsh, I. J., Davis, H., Silverman, S. R., Reynolds, E. G., Eldert, E., & Benson, R. W. (1952). Development of materials for speech audiometry. *Journal of Speech and Hearing Disorders, 17*, 321-337.

Objective: The purpose of this study was to create a series of audiological tests to replace the recorded PAL Auditory Tests 14 and 9 and Egan's phonetically balanced (PB) lists. The resulting tests were entitled C.I.D. Auditory Test W-1, C.I.D. Auditory Test W-2, and C.I.D. Auditory Test W-22.
Design: Test W-1 consisted of six randomizations of a single list of 36 spondaic words, which were narrowed down from a set of 84 spondaic words. These words were selected by a panel of judges based on familiarity and were then recorded by an adult male. Test W-1 was used to test speech audibility threshold of six listeners during the preliminary test of the original 84 words. Listeners were asked to repeat each word after it was played at 4, 2, 0, -2, -4, and -6 dB relative to the threshold that had been obtained for PAL Test 9. After the raw data were collected, words that were judged to be difficult were eliminated, resulting in 36 spondaic words. These words were arranged into six different randomizations. A secondary test similar to the preliminary test was conducted, this time with experienced and inexperienced listeners in order to develop tentative speech thresholds. Based on listener response, words deemed to be more difficult or easy than the mean were adjusted down or up by 2 dB.

Test W-2 was designed for a rapid estimate of the threshold of intelligibility. For Test W-2, the same 36 words were employed, but the intensity of the words was attenuated within each list of 3 dB every three words, to allow for faster test pacing. Each participant listened to all six randomizations of W-2 to obtain thresholds.

Test W-22 consisted of 200 monosyllabic words divided into four 50-word lists, each of which was phonetically balanced. This test was intended to determine a listener's discrimination loss for speech. Five people rated the words in the 20 Psycho-Acoustic Laboratory PB-50 lists for familiarity, resulting in a selection of 120 words judged to be very familiar. An additional 80 words were added from various sources, resulting in 200 total words. These words were cross-checked using Thorndike's list to ensure that they were among the most common English words. Once the words were selected and phonetically balanced, a male talker recorded each list onto magnetic tape; six randomizations of each list were then created. Three groups of listeners were used to assess the W-22 materials. The first group listened monaurally to all 24 lists at a high level before hearing each of the 24 lists in random order from 20 to 70 dB SPL in increments of 10 dB. The second group listened under similar conditions, except that the lists and levels were in different order and the level of 15 dB SPL was added. The third group first listened to all lists at 100 dB before listening to each word order at 50, 40, 30, 20, and 10 dB SPL. The final version of the W-22 test materials was played to 15 listeners in groups of 5 at 80 dB SPL to check the maximum articulation score and at 25 dB SPL to check scores close to threshold.

Study Sample: Six listeners, both experienced and inexperienced, were used for the preliminary test to create Test W-1 materials; each listener had normal pure-tone audiograms. During the second round of testing, six experienced and six inexperienced listeners were used. For Test W-2, 14 listeners who listened to the W-1 disc recordings were used. To test the lists created for Test W-22, three groups of 5 listeners with normal hearing were used.
Results: For Test W-1, the absolute thresholds for inexperienced and experienced listeners were 21 dB and 20 dB SPL, respectively. For Test W-2, the mean absolute threshold for the 14 listeners was 17.7 dB SPL. Analysis of variance showed no significant differences in difficulty between the W-2 discs. Results for Test W-22 indicated no consistent differences between scores on the four different lists.
Conclusions: The results indicated that the difference between thresholds for the W-1 and W-2 tests was attributable to the number of words for each experiment. The authors concluded that the intelligibility of spondee words increases more rapidly with an increase in intensity than does the intelligibility of monosyllabic words. They also surmised that the threshold level for a list of spondees is lower than the threshold for a list of phonetically balanced monosyllabic words.
Relevance to Current Work: This article is an example of early development of recorded and psychometrically equated speech audiometry materials.

Hudgins, C. V., Hawkins, J. E., Karlin, J. E., & Stevens, S. S. (1947). The development of recorded auditory tests for measuring hearing loss for speech. *Laryngoscope, 57*, 57-89.

Objective: The purpose of this study was to explore problems that frequently occur during speech audiometry testing and to produce a speech audiometry test appropriate for measuring all degrees of hearing loss.
Design: The authors delineate specific criteria that must be met for speech audiometric testing, including familiarity, phonetic dissimilarity, normal sampling of English speech sounds, and homeogeneity with respect to basic audibility. The authors list Auditory Test No. 9 and Auditory Test No. 12 as test that were found to have a uniformly high audibility compared to words with differing stress patterns.
Study Sample: Auditory Test No. 9 used 30 participants with typical pure-tone audiograms and three groups of 70 participants with atypical pure-tone audiograms. Auditory Test No. 12 used two groups of typically hearing participants and two groups of participants with atypical hearing. In total, these two tests used 68 typically hearing participants and 49 atypical hearing participants.
Results: Auditory Test No. 9 had a standard error of measurement of 2 dB for typically hearing participants and a range of 2.1 to 2.8 dB for the atypical hearing groups. Auditory Test No. 12 had standard errors of measurement of 2.2 to 2.8 dB for typically hearing participants and a range of 1.4 to 2.3 dB for the atypically hearing groups.
Conclusions: The auditory test word lists were found to be clinically valid for administration to participants with atypical hearing.
Relevance to the Current Work: The guidelines for speech audiometry tesing provided in this article were used throughout the current study.

Huttenlocher, J. W., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27,* 236-248.

Objective: The purpose of this study was to examine a child's vocabulary growth dependent upon exposure to speech.
Design: The researchers estimated the vocabulary size of each child. They then estimated the amount of parent speech through the amount of exposure of speech the child has a day. The next step was to estimate vocabulary growth through the children's spontaneous productions. They compared these steps results in order to gain an understanding of the relation of speech exposure to rate of vocabulary growth.
Study Sample: Two groups of parent-child pairs were tested. Each group had 11 children (6 boys and 5 girls). All children were between 14-16 months old at the time of initial testing.
Results: Researchers found that the amount of mother speech input at 16 months accounts for most of the variation between children's rate of vocabulary growth. Both quantity and quality of the parents' speech were found to be important indicators of the child's vocabulary growth.
Conclusions: Researchers found that gender is an important factor in rate of vocabulary growth as well as parent speech and vocabulary because girls have a greater capacity early on for vocabulary and language. There is still much research to be done to explain the other elements of why children's rate of vocabulary growth are so different from one another's.
Relevance to Current Work: This article provides insight into how much a child's vocabulary can differ.

Individuals With Disabilities Education Improvement Act of 2004, Pub. L. No. 108-446 (2004).

Summary: This act includes sections defining highly qualified teachers, opportunities for professional development, developing individual education plans, early intervention for young children, and eligibility for services.
Relevance to Current Work: This act states that children should be administered testing in their native language or in the language they feel most comfortable with.

Jacobs, A.M. (2012). *Test-retest reliability in the determination of the speech recognition threshold.* (Unpublished master's thesis). Brigham Young University, Provo, UT.

Objective: The purpose of this study was to look at the test-retest reliability of high-frequency usage and psychometrically equated spondaic words.
Design: The study used CID W-1 spondaic words which consists of 36 words. Talkers recorded the list of words and these recordings were assessed by native judges. The participants were presented with the words after familiarization at 50 dB HL. The SRT testing descended in 2 dB decrements.
Study Sample: There were 40 participants between 18 and 35 years old participated. These participants passed hearing screenings, tympanometry, and acoustic reflexes in both ears.

Results: The averaged SRT scores were found for both male and female talker recordings across test and retest values of 0.2 dB HL and -1.2 dB HL. Retest SRTs of 0.4 dB HL and -0.8 dB HL for the female talker. PTA for each participant were compared. For the female talker the SRT was 2.0 dB better than the PTA.
Conclusions: The CID W-1 word list has valid test-retest reliability.
Relevance to Current Work: The current study compared the SRT with the PTA and test retest reliability was explored.

James, C. J., Bowsher, J. M., & Simpson, P. J. (1991). Speech audiometry: Digitization effects and the non-equivalence of isophonemic word lists. *British Journal of Audiology, 25*, 111-121. doi:10.3109/03005369109079841

Objective: The purpose of this study was to determine the range of acceptable digitizing parameters for speech signals being used for speech audiometry testing.
Design: Tape recordings were digitally transcribed. Twelve different conditions of sampling parameters were used, using four-bit resolutions with three bandwidths up to 120bit, 8 kHz bandwidth signal quality. Three presentation levels were used: 0, -10, and -20 dB. The tone was played through a headset at 60 dB SPL. Test words were presented binaurally using headphones. The twelve word lists were randomly assigned to one of twelve conditions. Participants were asked to repeat anything they heard.
Study Sample: Six male and six female subjects were used with no history of hearing impairment with an age range of 21 to 25. Subjects were recruited from the local university. Each subject had pure-tone thresholds between 20 and -10 dBHL. Each subject was familiarized with the range of sounds prior to the experiment.
Results: The three six-bit resolution conditions had marked fluctuations in the standard deviations. Performance measures for bit resolution had statistically significant effects using ANOVA ($P < 0.05$). Post hoc comparison showed that the 12-bit, 10-bit, and 8-bit conditions were homogeneous. Performances for lists 3 and 11 were lower for all presentation levels.
Conclusions: Results show that testing conditions are most adversely affeted when the overall presentation level is reduced. It was found that phonetically balanced lists are not the best for intelligibility testing or hearing tests.
Relevance to Current Work: The present study deals with presenting word lists to children in measurement of the SRT.

Jerger, S., Jerger, J., & Abrams, S. (1983). Speech audiometry in the young child. *Ear & Hearing, 4*, 56-66.

Objective: The purpose of this study was to develop word and sentence lists for use in pediatric speech intelligibility testing.
Design: Sixty picture stimulus cards were developed. Thirty of the cards displayed images for monosyllabic words and the other thirty portrayed actions that could be described in a sentence. The words were selected based on children’s literature for nouns and verbs. Responses to the sentence stimuli cards were recorded. Varying characteristics of childrens’ response patterns were also recorded.

Study Sample: Eighty seven normally developing children between the ages of three and six years were tested.
Results: The childrens' performance on PSI word materials did not differ as a function with their receptive language skills. Relability of the PSI test ranged from 0.85 to 0.95 across normal-hearing and hearing-impaired children.
Conclusions: Performance of the two different tests differed significantly between the children of different ages. Format 1 sentences are appropriate to test children with low receptive language abilities and format 2 sentence are appropriate for testing children with high receptive language skills. The results suggest that this PSI test is vavlid for testing central auditory dysfunction.
Relevance to Current Work: This study tested children within the target age of the current study.

Kail, R. (1991). Developmental changes in speed of processing during childhood and adolescence. *Psychological Bulletin, 109,* 490-501.

Objective: The purpose of this study was to investigate the developmental increases in processing speed in young children and compare these increases with adults using nonverbal stimuli.
Design: A battery of 8 computer-administered tests were used to measure the processing speed in children of different ages and adults.
Study Sample: Participants included 34 children at 4 years, 37 at 5 years, and 38 at 6 years including 43 adults.
Results: The results showed significant age-related increases in processing speed which are not attributed to increased accuracy or error rate monitoring.
Conclusions: The researcher's findings suppor thte global developmental trend hypothesis which suggests that the processing speed extends beyond verbal skills.
Relevance to Current Work: The current study tested children of various ages in verbal tasks where increased processing speed would be an advantage in repeating recognized words.

Keller, L. (2009). *Psychometrically equivalent trisyllabic words for speech reception threshold testing in spanish* (Unpublished master's thesis). Brigham Young University, Provo, UT.

Objective: The purpose of this study was to develop, digitally record, and evaluate trisyllabic Spanish word lists for speech audiometry testing.
Design: Using a frequency usage dictionary, 150 trisyllabic Spanish words were selected for recording. One native Spanish-speaking male and one female talker were chosen to record the words. The words were presented to each participant in 2 dB increments until the participant responded correctly to 100% of the test items or the presentation level reached 16 dB HL.
Study Sample: A total of 20 participants (12 female, 8 male) ranging in age from 18 to 51 years (M = 25.5) were used. Each participant had pure-tone air conduction thresholds of 15 dB HL or less at octave and mid-octave frequencies (125 to 8000 Hz) and static acoustic admittance between 0.3 and 1.4 mmhos with peak pressure between -10 and +50 daPa.

Results: Ninety of the words were selected for the final word list. Thresholds for the male talker ranged from -0.5 to 14.8 dB Hl (*M* = 6.7) and for the female talker from -4.8 dB HL to 16.5 dB HL (*M* = 5.6).
Conclusions: The resulting adjusted words were found to be homogeneous in audibility. The words were found to be viable for clinical testing among Spanish-speaking adults to test their SRT.
Relevance to Current Work: The word lists as well as the methodology were used as a framework for the current study.

Kiselev, S., Espy, K., & Sheffield, T. (2009). Age-related differences in reaction time task performance in young children. *Developmental Cognitive Neuroscience Laboratory - Faculty and Staff Publications, 44,* 150-166.

Objective: The purpose of this study was to measure performance of reaction time in young children and adults to test age-related differences and if they supersede a global mechanism for task demands and processing requirements.
Design: Researches took the mean and standard deviation in RT for each participant. Mixed-factorial, multilevel modeling analyses were performed twice to focus on response time differences between children and adults and another set on the difference among the children of different ages. The study used a regression approach invented by Brinley and the transformation method proposed by Madden to analyze the age-related differences in processing speed. Areas tested were discrimination of color and spatial orientation, reversal of contingencies of previously learned stimulus-response rules, and stimulus-response complexity.
Study Sample: Participants included 54 4-year olds, 54 5-year olds, 59 6-year olds, and 35 adults. All participants are from Russia.
Results: The age-related differences in these response tasks were greater than the researchers had predicted. Color and spatial orientation discrimination resulted in large differences between the children of different ages.
Conclusions: Age-related differences were significant both between children and adults and among the different child age groups. The researchers concluded that slow processing speed among younger chldren are due to age and also task-specific demands that require different levels of cognitive processing while the children are in different developmental stages.
Relevance to Current Work: Response time was indirectly assessed in the present study as a reason for such significant threshold differences between children and adults.

Laroche, C., Vaillancourt, V., Melanson, C., Renault, M.E., Theriault, C., Soli, S.D., & Giguere, C. (2006). Adapting the HINT (hearing in noise test) for canadian francophone children, and preliminary data on the effects of age. *Journal of Speech-Language Pathology and Audiology, 30,* 95-109.

Objective: The purpose of this study was to develop the French Hearing in Noise Test (HINT) for children as an initial look at the effect of age on speech recognition thresholds for sentences.

Design: The PTA and SRT for each participant were measured. The test consisted of 17 phonemically balanced lists of 10 sentences taken from the French HINT. The test words are based on vocabulary words appropriate for 6-year-old children.
Study Sample: Groups of children with mean ages of 6, 7, 8, and 9 years were compared to a group of 'young adults' with a mean age of 11.
Results: A 7 dB difference in SRT was found between the children of 6 years old and the children at 11 year olds.
Conclusions: Children have very different SRTs at varying ages. Age-related differences are significant when it comes to measuring the SRT.
Relevance to Current Work: The current study looked at measuring the SRT with children of varying ages.

Larousse. (2011). *Mi primer diccionario Larousse* [My First Larousse Dictionary] (1^{st} ed.). Barcelona, Spain: Girard, F.

Summary: This is a Spanish dictionary intended for children ages 4 and up. The dictionary contains colorful pictures next to over 1680 spanish vocabulary words. The entries give the part of speech and an example sentence.
Relevance to Current Work: This pediatric Spanish dictionary was used to cross-reference the word list to decide how familiar with the words we could expect the children to be.

Martin, F., & Hart, D. B., (1978). Measurement of speech thresholds of Spanish-speaking children by non-Spanish speaking clinicians. *Speech Hearing Disorders, 43*, 255-262.

Objective: The purpose of this study was to investigate the feasibility of using prerecorded speech threshold procedures using a picture-pointing format in order to administer Spanish test to Spanish-speaking children by a non-Spanish-speaking clinician.
Design: The derived Spanish word list was compared for equivalency with English spondees using a group of bilingual adult judges. The test was then administered to young children. The SRT and PTA of each participant was measured and compared on the word lists.
Study Sample: Sixteen native Spanish-speaking children aged 3 to 6 years were tested. Each of these children had typical pure-tone thresholds.
Results: Both the English and Spanish lists used comprised of highly homogeneous word lists. The PTA and SRT were found to be equivalent for both the Spanish and English lists, proving them clinically valid. Two words were deleted from the list because of their aberrant scores. The differences in threshold between two participants ranged form 0 to 7.5 dB.
Conclusions: The test was found to be feasible, rapid, and reliable. Presenting a test in this manner was found to be clinically valid.
Relevance to Current Work: The present study tested Spanish-speaking children using a Spanish word list.

Martin, F. N., & Clark, J. G. (2009). *Introduction to audiology*. Boston: Allyn and Bacon.

Summary: This is an overview of the purpose of speech audiometry testing including audiological equipment, ideal testing conditions, clinician roles in testing, patient roles in testing, and speech-threshold testing. Speech-detection threshold and speech-recognition threshold definitions and history are given. Instructions are given for obtaining the SRT and recording results including use of masking. Range of comfortable loudness including uncomfortable loudness is given. Phonetically balanced word lists used in word recognition testing are described.
Relevance to Current Work: This source provides relevant information on the history and procedures for testing speech audiometry.

Neuman, A.C., Wroblewski, M., Hajicek, J., & Rubinstein, A. (2009). Combined effects of noise and reverberation on speech recognition performance of normal-hearing children and adults. *Ear & Hearing, 31,* 336-344.

Objective: The purpose of the study was to determine how combinations of noise levels and reverberation in classrooms affect the speech recognition performance of typically developing children to compare with results of adults with normal hearing.
Design: Speech recognition performance was measured using the Bamford-Kowal-Bench Speech in Noise test. A virtual test paradigm represented the signal reaching a student seated at the back of a classroom with varied reverberation. The signal-to-noise ratio required for 50% performance were determined. This was a cross-sectional developmental study.
Study Sample: A total of 63 children with normal hearing and typically developing speech and language were tested. There were 9 children in the following age groups 6-12 years.
Results: The SNR-50 increased significantly with increased reverberation and decreased significantly with increasing age. Children required positive SNRs for 50% performance and thresholds for adults were close to 0 dB. Children 6-8 showed a moderate SNR loss and children 9-12 years showed a mild SNR loss.
Conclusions: The results show the changes in speech recognition performance by age in elementary school. More SNR is required when the environment is reverberant. Younger children require better SNR.
Relevance to Current Work: This study compared adult and child speech recognition results.

Nissen, S. L., Harris, R. W., Jennings, L., Eggett, D. L., & Buck, H. (2005). Psychometrically equivalent Mandarin bisyllabic speech discrimination materials spoken by male and female talkers. *International Journal of Audiology, 44*, 379-390.

Objective: This article developed a set of psychometrically equivalent Mandarin bisyllabic word lists for use in word recognition testing.
Design: A set of 240 regularly used Mandarin bisyllabic words were recorded by native male and female Mandarin speakers. The word set was presented to native Mandarin listeners ranging from -5 to 40 dB HL for evaluation.

Study Sample: Twenty native Mandarin speakers between the ages of 19 and 35 years were used. All participants had typical hearing and reported speaking Mandarin on a daily basis.
Results: Two hundred words with the steepest logistic regression slopes were arranged into four psychometrically equivalent lists of 50 words each and eight half-lists of 25 words each. Each list was digitally adjusted in intensity so that each list's threshold would be equal to the mean threshold of all male and female lists and half-lists.
Conclusions: The authors concluded that the resulting lists and half-lists were psychometrically equivalent with respect to audibility and psychometric function slope. These materials are available on compact disc.
Relevance to Current Work: This study is similar in use of digital recordings by native speakers, selection and evluations of materials, and development of psychometrically equivalent audiologial materials.

Nissen, S. L., Harris, R. W., Channell, R. W., Conklin, B., Kim, M., & Wong, L. (2011). The development of psychometrically equivalent Cantonese speech audiometry materials. *International Journal of Audiology, 50*, 191-201. doi: 10.3109/14992027.2010.542491

Objective: The purpose of this study was to develop and evaluate psychometrically equivalent word recognition and speech recognition testing materals for native speakers of Cantonese.
Design: Commonly used chinese words were selected and recorded by a native Cantonese speaker. Two hundred and fifty Chinese bisyllabic words were divided into 10 lists of 25 words each as stimuli for the WR materials, and 90 trisyllabic words were selected for the SRT materials. The WR lists were presented at different intensity levels ranging from -5 to 40 dB HL in 5dB increments. SRT testing used randomized stimuli delivered to the listeners beginning at 6 dB below their PTA, decreasing in increments of 2 dB until no correct responses were given. The test continued by raising the presentation level in increments of 2 dB until the participant correctly repeated all stimulus items. Native Cantonese judges rated each response as correct or incorrect. Based on listener responses, the WR lsits were rearranged into four lsits and eight half lists relatively homogeneous in audibility.
Study Sample: Twenty Cantonese speakers with typical hearing between the ages of 19 and 29 years were used as participants.
Results: The mean psychometric function slopes for WR materials were 7.5%/dB for the male talker and 7.6%/dB for the female talker. For SRT materials, the mean psychometric slopes were 14.5%/dB for the male talkers and 14.9%/dB for the female talkers.
Conclusions: The Cantonese speech audiometry materals are available on compact disc.
Relevance to the Current Work: The design of this study was similar to the current work and was used as a reference.

Olsen, W. O. & Matkin, N. D. (1991). Speech audiometry. In Rintelmann (Ed.), *Hearing Assessment* (pp. 39-140). Austin, TX: Pro-Ed.

Summary: This audiological desk reference book covers definitions of PTA, SRT, and WRS. The authors explore the acceptable differences between a patient's PTA and SRT. The definition for speech audiometry is given as well as its clinical background and history with pure-tone audiometry. The authors give examples for how to calculate a patient's PTA. The authors present the proper procedures when testing for the SRT including the 2 dB and 5 dB steps. Olsen and Matkin also briefly touch on tympanometry and acoustic reflex.
Relevance to Current Work: The current study explored spech audiometry and differences between a patient's PTA and SRT.

Parragon. (2007). *Mi primer diccionario* [My First Dictionary] (1st ed.). Beijing, China: Root, B.

Summary: This is a Spanish dictionary intended for children ages 4 and up. The dictionary contains colorful pictures next to over 300 of the most commonly used child spanish vocabulary words. The entries give the part of speech and an example sentence. The dictionary is intended to help children learn how to read.
Relevance to Current Work: This pediatric Spanish dictionary was used to cross-reference the word list to decide how familiar with the words we could expect the children to be.

Penny, R. J. (2000). *Variation and change in Spanish*. New York, NY: Cambridge University Press.

Summary: The author presents several aspects of the Spanish language including language variations, dialectal definitions, mechanisms of change, variations in Spain, variations in Spanish America, variations in Judeo-Spanish, and standardization of the language. The definitions of idiolects and dialects are given.
Relevance to Current Work: This chapter informed research of the Spanish language for better word selection and understanding of the language.

Preece, J. P. & Fowler, C. G. (1992). Relationship of pure-tone averages to speech reception threshold for male and female speakers. *Journal of American Academy of Audiology, 3,* 221-224.

Objective: The purpose of this study was to compare the PTA and SRT results for male and female speaker recordings.
Design: The study used recorded male talker and female talker word lists on compact disc.
Study Sample: Two similar groups of clinic patients were selected to participate. Each group included 215 ears with normal hearing. The mean age for group 1 was 46.8 years and the mean age for group 2 was 45.4 years.
Results: No significant differences were found between the SRTs for the two groups suggesting the two speaker recordings are equivalent for clinical purposes. The SRTs for both speaker recordings were within 6 dB for both the two- and three-frequency averages.
Conclusions: Both female and male talker recordings are clinically equivalent.

Relevance to Current Work: The present study compared the PTA and SRTs of male and female speakers.

Ramkissoon, I. (2001). Speech recognition thresholds for multilingual populations. *Communication Disorders Quarterly, 22*, 158-162.

Summary: This article focuses on the development of speech audiometry testing within the United States. The author also reports the current need for measuring the speech recognition threshold among multi-lingual populations. An overview of the development of audiological testing and speech audiometry procedures is given to highlight sociolinguistic considerations for testing individuals with cultural and linguistic diversity. The author demonstrated how altering test materials might invalidate SRT results. Digit stimuli are presented as appropriate alternatives to spondee words because of their steep rise in intelligibility.
Relevance to the Current Work: This article provides evidence that supports the need for materials developed for non-native English speakers.

Ramkissoon, I., Proctor, A., Lansing, C. R., & Bilger, R. C. (2002). Digit speech recognition thresholds (SRT) for non-native speakers of English. *American Journal of Audiology, 11*, 23-28.

Objective: The purpose of this study was to create SRT test materials using digit pairs for use with non-native English speakers.
Design: Test stimuli were created using 56 digit pairs and 36 test items from the CID W-1. Participants received pure-tone audiometry hearing screenings to obtain a PTA to later compare to the SRT. Both the digit pairs and test items were disclosed to participants prior to testing through playing them at a comfortable loudness level. Testing began at 20 dB above the participant's pure-tone threshold at 1000 Hz. Participants received instruction to listen and repeat the stimuli they head. The presentation level was decreased by 2 dB at a correct response and increased by 2 dB at an incorrect response. Each participant's threshold level for speech was estimated.
Study Sample: Twelve non-native English spakers and twelve native English speakers participated. The participants ages ranged from 22 to 69 years old. Each participant had typical hearing. The non-native English speakers were native speakers of Spanish, Chinese, French, Farsi, and Russian. Each non-native English speaker reported living in the United States for less than one year.
Results: The authors found a strong correlation between the D-SRT and CID mean thresholds between digit stimuli and listeners' pure-tone averages. The average threshold was found to be higher in non-native English speakers than in native English speakers.
Conclusions: These results indicate that the strong correlation between a listener's PTA and their CID-SRT is unique to native speakers of English. Data pairs were found to be an acceptable alternative to measure the hearing threshold for speech, particularly for those listeners with limited English proficiency.
Relevance to the Current Work: This article provides an alternative method for testing speech recognition of non-native English speakers.

Ramkissoon, I., Estis, J. M., & Flagge, A. G. (2014). Digit speech recognition threshold (SRT) in children with normal hearing ages 5-8 years. *American Journal of Audiology, 23,* 182-189. doi:10.1044/2014_AJA-13-0039

Objective: The purpose of this study was to test children using digit stimuli and spondees in measurement of the SRT in order to determine the influence of changes in increment size using a descending method.
Design: For each participant, the SRT was measured during testing with paired digit stimuli and pediatric word stimuli. The children were tested using 5 dB and 2 dB descending step sizes. Pure-tone testing was administered prior to SRT testing to ensure normal hearing. The PTA thresholds for the digit stimuli and word stimuli were compared for each participant.
Study Sample: Thirty typically developing children with normal pure-tone thresholds between the ages 5;0 to 8;11 were tested.
Results: Analyses of variance were performed and revealed a significant difference in PTA between 2 dB and 5 dB steps. PTA was significantly lower when using 2 dB steps. The step size when testing digit and word stimuli made no significant differenct in the resulting SRT. The digit SRT and word SRT were highly correlated (r = .49-.72). Regression analyses showed comparable PTA and SRT for digit and word stimuli.
Conclusions: Digit pairs are an appropriate alternative for SRT measurement in children with typical development and normal hearing.
Relevance to Current Work: The current study also measured the SRT in children using word stimuli to compare the PTA and SRT.

Rudmin, F. (1987). Speech reception thresholds for digits. *Journal of Auditory Research, 27*, 15-21.

Objective: The purpose of this study was to develop lists of digits for use for speech audiometry testing for patients who are not native English speakers.
Design: Digits 1, 2, 4, 5 and 9 were selected due to their homogeneity and longer vowel durations. Patients were required to give two correct responses at each testing level because of the likelihood of correct guesses when using only five stimuli. The patient's threshold was first found then reduced in 10 dB steps for each presentation with 2 correct responses.
Study Sample: There were 130 patients used in this study. They were referred from ENTs. Data from seven of the patients had to be thrown out; of the remaining 123 (74 male, 49 female) patients, 43 were English, 40 French, and 40 Greek, Chinese, German, Yiddish, Arabic, Italian, and Polish. Ages ranged from 10-70 years.
Results: The digit SRTs were found to be equivalent to each patient's 3-frequency PTA.
Conclusions: The results show that using English digits is a reliable way to clinically test for a patient's SRT regardless of the patient's ethnolingual background.
Relevance to Current Work: This study used patients whose dominant language was not English and tested their SRT. This study also tested young children under the age of 10.

Schneider, B. S. (1992). Effect of dialect on the determination of speech-reception thresholds in Spanish-speaking children. *Language, Speech, and Hearing Services in Schools, 23*, 159-162.

Objective: The purpose of this study was to compare stimuli in Castillian, Caribbean, and Mexican dialects when testing the SRT of Spanish-speaking children.
Design: A list of 12 Spanish words used in previous studies by Martin and Hart (1978) were selected due to their phonetically dissimilar vowel sequence patterns. A separate tape recording was prepared in each of the three Spanish dialects. Each word had a carrier phrase. Each recording was randomly played to each subject, and the subject was told to point to the picture matching the word they had heard. The subjects were briefly introduced to the test words prior to testing. Words were presented 30 to 40 dB above the estimated PTA and decreased in 10 dB increments, presenting one word at each level.
Study Sample: Twelve children (6 boys, 6 girls) between the ages of 6 and 7 years born in Puerto Rico were tested in this study. Each child was judged Spanish-dominant through the Hartford Board of Education Home Language Survey. Each subject had typical pure-tone thresholds in the test ear of 15 dB HL or less for at least three frequencies.
Results: The differences between the three dialects were statistically significant ($p < 0.01$). Intrasubject differences ranged from 0 to 6 dB. Four of the subjects showed exact agreement aross all three dialects. The remaining eight subjects showed differences of 2 dB or less among the three dialects. The SRT found using all three dialects were highly correlated with the PTA.
Conclusions: The results indicated a reliable test because of the thresholds correlating with the subjects' PTA. This test indicates that Spanish dialect will not make a significant difference in the patient's SRT results.
Relevance to Current Work: This study tested the PTA and SRT of young Spanish-speaking children.

Spitzer, J. B. (1980). The development of a picture speech reception threshold test in Spanish for use with urban U.S. residents of Hispanic background. *Journal of Communication Disorders, 13*, 147-151. doi:10.1016/0021-9924(80)90031-3

Objective: The purpose of this study was to develop SRT testing materials for use by non-Spanish speaking audiologists for testing Spanish-speaking patients.
Design: The authors selected familiar nouns that were not known to have a usage idiosyncratic to a particular region. The resulting list consists of pictures of people, body parts, clothing, food, animals, and common objects. Most of the words follow a trochaic stress pattern. A carrier phrase is played prior to each word, which indicates to the subject to point to the named object. Subjects were familiarized to the words prior to testing at a high SL. Words were played beginning at 40 dB HL and descended in 10 dB steps.
Study Sample: The participants used are not oulined in this study, although from context it can be assumed they were Spanish-speaking children.
Results: The testing showed good correspondence between the SRT and PTA of each subject within 10 dB of each other.

Conclusions: This list of words is a feasible approach to measuring the ST in subjects from Hispanci backgrounds.
Relevance to Current Work: This study tested Spanish speaking children.

Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., & Gwaltney, C. A. (1999). Monosyllabic word recognition at higher-than-normal speech and noise levels. *Journal of the Acoustical Society of America, 105*, 2431-2444.

Objective: The purpose of this study was to measure the speech intelligibility factored with increases in noise level during a word recognition test of monosyllabic words.
Design: The word lists used were taken from the Auditec of St. Louis recordings of the NU#6 word test spoken by a male talker. A noise masker was used. A total of eight speech RMS levels (64, 69, 74, 79, 84, 89, 94, 99 dB SPL) were evaluatd along with ten S/N ratios (-4, -1, 2, 5, 8, 12, 16, 20, 24, 28 dB). Each participant experienced all 16 conditions in one session.
Study Sample: Three groups of participants were tested. The first group consisted of 72 normal-hearing young adults (23 males, 49 females). The second group consisted of 32 hearing-impaired adults (22 males, 10 females) under the age of 70 years. The third group consisted of 12 hearing-impaired adults (7 males, 5 females) over the age of 70 years. Participants in group one had pure-tone thresholds of 15 dB HL or better from 250 to 8000 Hz. Participants in groups two and three had pure-tone thresholds of 20 dB HL or worse.
Results: Group 1: Performance declined as S/N ratio increased. Reductions were most large when the S/N ratio was below 16 dB. Group 2 & 3: Results of these two groups did not differ significantly.
Conclusions: The results concluded that speech intelligibility in noise decreases when speech levels exceed 69 dB SPL and the S/N ratio remains constant. The negative effects of added noise level are greater when the speech level is high and vice versa. Normal-hearing and hearin-impaired subjects are affected the same by increased speech and noice levels. The effective dynamic range of speech may be larger than the previously assumed 30 dB.
Relevance to Current Work: This work concluded that recording onto CDs adds reliability to test scores, rather than performing tests through live-voice presentation.

Thibodeau, L. M. (2007). Speech audiometry. In R. J. Roeser, M. Valente & H. Hosford-Dunn (Eds.), *Audiology diagnosis* (2nd ed., pp. 288-313). New York, NY: Thieme.

Summary: The author covers terminology of speech audiometry, including differences between detection threshold, recognition threshold, and reception threshold. The rationale for testing speech audiometry is given, including the effects of reduced absolute and differential sensitivity and its contributions to differential diagnosis. The inherent applications to hearing aid fittings and aural rehabilitation are noted. Thibodeau then describes the proper administration of a speech audiometry test. She describes the response format, speech stimuli, presenation mode, presentation level, calibration for speech stimuli, procudres, and analysis of results. The author defines speech awareness, discrimination, identification, and loudness perception. Speech audiometry can be used

in differential diagnosis for conductive, sensorineural, and retrocochlear hearing loss. The author finishes by describing adjustments to be made in special age groups, including young children and the elderly.
Relevance to Current Work: This chapter illustrates the importance and proper use of speech audiometry testing and its application in use with young children.

U.S. Census Bureau. (2011). *The Hispanic population: 2010.* Retrieved from: http://www.census.gov/prod/cen2010/briefs/c2010br-04.pdf

Summary: This census brief introduces the hispanic population summary for the year 2010. The information includes understanding the Hispanic origin data from the census and how to read the graphs. Tables and graphs are included to show the increases in the population growth over time. The census information details which ethnicities and dialects of the Spanish language are primarily immigrating to the Untied States. The population growth is further broken down into states.
Relevance to Current Work: The census information directly impacted the importance of the current study.

Wilson, R. H., & Strouse, A. (1999). Psychometrically equivalent spondaic words spoken by a female speaker. *Journal of Speech, Language, and Hearing Research, 42*, 1336-1346.

Objective: The purpose of this study was to psychometrically equate spondaic word thresholds from the CID W-1 test to reduce threshold variability of the individual spondaic words.
Design: Two experiments were conducted in this study. In the first experiment, listener responses established psychometric functions for the 36 spondaic words spoken by both a male and female talker. In the second experiment, the talker recordings were then digitally adjusted to produce equal thresholds. The psychometric functions for the 36 adjusted spondaic words were then established.
Study Sample: Two groups of 20 listeners with normal hearing were used in this study. Listeners used range in age from 19 to 30 years.
Results: The mean thresholds for both experiments were equivalent; standard deviations for word thresholds in the second experiment were significantly smaller than those of the first experiment.
Conclusions: The standard deviations obtained showed that the resulting materials are psychometrically equivalent. The spondaic words are available on compact disc.
Relevance to the Current Work: This study similarly included the process of creating homogeneous speech audiometry materials.

Appendix B

Informed Consent

RESEARCH CONSENT FORM

Introduction

This research study is being conducted by Richard Harris, PhD at Brigham Young University; Brenda Peterson, BS Communication Disorders, Communication Disorders graduate student at BYU; and Jessica Graham, BS Communication Disorders, Communication Disorders graduate student at BYU to evaluate a word list recorded using improved digital techniques. You were invited to participate because you have a child who speaks Spanish and is between the ages of four and eight years.

Procedures

If you agree to participate in this research study, the following will occur:

- You will receive a hearing screening where you will hear beeps and indicate whether or not you heard them.
- You will receive an articulation screening where you will name pictures of objects.
- You will listen to Spanish words and repeat the words you hear.
- The total time commitment will be approximately 60 minutes divided into two 30-minute sessions.
- This will take in the hearing lab, room 110 of the Taylor Building on BYU Campus.

Risks/Discomforts

There are no known risks associated with this study. The researchers will be present at all times to make sure that your child is not experiencing any problems in the study conditions. If the child indicates in any way that he/she does not want to participate by crying or another behavior, we will stop immediately.

Benefits

The primary benefit to your child is finding out whether he/she has normal hearing or not throughout the course of the study. There may be benefits to society in general in that this study may result in more effective treatment methods for Spanish-speaking children participating in hearing exams.

Confidentiality

Your child's participation will be confidential. The data will be stored in locked file cabinets within locked labs within the BYU Speech & Language Clinic. Only the researchers will have access to the data. The names will be removed from research materials. Neither your or your child's name will ever be used in association with this research.

Information will be kept for three years after the study is completed. The files will remain in a locked cabinet only accessible by the researcher. Internet data will be saved as a Microsoft Excel document with no subject identifiers. Participants will be listed by number with no names or identifying referents.

Compensation

You will receive $20 for your participation; compensation will not be prorated. Your child will receive a free hearing exam and you will receive the results. Your child will also receive a sticker or other small prize after each session.

Participation

Participation in this research study is voluntary. You are free to decline to have your child participate in this research study. You may withdraw your child's participation at any point without losing the compensation.

Questions about the Research

Please direct any further questions about the study to Richard Harris at (801) 422-6460 Richard_harris@byu.edu. You may also contact Brenda Peterson or Jessica Graham at (385) 414-5022 or (949) 355-3578, bren.krog@gmail.com or jessicaleerandle@gmail.com.

Questions about Your Rights as Research Participants

If you have questions regarding your rights as a research participant contact IRB Administrator at (801) 422-1461; A-285 ASB, Brigham Young University, Provo, UT 84602; irb@byu.edu.

Statement of Consent

I have read, understood, and received a copy of the above consent and desire of my own free will to participate in this study.

Parent's Name (Printed): ____________________ Signature: ____________________ Date: ______________

Appendix C

Spanish Trisyllabic Word Dictionary Definitions

Word	Definition [b]	Part of Speech	Dictionary [a]
comprender	to understand	verb	1
derecho	right; straight	adjective	1, 2
ejemplo	example	noun	1, 2
figura	figure; form	noun	1,2
levantar	to lift; pick up	verb	1, 2
minuto	minute	noun	1, 2
momento	moment	noun	1
pequeño	small; young	adjective	1, 2, 3
perfecto	perfect; ideal	adjective	1, 2, 3
persona	person	noun	1, 3
pregunta	question	noun	1, 3
primero	first	noun	1
recoger	to pick up; to gather	verb	2
recordar	to remember; recall	verb	1, 2
tamaño	size	noun	1, 2

[a] 1=See Larousse, 2011; 2=See Everest, 2013; 3=See Root, 2007. [b] See Oxford, 2015.