2020

COMPARISON OF LANGUAGE PERFORMANCE USING EXPOSURE AND NON-EXPOSURE NORMS IN ENGLISH LEARNERS IDENTIFIED WITH SPEECH LANGUAGE IMPAIRMENT

ANNA TELLO

COMPARISON OF LANGUAGE PERFORMANCE USING EXPOSURE AND NON-EXPOSURE NORMS IN ENGLISH LEARNERS IDENTIFIED WITH SPEECH LANGUAGE IMPAIRMENT

A dissertation submitted in partial fulfillment

of the requirements for the degree of

DOCTOR OF PSYCHOLOGY

to the faculty of the

DEPARTMENT OF PSYCHOLOGY

of

ST. JOHN'S COLLEGE OF LIBERAL ARTS AND SCIENCES

at

ST. JOHN'S UNIVERSITY

New York

by

Anna C. Tello

Date Submitted: March 7, 2020 Date Approved: April 21, 2020

_____	_____
Anna C. Tello	Samuel O. Ortiz, Ph.D.

ABSTRACT

COMPARISON OF LANGUAGE PERFORMANCE USING EXPOSURE AND NON-EXPOSURE NORMS IN ENGLISH LEARNERS IDENTIFIED WITH SPEECH LANGUAGE IMPAIRMENT

Anna C. Tello

English Learners (ELs) in the U.S. have historically been misidentified as having a Speech Language Impairment (SLI). School psychologists and speech language pathologists must determine whether ELs present with language differences that are due to normal patterns of language acquisition or language impairment. Not only is the process complicated by the presence of two or more languages rather than just one, it often involves the use of standardized tests that do not control adequately for the differences in developmental language proficiency. Based on the results of this study, it was found that using tests that are developed with language exposure norms (e.g., the Ortiz PVAT) seem to produce scores that are substantially higher when used with ELs as compared to the scores that are derived from traditional tests using non-exposure norms. This study focused on an analysis and comparison between traditional measures of verbal ability (e.g., WISC-V Verbal Comprehension Index (VCI)), other language assessments (e.g., CELF-5 and PLS-5) and the Ortiz PVAT scores. It was found that both verbal ability VCI and language scores were consistently over one standard deviation lower than the Ortiz PVAT scores. Furthermore, when ELs were tested with the Ortiz PVAT their mean values were consistently within the average range. This suggests that current methods may be misidentifying English Learners as having a Speech Language

Impairment where none exists, which negatively affects eligibility and intervention decisions and, ultimately, their academic trajectory. Overall, the results indicate that the Ortiz PVAT provides an alternative to current evaluation instruments because it targets receptive vocabulary while controlling for language exposure in ELs. This alternative method allows practitioners to use the Ortiz PVAT to assess English Learners in a manner that assists in determining if their performance is reflective of typical language acquisition or if it is suggestive of a language disorder.

ACKNOWLEDGEMENTS

First, I would like to thank God for giving me the patience and fortitude to continue regardless of the countless obstacles placed before me. Next, I would like to take this opportunity to thank some of the most important people in my life who have helped me achieve this life-long dream. To my family for their constant encouragement and words of wisdom. To my parents who sacrificed so much to make my dreams become a reality. Mom, you are a true inspiration, you are the kindest and strongest person I know. I am so grateful for your unconditional "amor" and relentless confidence. To my sister, Vanessa, your dedication and perseverance as a mom gave me the inspiration to work on this project. To my brother John, thank you for reassuring me that I can always count on you, "no matter what"! To my husband, Eric, thank you for listening, crying, and especially laughing with me. You brighten up my day with your endless hugs and reminders that "everything will be ok". To my remarkable daughter Mia and baby boy, Eric, please remember to reach for the stars and that mamá "loves you very much!" To my dad in heaven, who pushed me to strive for greatness and now I am the first person in our family to receive a doctoral degree. I would like to thank the rest of my family and friends for all their prayers, support, friendship and love.

I would also like to express my gratitude to my mentor, Dr. Samuel Ortiz, whose constant guidance and encouragement has helped me achieve this incredible journey. You are a true educator, innovator and pioneer in school psychology. To Dr. Marlene Sotelo-Dynega, Dr. Jennifer Mascolo, and Dr. Robyn Kurasaki your thoughtful feedback and words of support have been invaluable. Finally, I would like to thank all the children who participated in this study and hope that this project improves the way school

psychologists and speech language pathologists work with linguistically and culturally diverse students.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

Although considerable research has been conducted among English Learners (ELs) and Speech Language Impairment (SLI), the body of literature reveals mixed and sometimes conflicting results. A clear understanding of how to best provide assessment and intervention for ELs with suspected or confirmed SLI is predicated on understanding typical development of an EL's profile. In terms of this study, ELs will be defined as any child or adolescent that has had any significant exposure to another language other than English. ELs are culturally and linguistically diverse and represent an ever-increasing percentage of the U.S. student population, comprising the fastest growing subgroup by 2023 (Genesee, Lindholm-Leary, Saunders, Si Christian, 2005). An examination of school characteristics and educational outcomes reveals pervasive disparities in resources and overrepresentation of ELs students in special education relative to their mainstream monolingual English-speaking peers (Brayboy, Castagno & Maughan, 2007). According to research, referrals of ELs into special education are due to multiple reasons including poorly designed language assessments and weak psychoeducational practices (Figueroa & Newsome, 2006). Results are particularly dire for ELs placed into special education since the implications range from high grade retention and highest dropout rates of any school-aged youth (Rueda & Windmueller, 2006).

Additionally, the lack of effective instruction negatively influences assessment results, which are further confounded by the fact that tests that are designed for native English speakers lack psychometric properties when used to identify ELs with SLI (Abedi, 2006). Furthermore, many school practitioners have difficulty making the distinction between emergent English proficiency versus disability (Keller-Allen, 2006)

and it is common for normal language acquisition to be confused with learning problems (Artiles & Klingner, 2006). Factors contributing to this confusion include a limited array of appropriate instruments, lack of professional training in linguistic and cultural differences, and a shortage of bilingual educators and evaluators (Figueroa & Newsome, 2006). Since it is not uncommon for ELs to perform poorly on psychoeducational and speech measures that have high language demands, this population is even more vulnerable to misidentification and having a disability (Abedi, 2006).

A factor that contributes to the misidentification of ELs as having SLI stems from the problems in distinguishing a language difference vs. a language disorder. Paradis (2005) found that children learning English as a Second Language (ESL) may show similar characteristics to children with Specific Language Impairments (SLI) when assessed by language tests that are not valid, reliable, and free of bias. Thus, typically developing students learning English as a Second Language may be diagnosed as having a language disorder when, in reality, they are showing signs of typical second language acquisition. According to the American Speech-Language- Hearing Association (ASHA), clinicians working with diverse and bilingual backgrounds must be familiar with how elements of language differences and second language acquisition differ from a true disorder (ASHA, 2004). Not only is the process of assessing ELs complicated by the presence of two languages rather than just one but also it often involves the use of standardized tests that do not adequately control for differences in developmental proficiency in the language being measured. Typical approaches, such as testing the limits, use of translators, and reliance on native language tests, have not proven to be successful in generating valid scores (Ortiz, 2014). Coupled with inequitable

interpretation, it is not surprising that ELs are disproportionately represented in the disability category of SLI.

Recent advancements with access to language-normed samples suggest that it may be possible to control for these developmental language differences (Ortiz, 2018). In this approach, amount of lifetime exposure to English is used as a formal stratification variable within the context of testing in English, as a common metric for evaluating the process of normal vocabulary acquisition in the second language. By providing a true peer comparison group that addresses differences in experience and language development, it is possible to generate scores that hold the necessary degree to support the validity of subsequent identification of SLI in ELs. The purpose of this investigation was to determine if tests without exposure norms for identifying SLI in ELs produce scores that are consistent with scores measuring performance on tests with exposure norms. In this way, the concept of controlling for amount of exposure to English and its impact on test results may be examined in light of its utility in guiding diagnostic assessment of ELs who are suspected of having SLI.

Literature Review

Current immigration trends in the United States have increased the number of ELs student population, which represents a rapidly growing community that is culturally and linguistically diverse. The percentage of public-school students in the United States who were ELs was higher in fall 2016 (9.6 percent, or 4.9 million students) than in fall 2000 (8.1 percent, or 3.8 million students) (National Center for Education Statistics, 2016). This number reflects a 45% increase in the ELs population since 2000–2010 (U.S. Department of Education, 2017). In New York City there are about 154,276 ELs enrolled during the 2016-2017 school year. Of the 154,276 ELs who enrolled in NYC city schools the largest population of ELs attended schools in Queens with 46,898 students (30.40%), followed by Brooklyn 44,498 (28.84%) and Bronx with 40,506 ELLs (26.26%) (U.S. Department of Education, 2017).

In 2015, there were about 3.8 million Hispanic-Latino EL students, which constituted over three-quarters (77.7 percent) of ELs student enrollment overall. Asian students were the next largest racial/ethnic group among ELs, with 512,000 students (10.5 percent of EL students). In addition, there were 295,000 White EL students (6.1 percent of EL students) and 178,000 Black EL students (3.7 percent of EL students). In each of the other racial/ethnic groups for which data were collected (Pacific Islanders, American Indians/Alaska Natives, and individuals of two or more races), fewer than 40,000 students were identified as ELs (U.S. Department of Education, 2017).

Although special language programs are provided for more than 400 language groups, approximately 80% of ELs are Spanish speakers (National Center for Education Statistics 2016). This percentage has increased from previous years and is expected to

grow at a faster rate than the general student population (National Center for Education Statistics 2016). One projected estimate by the U.S. Census Bureau (2015) predicts that language minority children will become the majority in U.S. schools by 2023. Hence, this population of children presents challenges for practitioners such as school psychologists and speech-language pathologists who provide language assessments and interventions.

As the EL population in U.S. schools continues to grow, practitioners are encountering an increased need to conduct evaluations designed to assess potential language disabilities in students who are not native English speakers. Developments toward addressing these problems were addressed by The National Association of School Psychologists (NASP) which further aligned its methodology of how to work with ELs with the Individuals with Disabilities Education Improvement Act (IDEA, 2004). One of the goals of IDEA is to reduce the misidentification of ELs by emphasizing the use of culturally and linguistically appropriate assessments and evaluations (IDEA, 2004). However, according to NASP, there continues to be an overrepresentation of ELs in SLI due to the unclear assessment and testing standards that are not psychometrically, culturally, and linguistically appropriate for evaluating ELs (NASP, 2015).

Although many educators, school psychologists, speech pathologists and even parents believe that ELs should be placed into special education, the research demonstrates the negative effects of this practice. One of the strongest criticisms for this practice is that placement in special education affects ELs at a disproportionate rate and impacts their educational and life trajectories. These students are often relegated to less rigorous learning environments that stifle both their educational and social development. Once a child is labeled as having a disability and is placed within special education, their

chances of reaching their full potential becomes limited (Keller-Allen, 2006). Research has demonstrated that placing ELs in a special education setting has startling statistics relating to increased drop-out rates, adult unemployment, increased mental health problems, lower adult SES and even increased risk for incarceration (Hanushek & Rivkin, 2002). Furthermore, there is supporting research that a large percentage of the ELs who are already placed into special education can be mainstreamed if accurately evaluated and provided with the appropriate educational services.

Although it is still unclear how best to evaluate ELs students, there are several methods that are currently being used to evaluate children for SLI. While there are multiple scales, surveys, parent questionnaires and assessments, this review will focus on the top five most popular approaches that practitioners use to evaluate SLI among ELs.

The first measure is the Peabody Picture Vocabulary Test, Fourth Edition (PPVT–4) scale. The PPVT-4 scale is an English only norm-referenced, wide-range instrument for measuring the verbal ability in standard American English, receptive vocabulary of children and adults ages 2 to 90+ years old (Hoffman, Templin & Rice, 2012). It measures listening and understanding of single-word vocabulary and receptive processing. If administered to a school-age child, the PPVT-4 can estimate the child's scholastic aptitude. The authors claim that the test can assess for high and low verbal abilities in addition to identifying possible learning disabilities and language disorders in children (Hoffman, Templin & Rice, 2012).

However, the PPVT-4 has several psychometric problems that begin with the authors' assessment of its construct, content validity, and discriminant accuracy. Construct validity is the degree to which the test measures what it aims to measure and if

it assesses the extent to which a test can be used for a specific purpose, such as to identify children with a language disorder (Hoffman, Templin & Rice, 2012). However, construct and content validity were found to be insufficient because the test only assesses receptive vocabulary; thus, conclusions about language proficiency cannot be made with confidence (Dunn & Dunn, 2007). Authors of the PPVT-4 did not report discriminant accuracy. Similarly, diagnostic accuracy was judged to be poor due to the inclusion of individuals with disorders in the standardization sample, which reduces test sensitivity (Peña, Spaulding, & Plante, 2006). Unfortunately, the PPVT-4 tests the experiences of the examinee rather than their inherent abilities to acquire new vocabulary and language (Campbell, et.al., 1997).

The second measure widely used is the Expressive Vocabulary Test, Second Edition (EVT-2). Unlike some other measures of vocabulary, the EVT-2 supplies two equivalent forms of the test which contain different vocabulary items—helping ensure the individual has not "learned" the test (Williams, 2007). EVT-2 also includes a unique Growth Scale Value (GSV) which is sensitive to small changes in vocabulary acquisition over time. Each form contains sample items and 190 test items arranged in increasing difficulty (Williams, 2007). Psychometrically, it was developed over a five-year period and was co-normed along with the PPVT-4 test with a national sample of individuals ranging in age from 2 to 90 years old (Hayward, et al., 2008). More than 3,500 subjects were used for the normative scores and the sample was tightly controlled to matched to the U.S. Census on gender, race/ ethnicity, region, and clinical diagnosis or special education placement (Williams, 2007). Regarding word retrieval, the comparisons with

the PPVT-4 show high correlations, 60% shared variance, and provide evidence that word retrieval is the construct being measured (Hayward, et al., 2008).

However, the EVT-2 test was standardized on monolingual English-only norms. The author states that caution should be used when assessing individuals for whom English is a second language and that, in such situations, results should be used as a baseline measure of vocabulary only (Williams, 2007). As per the author's own guidelines, this measure does not have the psychometric norms to be used with ELs and hence it is unreliable and invalid in regard to the assessment of ELs (Williams, 2007).

The third measure commonly used is the Preschool Language Scales Fifth Edition (PLS-5). It is most commonly used to assess children for language difficulties in early intervention programs and preschools. It was designed for use with children aged birth through 7:11 to assess language development and identify children who have a language delay or disorder. The test also aims to identify receptive and expressive language skills in the areas of attention, gesture, play, vocal development, social communication, vocabulary (Dollaghan & Horner, 2011). There were 1,400 children who participated in the standardization normative sample, collected in more than 45 states in the United States. The standardization sample matches U.S. Census figures for region, race/ ethnicity, and level of caregiver education.

Nevertheless, results from the PLS-5 are likely not to be valid due to an insufficient reference standard and insufficient discriminant accuracy to properly identify children with a language delay or disorder (Dollaghan & Horner, 2011). The PLS-5 manual states that it is a valid test with sensitivity values; however, it is critical to note that these numbers are misleading as the clinical populations chosen for the sensitivity

measure were chosen through an unspecified measure. Specificity measures reported as fair were invalid since no reference standard was applied to the non-clinical sample for the specificity measure (Dollaghan & Horner, 2011). Only those individuals who were clearly language disordered/ delayed were included in the sensitivity sample and only those who were not receiving speech and language services were included in the specificity sample. Mildly delayed children and children who were borderline typically developing were excluded from diagnostic accuracy measures reported (Dollaghan & Horner, 2011). Diagnostic accuracy measures are considered irrelevant as they were not based on a population representative of a real-world clinical population. In addition to unacceptable measures of validity, accuracy and reliability in determining the presence of language disorder or delay, the PLS-5 contains significant cultural and linguistic biases which preclude it from being appropriate for children from diverse backgrounds. This is noted in federal legislation which requires testing materials to be valid, reliable and free of significant bias (IDEA, 2004). As many parts of this test are largely vocabulary based, it will likely falsely identify children as language delayed or disordered from non-mainstream cultural and linguistic backgrounds or who come from lower socioeconomic status (Harry & Klinker, 2006; National Research Council, 2002). In conclusion, this measure has multiple psychometric problems and it has not been normed on ELs, making it unreliable and invalid for use with this population.

Another measure often used is the Clinical Evaluation of Language Fundamentals Fifth Edition (CELF-5), which was designed to assess a student's language and communication skills in a variety of contexts. Although the CELF-5 is a lengthy test to administer, it is one of the most well-known and widely used assessment tools for

children with SLI because it identifies a student's language strengths and weaknesses (Dollaghan & Horner, 2011). It is commonly used in schools to determine eligibility for special education services or accommodations and provide performance-based assessment that corresponds to educational objectives.

The CELF-5 is appropriate for students ages 5:0 through 21:11, and psychometrically, the standardization sample was based on the March 2010 US Census and was stratified by age, sex, race/ethnicity, geographic region, and parent education level (Dollaghan, 2007). Of the 3,000 participants, 27% spoke a dialect other than Standard American English (SAE), 4% were classified as gifted or talented, and 11% had some clinical diagnosis (Dollaghan, 2007). The authors of the CELF-5 measured content and construct validity using a study of students diagnosed with and without language disorders. However, several factors contribute to lack of content validity. First, there is a lack of information regarding how individuals who participated in normed studies were identified as typically developing or language impaired, making it invalid (Betz, Eickhoff, & Sullivan, 2013).The pilot model used sample sizes smaller than what is considered acceptable in the field. Based on the information provided in the test manual, construct validity is also insufficient, since, there were only 67 children with language disorders in the sensitivity group we simply cannot rely on the sensitivity and specificity information provided in the CELF-5 to support its construct validity (Betz, Eickhoff, & Sullivan, 2013). Overall, this measure has several problems, but more importantly, it was not normed on ELs making it unreliable and invalid for assessment of this population, according to IDEA standards (2004).

The last measures are the Weschler scales, which are cognitive tests that school psychologists use to assist verbal scores on language evaluations completed by speech pathologists. The Wechsler Intelligence Scale for Children, Fifth Edition (WISC-V; Wechsler, 2014) is the most used cognitive test to support SLI in school-aged children age 6:0–16:11. Although the WISC-V generates five composite scores indices, the Verbal Comprehension Index (VCI) is made up of the Similarities and Vocabulary subtests and primarily used to identify SLI in ELs (Wechsler, 2014). The core subtests which comprise the VCI require children to define pictures or vocabulary words and describe how words are conceptually related (Wechsler, 2014). Children with expressive and or receptive language deficits often perform poorly on VCI.

Another Weschler scales used for younger preschool children is the Wechsler Preschool and Primary Scale of Intelligence, Fourth Edition (WPPSI-IV). The WPPSI-IV was developed for children ages 2:6 to 7:7 years old and only need only six subtests to calculate a FSIQ (Wechsler, 2014). The Verbal Comprehension Index (VCI) is composed of Information and Similarities. The Information subtest measures capacity to acquire, retain, and retrieve general factual knowledge. In the Similarities subtest, the child selects the picture that is from the same category as the other two pictures depicted and then verbally describes how the two objects or concepts are similar (Flanagan, Ortiz & Alfonso, 2013).

Overall, the VCI on the WISC-V and the WPPSI-IV help measure verbal reasoning, understanding, and concept formation, in addition to a child's fund of knowledge and crystallized intelligence (Flanagan, Ortiz & Alfonso, 2013). Crystallized intelligence is the knowledge a child has acquired over his or her lifespan through

experiences and learning. Furthermore, according to Goldenberg (2008), English Learners often exhibit poorer performance on the VCI due to the high language necessary to do well on these verbally loaded tests. However, low scores in these subtests tend to be inappropriately used to support classification of SLI among ELs.

Despite providers reporting use of a combination of assessment tools during their evaluations, standardized testing is the most widely used method to diagnose speech/language difficulties, with an estimated number of 90%–100% of speech language pathologists (Caesar & Kohler, 2009). In addition, most providers have ranked standardized testing as among the most important data collected during a child language assessment (Betz, Eickhoff, & Sullivan, 2013). Consequently, all these standardized assessments are frequently used to make important diagnostic and management decisions for children, including for ELs with language impairment in the academic setting. For accurate diagnosis and provision of effective intervention, it is important that assessments be chosen for use have evidence of good psychometric quality (Flanagan, Ortiz & Alfonso, 2013). However, both speech pathologists and school psychologists are not selecting assessments based on valid and reliable psychometric qualities that are normed for ELs. Therefore, emphasis needs to be placed on the selection of assessments that are evidence-based and appropriate to the needs of the child and remembering that the implications can be detrimental to ELs' academic success. Practitioners need to advocate for improvements to the quality of both currently used assessments and those developed in the future, when evaluating ELs with speech language difficulties.

English Learners and Language Acquisition

It is important to remember that during the language acquisition process, ELs usually achieve conversational proficiency within one or two years, but their ability to reach grade-appropriate academic proficiency can take much longer (Hakuta, Butler & Witt, 2000). In terms of linguistic differences, variables such as type of native language(s), proficiency of native and English language abilities (in oral language skills, reading, writing, and listening) as well as the type of instructional program provided at school, have significant impact on long-term academic performance (Hakuta, Butler & Witt, 2000). Cummins (1984) proposed two types of language proficiencies: Basic Interpersonal Communication Skills (BICS) and Cognitive Academic Language Proficiency (CALP). BICS is characterized as more conversational language that is cognitively undemanding and embedded in context that takes one to three years to develop. On the other hand, CALP is characterized by more complex language skill that is required for academic learning and thus requires students to carry advanced interpersonal conversations with clarity and efficiency (Rhodes, Ochoa & Ortiz, 2005).

According to Cummins (1984), EL students can take at least five to seven years to gain a level of linguistic proficiency (i.e., CALP) in the second language that is comparable to their native English-speaking peers. Research further indicates that by the third year ELs have not yet fully developed their CALP in order to perform well on mandated state tests given in English (Rhodes, Ochoa, & Ortiz, 2005). However, educators and practitioners continue to widely mistake the progression of ELs language acquisition with impairment. Many believe that if an EL has had three years of English language and can demonstrate fluent social conversational skills, they should then be

performing at the same level as their monolingual peers. This assumption fails to consider that ELs students have not mastered CALP, which usually takes a longer period to achieve (5-7 years). Similarly, even if the tests were given in their native language, ELs still may not be able to perform well compared to their native English-speaking peers (Rhodes, Ochoa, & Ortiz, 2005). Since, CALP proficiency in the native language is first required to develop CALP proficiency in the second language (Rhodes, Ochoa & Ortiz, 2005). Most ELs do not master CALP in their native language since they begin school in U.S. where English only educational standards exist not allowing for the progress of their native language (Hakuta, Butler & Witt, 2000).

In addition, school psychologist and speech language pathologist use varied methods of modification to assess children of diverse backgrounds. According to Ortiz (2014), many practitioners learn to "test limits", which is defined as the changing of standardized testing by providing clues, eliminating culturally loaded content, changing the modality, and modifying time constraints. The purpose of modification is to help the assessor identify a child's true ability on a test. However, according to Flanagan, Ortiz & Alfonso (2013), these modifications violate standardization, introduce error, as well as prevent valid and defensible interpretation. Therefore, if a practitioner wants to know how an EL would perform under such conditions, tests need to first be administered in the standardized manner and then re-administered with alterations (Ortiz, 2014). However, this recommendation is rarely practiced in the field as practitioners have large referral cases loads with specific deadlines and often have little or no training in best practices for evaluating ELs that present with speech language difficulties.

Socioeconomic Status, Linguistic-Prior Knowledge and Cultural Factors

There are other factors that further the problems with the traditional evaluation methods and tools discussed. These range from linguistic, prior knowledge, cultural, and social economic biases that affect the psychometric results of these measures, possibly making them invalid for testing ELs. Each one of these factors affects the language development of ELs, as it creates a disproportionate developmental disadvantage as compared to monolingual English speakers.

Socioeconomic Status

Recent studies have shown that socioeconomic status (SES) positively correlates with vocabulary knowledge and language acquisition. According to DeNavas-Walt & Proctor (2015), Hispanic children are more likely to live in poverty than non-Hispanic mainstream children; one in every five Hispanic children in the United States lives in poverty in comparison to one in every 10 non-Hispanic mainstream children. This finding is especially critical for this study given that SES is associated with low language skills due to reduced quantity and quality of input (Chodrogianni & Marinis, 2011). Therefore, children from low SES backgrounds typically perform worse than expected on standardized language measures (Gilliam & de Mesquita, 2000). Low scores on standardized language tests could be the result of an ELs profile, low parental education, or a combination of any number of risk factors (Chodrogianni & Marinis, 2011).

According to Campbell and Dollaghan (1997), vocabulary development is strongly tied to a child's experiences and exposure to language, which is influenced by their SES and cultural background. Studies suggest that a child's vocabulary correlates with their family's socio-economic status. Parents with low SES used fewer words per

hour when speaking to their children than parents with professional skills and higher SES (Hart & Risley, 1995). Typically developing EL children frequently show smaller vocabularies or limited vocabulary for certain topics in each of their languages than monolingual children on standardized and non-standardized tests, due largely to their limited use or exposure to the languages across certain contexts (Genesee & Crago, 2004). Thus, children from families with a higher SES will have larger vocabularies and will show a higher performance on standardized child language tests. Gilliam and de Mesquita (2000) found that EL children from low SES homes performed worse than higher SES peers on norm-referenced vocabulary tests such as the Peabody Picture Vocabulary Test IV (PPVT-4) and Expressive Vocabulary Test II (EVT-2). However, it is important to note that lower vocabulary scores possibly due to low SES, may not necessarily suggestive of language delays or disorders, especially in young children, as it may be part of normative language acquisition (Fernández,1992). Consequently, practitioners do not account for these factors when assessing vocabulary development and disproportionately identify ELs with SLI (Gilliam & de Mesquita, 2000).

Several questions on the CELF-5 tests such as Formulating Sentences, Word Classes and Word Definitions, place a heavy emphasis on vocabulary, which may be more difficult for a student from low SES. For example, item four on the Word Definitions test requires a student to define the word “cactus” (Horton-Ikard & Weismer, 2007). This word is not a commonly used word in most regions of the U.S. Therefore, it may pose a challenge for children from low SES who have decreased opportunities for vacations or whose parents have less experience with uncommon vocabulary items (Klingner, Artiles & Mendez-Barletta 2006). As a result, these students demonstrate

reduced lexical diversity, and therefore, test focusing on vocabulary may result in reduced scores for children from low SES relative to their higher SES peers (Paradis, 2005).

Linguistic Factors and Prior Knowledge

A study by Paradis (2005) found that EL children may show similar characteristics to children with speech language impairments when assessed by language tests that are not valid, reliable, and free of bias. A child's performance on these measures may also be affected by their prior knowledge and experiences. For example, on the CELF-5, subtest questions with words such as "giraffe" or "snowman", may not be known by all children (Paradis, 2005). These subtests contain low frequency vocabulary words that students from certain areas may have less experience with (Genesee & Crago, 2004). It is also important to consider that the format of the test may affect a child's performance if they do not have prior experiences with the specific type of testing. According to Peña, Spaulding & Plante (2006), children from culturally and linguistically diverse backgrounds do not perform as well on assessments that contain tasks such as labeling and known information questions, as they are not exposed to these tasks in their culture. Thus, typically developing students learning English as a second language may be diagnosed as having a language disorder, while in fact they are showing signs of typical second language acquisition (Rhodes, Ochoa & Ortiz, 2005).

Similarly, an EL's performance on a test may be affected by prior exposure to books. According to Peña and Quinn (1997), many ELs are not exposed to books, print, take-apart toys, or puzzles which are required on many standardized tests. For example, for a student to succeed on the CELF-5, they must possess skills such as print awareness

or the ability to recognize that pictures and symbols convey meaning (Paradis, 2005). These skills are crucial pre-literacy skills that develop through early exposure to books and print materials and do not develop in the absence of such exposure. Delayed pre-literacy skills lead to reduced metalinguistic ability, a critical skill for the Formulating Sentences subtest on the CELF (Paradis, 2005). This subtest requires students to create sentences using a provided target word as well as the need to have prior knowledge of the stimulus word to provide appropriate answers. A student who has had limited exposure to language through books, word games and print material may be challenged on this test (Paradis, 2005). Some ELs may have difficulty with these activities due to lack of prior exposure to language.

Cultural Factors

According to Peña & Quinn (1997), tasks on language assessments do not consider cultural variations in socialization practices. For example, the ELs response to the type of questions that are asked (e.g., known questions) and how the child is required to interact with the examiner during testing may be affected by the child's cultural experiences and practices (Paradis, 2005). Have there been any consideration that during test administration; children are expected to interact with strangers. Although common for mainstream American culture, in other cultures, it may be customary for a child not to speak until spoken to (Paradis, 2005). If a child does not respond to the clinician's questions because of cultural traditions, they may be falsely identified as having a language disorder (Paradis, 2005).

On the CELF-5 for example, many of the pragmatic activities on the Pragmatics Activities Checklist (PAC) relate to culturally specific nonverbal skills including gaze,

gesture, expression and body language (Paradis, 2005). The PAC includes games as well as a requirement to interact with an unknown adult. The PAC is highly culturally dependent and not truly indicative of a speech deficit (Paradis, 2005). These items could be difficult for a child from a cultural background that differs from the examiner or from mainstream American culture.

Overall, the traditional measures currently used to assess ELs with SLI have significant problems as they do not take into consideration all aspects of the EL student. A child that is exposed or learning two languages is different than one that has sole exposure to the mainstream language. Traditional assessments do not consider an ELs' experience and level of exposure to English. Studies have also shown socioeconomic, cultural and linguistic bias, since ELs' experience differ markedly from monolingual English-only speakers in the US. Finally, these measures are psychometrically inaccurate due to English-only norm samples, making them invalid and unreliable when assessing ELs. Consequently, the overidentification of ELs into special education and the misidentification of SLI classification clearly have a negative impact on ELs academic achievement and future life trajectory.

Nevertheless, there have been recent developments for the evaluation, testing and assessment of ELs that attempt to take into account the challenges discussed. The Ortiz Picture Vocabulary Acquisition Test (Ortiz PVAT) is a receptive vocabulary assessment that accounts for the length of language exposure among ELs (Ortiz, 2018). The Ortiz PVAT is an alternative to current tests because it makes it possible to evaluate an individual from any language background and determine whether an individual's English vocabulary performance follows expected patterns (as either a native English speaker or

an English Learner) (Ortiz, 2018). By using language exposure norms, it provides an advanced tool for evaluators to best assess for language difference versus a language disorder (Ortiz, 2018).

Specifically, the Ortiz PVAT has dual norms for both a monolingual native-English speaking sample and an English Learner sample stratified in the same manner, with the exception that race and ethnicity are replaced as stratification variables by language spoken (53 different ones are included) and more importantly by amount of lifetime exposure to English (Ortiz, 2018). The Ortiz PVAT first helps determines how much English experience an EL has acquired and then comparing it with other ELs with the same amount of experience in learning English (length of language exposure). Thus, the differences in language development are effectively controlled and provide a more equitable and fairer basis for evaluating potential differences in receptive vocabulary acquisition as related to overall language acquisition (Ortiz, 2018). In addition, the Ortiz PVAT has two versions that can be used at pre and post assessment to measure a student's progress throughout the academic school year. If the Ortiz PVAT's novel dual norms approach is indeed able to provide true language comparison based on experience for ELs, then it may provide an improved method regarding fairness for current and future assessment practice. In turn, this approach may allow for a possible reduction of the overidentification of ELs with SLI by being able to better distinguish difference vs. disorder.

Present Study

The purpose of this study is to determine if tests without exposure norms (e.g., WISC-IV VCI, CELF language scores) produce scores and data that are consistent with scores measuring performance on tests with exposure norms (e.g., Ortiz PVAT) for identifying SLI in ELs. Traditional methods for SLI identification have focused on English only norms which, according to recent studies, exclude ELs' true abilities. The effects that English-only norms have a significant impact on the use of evaluating measures for SLI, such as the CELF, PLS, and WISC-V, the latter being used primarily by school psychologists as opposed to speech-language pathologists. Studies have verified that these norms can increase or decrease scores depending on how they are used and interpreted. Hence, many ELs are incorrectly identified as having a disorder such as SLI rather than a difference of normal language acquisition. This difference among ELs may be affected by the length of time learning English as well as the child's age, cultural and linguistic diversity (Ortiz, 2018). The Ortiz PVAT provides an alternative to current tests because it assess for the language exposure and dual norms that can accurately assess both English speakers and ELs. Norming for language exposure allows for an enhanced cultural and linguistic competence for assessment and evaluation. This would allow for a more inclusive and fair measure of evaluation to assess whether current ELs identified with SLI have a true disorder or if their scores relate to normal second language acquisition. Furthermore, it is expected that ELs will score lower on tests with non-exposure norms (e.g, the EVT-2, PPVT-4, PLS-5, WISC-V and CELF-5) as compared to tests with exposure-based norms (e.g., the Ortiz PVAT).

The clinical significance of this study is to increase awareness of the potential value of using assessments to help diagnose and identify best practices when identifying EL children with communication difficulties. It is possible that the Ortiz PVAT may have relevance in improving our understanding of ELs, whose communication difficulties appear to change over time and whose poor functioning cannot be explained by their scores on other language assessments.

Hypothesis: This study aimed to determine if there was a statistically significant difference in how ELs with SLI scored on the non-exposure tests vs. normed referenced exposure-based test (Ortiz PVAT). If the scores obtained from tests that rely on non-exposure norms (e.g., WISC VCI, CELF language scores) to identify SLI in ELs are comparable to that obtained via use of tests with exposure norms (e.g., Ortiz PVAT), then there would appear to be little benefit regarding the time, effort, and complexity in developing norms that control for varying levels of language exposure and experience. Conversely, if use of a test that controls for differences in English exposure and development produces values that differ significantly from tests that do not use them, and where such scores and differences do not suggest language impairment, it would suggest that the use of tests without exposure norms are likely to be highly discriminatory and result in higher rates of misidentification of language impairment.

Methods

Participants and Procedures

This study included ELs that have been identified with a SLI through traditional methods by speech pathologists and school psychologists. Identification of a SLI was confirmed through a records review of Individual Education Plans (IEPs), speech and psychoeducational evaluations, school data, parental confirmation and via other measures. The participants ranged from age of 2.6 to 18 years old in accordance with range from the Ortiz PVAT. According to parent interviews, most children in our sample came from low-SES income families and were eligible for free or reduced lunch programs at their school. In terms of this study, ELs were identified as a student with significant exposure to a language other than English (i.e., English cannot be the child's only language). The participants were recruited from clinics, schools, preschools and friends within the metropolitan areas of New York, Long Island and New Jersey. Using convenience sample recruitment, a total of 59 participants were recruited. Each participant was exposed to another language aside from English and had already been identified with SLI. After participants were identified, the parents were given a written consent which detailed the process of the study including consent of their children's participation. Verbal assent was also attained from each participant. Once the study was completed, each participant received a small incentive for their participation. A summary of the descriptive characteristics are presented in Table 1.

The sample included 59 students already classified with SLI. Based on their overall scores, 58 of the students had a receptive-expressive language impairment and only one had receptive language impairment. Most examinees were male (n = 41, 69.5%)

compared to female (n = 18, 30.5%), ranging in age from 3 to 18 years (M = 8.1, SD = 2.9). The students' grade levels ranged from pre-kindergarten to one year at college. Specifically, six students (10.2%) were in pre-kindergarten, three (5.1%) were in kindergarten, two (3.4%) were in first grade, 14 students (23.7%) were in second grade, 15 students (25.4%) were in third grade, eight (3.6%) were in fourth grade, six (10.2%) were in fifth grade, six (10.2%) were in sixth grade, and then there was only one student (1.7%) representing each of the higher grades. English exposure was also measured and ranged from 25% to 76% based on testing date and time spent in the country (M = 49.5, SD = 12.2, mode = 50.0). In total, 43 students (72.9%) were Hispanic/Latino, six (10.2%) of the students were East Asian, seven (11.9%) were European/White, and three (5.1%) were Middle Eastern. Languages spoken correspond to racial demographics, with 43 (72.9%) of students being Spanish speaking, six (10.2%) speaking either Mandarin or Cantonese Chinese, six (10.2%) speaking Russian, three (5.1%) speaking Arabic, and one student (1.7%) whose native language was Hebrew.

Measures

Verbal Ability Scores (Non-Exposure Norms)

One of the most popular cognitive ability assessments for children among psychologists is the Wechsler Intelligence Scale for Children (WISC). It was developed for ages 6 and 16, used to obtain a comprehensive assessment of general intellectual functioning. The WISC has been revised frequently over the last seven decades to update norms that reflect population changes (Wechsler, 2014). As part of a standardization study, 2,200 participants were included and closely matched to 2012 US census data on race/ethnicity, parent education level, and geographic region and balanced with respect to gender (Wechsler, 2014). It is comprised of ten core subtests and five indexes measuring a global range of abilities including verbal comprehension, working memory, fluid reasoning, processing speed, as well as visual-spatial abilities. For the purpose of this study, data was collected from the WISC-IV, WISC-V, and The Wechsler Preschool and Primary Scale of Intelligence (WPPSI). The WPPSI was developed for children ages 2.6 to 7.7 years of age. The Verbal Comprehension Index (VCI) is composed of Similarities and Vocabulary. The VCI measures verbal reasoning, understanding, concept formation, in addition to a child's fund of knowledge and crystallized intelligence. The core subtests which comprise the VCI, measure receptive language, requiring a child to define pictures or vocabulary words and describe how words are conceptually related (Wechsler, 2014). According to Goldenberg (2008), English Learners with receptive language deficits often exhibit poorer performance on the VCI. The WISC (VCI) is much the same as the WISC-IV and the WWPSI (Information and Similarities subtest), as it measures factors like reasoning with language, understanding concepts and accumulated knowledge.

Language Scores (Non-Exposure Norms)

The Clinical Evaluation of Language Fundamentals (CELF) assess a child's language and communication skills and is the primary test used by speech pathologists in schools. The CELF provides a measure of specific aspects of language formulation and content, depending on the test task and child's response. Test scaled scores are used to compare the child's performance to the typical performances of the same-age norm group. Scaled scores of 7 and 13 are 1 SD below and above the mean, respectively, and are traditionally seen as the lower and upper limits of the average range of performance. The Core Language score and the Index scores are composite scores that reflect a child's abilities in a skill area (e.g., receptive language). The CELF-V was standardized on a sample based on the March 2010 US Census and was stratified by age, sex, race/ethnicity, geographic region, and parent education level (Dollaghan, 2007). Of the 3,000 participants, 27% spoke a dialect other than Standard American English (SAE), 4% were gifted or talented, 11% had clinical diagnosis (Dollaghan, 2007). In terms of this study, language scores were gathered from the CELF-IV and CELF-V which are comparable to each other as a standardized measure of language.

The Preschool Language Scales (PLS-5) is another highly used assessment primarily used by preschools and early intervention programs to assess children for language difficulties. It was designed for use with children aged birth through 7:11 to evaluate language development and identify children who have a language delay or disorder. The test aims to identify receptive and expressive language skills in the areas of attention, gesture, play, vocal development, social communication, vocabulary (Dollaghan, Horner, Oetting, 2011). The PLS has also been updated several times to

update norms. The most recent version, PLS-5, helps the clinician in determining strengths and weaknesses in these areas in order to determine the presence and type of language disorder (e.g. receptive, expressive, and mixed), eligibility for services and to design interventions based on norm-referenced and criterion referenced scores. Scores from the PLS-4 and PLS-5 were included in this study.

Language Scores (Exposure norms)

The Ortiz PVAT (2018) is a computerized test that measures receptive vocabulary acquisition evaluating parts of speech as well as basic interpersonal and academic language proficiencies (BICS and CALP). This test has two versions for progress monitoring and used dual norming samples, including 1,530 English speakers and 1,190 English learners aged 2:6 – 22:11, with the critical inclusion of English exposure as a variable for ELs. The samples were stratified based on age, gender, race/ethnicity, parental education, and geographic region (Ortiz, 2018). Each of the 59 participants were administered the Ortiz PVAT and their scores were analyzed as part of this study.

Range Classifications

During the process of data collection, it was noticed that most study participants were assessed "bilingually", and their language and verbal ability scores were reported in a qualitative manner. This is in accordance to NY chancellor's regulations for assessment of EL students that scores are reported in a descriptive manner and are advised to be interpreted with caution due to the absence of appropriate local norms and deviations from standard procedures to accommodate "English Learners". Due to current methods for assessing ELs, most of the scores gathered were reported in a descriptive qualitative manner (i.e., extremely low, borderline, below average, average).

For the purposes of analyses, to obtain scores for language and verbal ability, archival review was used to determine the classification range provided by reports documenting the evaluation and for which a mixed-receptive language impairment was rendered. To obtain the exposure-based score, the Ortiz PVAT was administered individually to each subject. For the language score, a report of speech-language evaluation was used and the classification regarding performance was used to establish the approximate score as the actual value was not often listed. Therefore, if the category, for example, was "extremely low" the range is noted as < 70, so that the low value was set at 60, the mid-value was set at 65, and the high value was set at 70. Again, this procedure was necessary due to the relative absence of scores that are reported in "Bilingual" evaluations of English learners. The same method was used to establish scores for the measure of verbal ability using the VCI. For example, a score that was in the "low average" category would have ranged from a low of 80, to a high of 89, with 85 being the mid value. For the Ortiz PVAT, exact values were obtained via direct testing and thus were used as they were derived. In sum, the use of the three ranges of the low, mid and high values was intended to provide more of a nuance in evaluating differences in performance apart from the just a mid-value which would likely have been the closest in terms of magnitude to the actual value.

Results

The main hypothesis of this study was to determine if tests without exposure norms (e.g., Weschler (VCI) and Language (CELF &PLA)) for identifying SLI in ELs produces scores and data that are consistent with scores measuring performance on tests with exposure norms (e.g., Ortiz PVAT). The rationale behind the hypothesis is that the Ortiz PVAT differs from other tests because it controls for language exposure whereas all other tests typically used to identify SLI, do not. Thus, it was hypothesized that speech-language assessments that relied on, or in part, on a language score or other verbal ability measure drawn from non-exposure norms, would be lower than a score from the Ortiz PVAT which uses exposure-based norms.

As discussed previously, for the purposes of analyses, this study had to rely on descriptive ranges for determining a series of values representing the lowest, middle, and highest possible scores with which to conduct appropriate comparisons using the general language and verbal ability scores. Table 2 and Figure 1 show descriptive statistics for the values that were obtained via this process as well as the scores taken directly via the Ortiz PVAT.

Preliminary analysis of the mean scores for the sample was conducted to shed further light on the nature of the results regarding performance on each of the tests. These results are presented in Table 3 where the bivariate correlation comparisons conducted for the variables of interest are listed. Of note are the very high associations between and within the score groupings according to language and verbal ability. The VCI scores correlated exceptionally high with each other and the language scores also correlated exceptionally high with each other. This is likely the result of having to use range

classifications for estimating performance such that all scores went down (when using the lowest value) the same magnitude and all scores went up (when using the highest value) the same magnitude. There was very little correlation between the language and verbal ability scores, and the associations are near zero. There was, however, a slightly positive and very consistent correlation between the language and verbal ability scores and the receptive language scores from the Ortiz PVAT. Although the association is very low, it remains rather consistent across all scores and begins to suggest that, in general, scores on the Ortiz PVAT appeared to increase slightly regardless of the value used for general language or verbal ability.

To test the main hypothesis posited earlier, multiple paired-sample *t*-tests were conducted to determine if there was a statistically significant difference in how students scored on the non-exposure norm-based tests (language and verbal ability) vs. the exposure-based test (i.e., receptive language). Since there were multiple simultaneous *t* tests the Bonferroni correction was applied, which lowered the alpha value from $\alpha = .05$, $df = 58$ equal to .001 to avoid any false positives making finding significance more stringent. A summary of the findings from these analyses is listed in Table 4.

Results showed that the Ortiz PVAT score was statistically significantly higher from the possible low, mid, and high verbal ability VCI scores. The Ortiz PVAT scores ($M = 100.31$, $SD = 7.99$) were 27.20 points (95% CI = [-30.27, -24.13]) higher than the low cutoff VCI ($M = 73.10$, $SD = 9.93$); $t = -17.73$, $p < .001$. The Ortiz PVAT scores were 22.71 points (95% CI = [-25.85, -19.57]) higher than the mid cutoff range for the VCI ($M = 77.59$, $SD = 10.14$); $t = -14.47$, $p < .001$. Lastly, the Ortiz PVAT scores were 19.05 points (95% CI = [-22.36, -15.74]) significantly higher on average than the high

cutoff VCI (M = 81.25, SD = 10.81); t = -11.53, p < .001. These results indicate that the null hypothesis regarding no differences must be rejected in favor of the alternative hypothesis because scores from a test that uses exposure-norms (Ortiz PVAT) were substantially and significantly higher than scores from tests using non-exposure norms. In addition, the Ortiz PVAT had the highest minimum score and the highest maximum score that were obtained, which shows that the overall range of performance in the Ortiz PVAT was substantially higher than all other scores. Likewise, the magnitude of the differences were statistically significant. These results suggest that the Ortiz PVAT scores differ significantly from the standard verbal ability VCI and provides further evidence for notable differences in results obtained in these tests.

The Ortiz PVAT score was also significantly higher than the low, mid, and high cutoff scores for the language tests. The Ortiz PVAT scores were 32.85 points (95% CI = [-35.8084-29.8865]) higher on average than the low range language test scores (M = 67.46, SD = 9.40); t = -22.21, p < .001. The Ortiz PVAT scores were 28.27 points (95% CI = [-31.19, -25.35]) higher than the language test mid cutoffs (M = 72.03, SD = 9.21); t = -19.37, p < .001. Lastly, the Ortiz PVAT scores were 24.61 points (95% CI = [-27.53, -21.69]) higher than the language score high cutoffs (M = 75.69, SD = 9.20); t = -16.85, p < .001.

Not only, were the Ortiz PVAT scores statistically significantly higher than the low, mid, and high cutoff scores for both the verbal ability VCI and language scores, they were substantially so. A Cohen's d analysis found that the size of the differences in the study were exceptionally large. The effect sizes ranged from 2.86 to 3.77 which are considered large effects. These results show that an exposure-based normed test (Ortiz

PVAT) yielded notably different results from non-exposure normed tests for classification of language-based disorders. A summary of Cohen's *d* effect sizes are presented in Table 5 and Figure 2.

Discussion

In the study, all the participants had significant exposure to another language and were already identified with a SLI. The smallest amount of lifetime exposure to English being 25% and the greatest amount being 76%. The participants ranged from preschool age, the youngest being three years old to the oldest being 18 years old in college, with a higher frequency of 8 years old participants in third grade.

Based on the results of this study, it was found that the use of tests that are built using exposure-based norms (e.g., the Ortiz PVAT) seem to produce scores that are substantially higher and largely within the average range when used with English learners, in comparison to the scores that are derived from tests using non-exposure norms. The results showed that participants on average scored over one standard deviation below the mean on both the verbal ability VCI and the language scores compared to Ortiz PVAT scores.

In terms of variability VCI scores ranged around 44 points, while language scores ranged around 25 points and PVAT scores ranged 39 points. Although all VCI scores were significantly low (i.e., VCI Low 73.10, VCI Mid 77.59 and VCI High 81.25) the scores were even lower for the Language sores (i.e., Lang Low 67.46, Lang Mid 72.03 and Lang High 75.69). This finding can be viewed as evidence that the group was indeed homogeneous as all participants were already identified as speech language impaired and not as learning disabled.

Furthermore, although there was a statistically significant difference between the scores among verbal ability VCI, language and the Ortiz PVAT there were other interesting factors. There was a close association between language and verbal ability

scores used in identification of ELs as having mixed receptive-expressive language disorders as the difference between them was only about 6 points (between the verbal ability VCI and the language scores) but both scores were in classification ranges that suggested impairment (SS=77 and SS=72, respectively). In contrast, on the Ortiz PVAT, which specifically measures receptive language, the mean score was squarely average at 100.31. There were only 3 cases where performance on the Ortiz PVAT was low enough to even suggest the possibility of a speech-language impairment and those scores were between 82-84 and represent only about 5% of the entire sample. Thus, there appear to be conflicting results which are problematic since children are scoring significantly lower on the verbal ability VCI and language assessments which identifies them as having a mixed expressive-receptive language impairment as compared to scoring within the average range in receptive language. This may indicate that since the Ortiz PVAT controls for language exposure, it provides a better measure for assessment of receptive language among English learners.

There are several reasons that can account for the differences in the scores between Ortiz PVAT, verbal ability VCI and language scores. First, although the participants had an overall identification of speech-language impairment, they had a specific classification of mixed receptive and expressive language disorder—that means, the child needed to have had both problems, not just one. Hence, if they only had one, they would have either a receptive problem or an expressive problem. According to the DSM-V (American Psychiatric Association, 2013), mixed receptive-expressive language disorder is a communication disorder in which both the receptive and expressive areas of communication are affected). Children with this disorder have difficulty understanding

words, sentences receptively and have deficiencies with expressive language. This classification indicates a receptive problem which should then correlate highly with the Ortiz PVAT receptive language score. However, there was a low correlation between these measures.

In a prior study by Adamek (2019), there was evidence of high correlations between the Ortiz PVAT and the VCI and VCI subtests in Polish-English bilingual, non-referred, normal ability samples in the U.S. This also provides supporting evidence that the Ortiz PVAT and VCI should have had moderate correlation between them. One of the reasons for a low correlation may be due to a restricted range as most participants had low scores that were not incredibly different from one another. Furthermore, the technical manual of the Ortiz PVAT shows a correlation between the WISC-V VCI and the Ortiz PVAT in a non-referred, normal ability sample, so once again there should be a correlation among them. However, although the differences were statistically significant the correlations between the Ortiz PVAT, verbal and language scores were small. This may be because of the nuances of each test, for example the WISC-V (Vocabulary and Similarities) measures more expressive abilities while other language tests may measure broader abilities aside from receptive language skills which is the ability measured by the Ortiz PVAT.

Still, a consideration of statistically error can be presumed, nonetheless, a Bonferroni correction was applied during the analysis which lowered the alpha value from $\alpha = .05$, $df = 58$ equal to .001 to avoid any false positives making it significance more stringent to find. Lastly, the effect sizes in the study were extremely large ranging

from 2.86 to 3.77 which could not be attributed to measurement error or some type of systematic problem in terms of reliability or classification for the Ortiz PVAT.

By all accounts, the scores should have been the same even in a group with speech impairment in receptive language. That there was not, suggests that the only real difference is that Ortiz PVAT uses exposure norms and the others used non-exposure norms. As a way of investigating this issue further, post hoc analyses were conducted to examine the performance of the sample relative to amount of exposure. If the Ortiz PVAT results in scores that are different in magnitude based on exposure, then this might also account for the lack of correspondence in scores. To that end, the sample was divided into two groups, those with exposure that ranged from 1% - 49% and those with exposure that ranged from 50% to 99%. If exposure makes no difference, then it is possible that the means of the two groups could have varied substantially such that one or the other may have been more consistent with the low scores of the other tests. The results of this post hoc analysis is presented in Table 6.

A review of the results in Table 6 show that the two groups differed by only 2.88 points with the low exposure group actually having a higher mean than that of the high exposure group. However, comparison of the two means showed that there was no statistically significant difference between the two values (p = .56; NS). This means that in fact, exposure differences were so well controlled on the Ortiz PVAT that its use with the sample did not reveal any actual or meaningful differences in terms of performance regardless of the individual's degree of exposure to English. This further supports the idea that Ortiz PVAT is generating fair and accurate scores that seem to be valid for this population in contrast to the very low scores that are suggesting disorder when derived

from non-exposure norms. As noted, within the sample, the performance of the subjects was within the average range according to the Ortiz PVAT in 95% of the cases. This strongly suggests that the vast majority of English learners who are being evaluated with current non-exposure based standardized tests, may in fact be over identified as having speech-language impairment when in reality they are average in their English language acquisition.

In sum, there was a large statistically significant difference between the verbal ability VCI, language and the Ortiz PVAT scores. The differences in the scores of non-exposure measures were over one standard deviation lower than the scores on the language exposure measures (Ortiz PVAT). Interestingly the scores for the Ortiz PVAT were squarely average at 100.31 suggesting that participants were scoring relatively average when language exposure was accounted for. This further implies that if Els are being assessed with verbal ability VCI and language measures they would score within the below average range and then classified with a SLI. On the other hand, when an ELs was assessed using the Ortiz PVAT their score would fall within the average range without a disability. These findings suggest that the use of a normed reference test (Ortiz PVAT) that accounts for English exposure was vital for best practice when assess ELs to distinguish “difference from disability”.

Study Limitations

Although the current study expanded evidence that tests that have exposure-based norms (e.g., Ortiz PVAT) provide best methods for evaluating ELs that exhibit difficulties with speech language, the study had its limitations. Due to the current methods for assessing ELs in New York most of the scores gathered were reported in a descriptive qualitative manner (i.e., extremely low, borderline, below average, average). This is in accordance to NY state chancellor's regulations for assessment of EL students that scores are reported in a descriptive manner and are advised to be interpreted with caution due to the absence of appropriate local norms to accommodate "English Learners". In the study about 20% of the actual scores were reported, the remaining 80% were presented in a qualitative manner. As a result, the use of low, mid and high ranges were established for each descriptive score based on the information gathered from psycho-educational and speech language reports. Ideally, having the exact standard scores for each participant, would have given the study more precise comparison. However, the ranges were not only statistically significant but also yielded such high effect sizes for each subgroup (low, mid, high) that having a precise score may just have added to the size in magnitude. Gathering exact standard scores would have also been problematic as most ELs are evaluated "bilingually" by school psychologists and speech pathologists, which did not report standard scores.

Another limitation for this study was the overall sample demographics. While the sample size of n =59 was large enough to yield over 85% power, there were several limitations. A nonclinical sample was not included, which would have been helpful to expand the range of functioning. A nonclinical sample may have provided higher

functioning participants which would have allowed for a full range of scores rather than clinical sample scores which were already classified as Speech Language Impaired. Also, the sample was all from the New York area, a metropolitan city, which does not represent the geographic diversity of the entire country. The sample had specific geographic boundaries which limits its generalizability. In addition, gathering more demographic information such as level of parent education, parental English level/ability, and social economic statues could have been good variables for future studies. Due to time constraints these variables were not fully evaluated but assessing for the level of native language dominance of parental figures for example could have impacted the performance of participants within this study. Since the quality of native language expose and social economic status can help the process of second language acquisition. Future research can explore gathering more specific sample demographics and attaining a non-clinical sample.

The last limitation in the study was the differences in the tests constructs while each test was good for general language development there are variations within each test. The clinical sample in this study were already classified with mixed receptive and expressive language disorder but each test utilized, measured slightly different abilities such as expressive, receptive, and mixed expressive receptive. Thus, future studies can better control for these specific variations.

.

Implications for School Psychologists

Findings from this study suggest, that school psychologist and speech pathologist should be mindful when selecting standardized tests that are representative of the population they are working with, in this case ELs with speech language difficulties. ELs demonstrate an array of strengths and weaknesses related to language exposure which produce varying results and not accounting for language exposure has detrimental effects because not all ELs have the same language exposure. Children that are acquiring another language will follow a natural process that takes time to develop. This does not mean that ELs are disabled. Historically, however, the over and misidentification of ELs into special education has had detrimental effects on their achievement in school, graduation rates and future outcomes. Hence, regardless of the limitations mentioned above, this study can help inform clinical assessment when interpreting the psychometric properties of measures, instruments and selecting the most valid and reliable tools for working with ELs that have speech language difficulties.

Since there are estimates that by 2023, ELs will be the largest subgroup in school systems around the U.S. it is vital that tests developers and researchers improve norming samples that assess for language exposure and have psychometrically appropriate norms. Current standardized practices of evaluation of ELs with verbal ability VCI and language scores are not an appropriate measure for identification of SLI among ELs. These measures should not be used for ELs since they are not specifically normed on ELs and do not account for language exposure. Scores should not be interpreted with the statement with "caution" because they are still interpreted as valid due to clinical interpretation by providers which continue to increase the overidentification of ELs into

special education. This study has demonstrated that the Ortiz PVAT is an alternative to current methods of assessing ELs with Speech Language Impairments. Practitioners are encouraged to use the Ortiz PVAT, since there is supporting evidence that it is a user friendly, computerized and normed reference for language exposure according the child's age. Overall, the unique language exposure norms of the Ortiz PVAT can best assess "language difference vs. disorder" among English Learners with speech language difficulties.

Appendix

Table 1.

Sample Characteristics (*N* = 59)

Category	*n*	%
Gender		
Male	41	69.5
Female	18	30.5
Ethnicity		
Hispanic/Latino	43	72.9
East Asian	6	10.2
White	7	11.9
Middle Eastern	3	5.1
Language		
Spanish	43	72.9
Mandarin or Cantonese	6	10.2
Russian	6	10.2
Arabic	3	5.1
Hebrew	1	1.7

Table 2.

Descriptive Statistics of Verbal Ability (VCI) Language and Ortiz PVAT scores (N = 59)

Test	Mean	SD	Min	Max	Range	Mode
VCI Low Range	73.10	9.93	58	102	44	70
VCI Medium Range	77.59	10.14	58	102	44	75
VCI High Range	81.25	10.81	58	110	52	79
Language Test Low Range	67.46	9.40	60	84	24	60
Language Test Medium Range	72.03	9.21	65	85	20	65
Language Test High Range	75.69	9.20	69	89	20	69
Ortiz PVAT Score	100.31	7.99	82	121	39	100

Table 3.

Correlations among Verbal Ability (VCI), language and Ortiz PVAT scores

(*N* = 59)	1	2	3	4	5	6	7	8
2. VCI Low	…	…	**.98**	**.92**	.07	.08	.09	.15
3. VCI Mid	…	...	…	**.99**	.03	.04	.05	.13
4. VCI Hi	…	…	…	…	-.01	.00	.02	.11
5. Lang Low	…	…	…	…	…	**.99**	**.96**	.15
6. Lang Mid	…	…	…	…	…	…	**.99**	.16
7. Lang Hi	…	…	…	…	…	…	…	.15
8. PVAT	…	…	…	…	…	…	…	…

Note. Bold indicates $p < .001$. VCI Low = Low range VCI score, VCI Mid = Midrange VCI score, VCI Hi = High range VCI score, Lang Low = Low range language test score, Lang Mid = Midrange language test score, Lang Hi = High range language test score, PVAT = Obtained PVAT score

Table 4.

Paired sample t-tests among differences in test scores for Verbal Ability VCI, language and Ortiz PVAT scores

Test comparison	*Mean diff*	*SE*	*p*	*t*
VCI Low – Ortiz PVAT	-27.20	1.53	< .001	**-17.73***
VCI Mid – Ortiz PVAT	-22.71	1.57	< .001	**-14.47***
VCI High – Ortiz PVAT	-19.05	1.65	< .001	**-11.53 ***
Lang Low – Ortiz PVAT	-32.85	1.48	< .001	**-22.21***
Lang Mid – Ortiz PVAT	-28.27	1.46	< .001	**-19.37***
Lang High – Ortiz PVAT	-24.61	1.46	<.001	**-16.85***

Note. df for all tests are equal to 58. Bold and * indicates $p < .001$ per calculated Bonferroni correction.

Table 5.

Cohen's d effect sizes for paired sample t-tests among test scores

(*N* = 59)	1	2	3	4	5	6	7
1. VCI Low	…	.45	.79	.58	.11	.27	**3.02**
2. VCI Mid	…	…	.35	1.04	.57	.20	**2.49**
3. VCI High	…	…	…	1.36	.92	.55	**2.01**
4. Lang Low	…	…	…	…	.49	.88	**3.77**
5. Lang Mid	…	…	…	…	…	.40	**3.28**
6. Lang High	…	…	…	…	…	…	**2.86**
7. PVAT	…	…	…	…	…	…	…

Note: Cohen's *d* = Small = .2; Medium = .5; Large = .8

Table 6.
Independent samples t-test between high and low English exposure Ortiz PVAT scores

English Exposure	*n*	mean	*SD*	*Mean diff*	*df*	*SE*	*t*
Low	37	101.38	8.36	2.88	57	2.14	1.35
High	22	98.50	7.13				

Note. No Significance Indicated $p < .001$

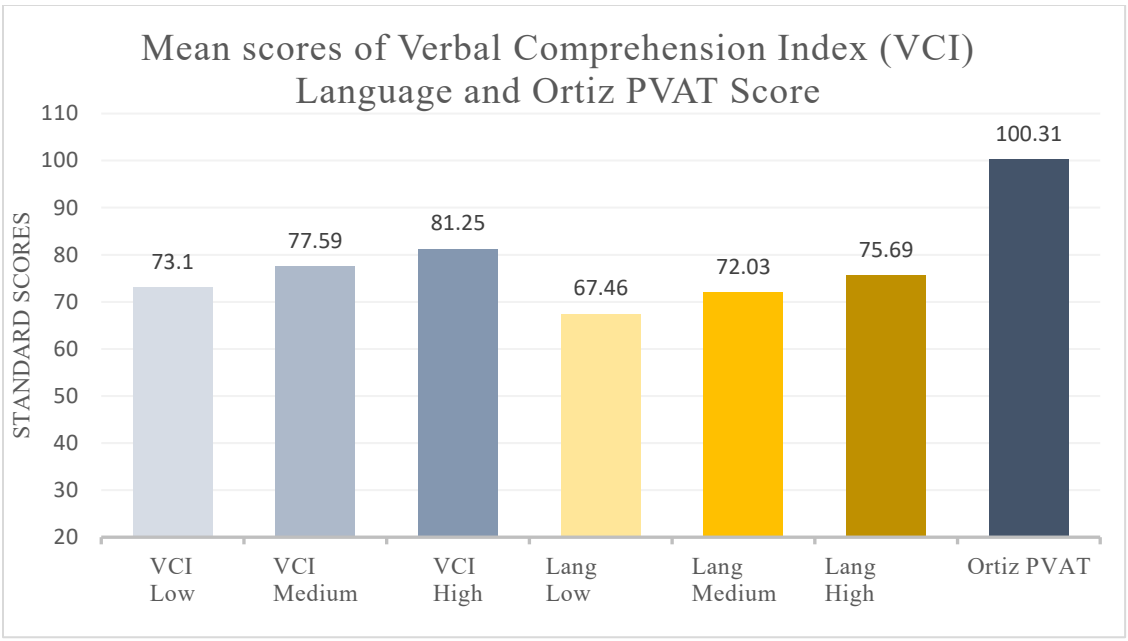

Figure 1. Overall mean standard scores of Verbal Ability (VCI) (low, mid, high ranges), Language Scores (low, mid, high ranges) and Ortiz PVAT scores.

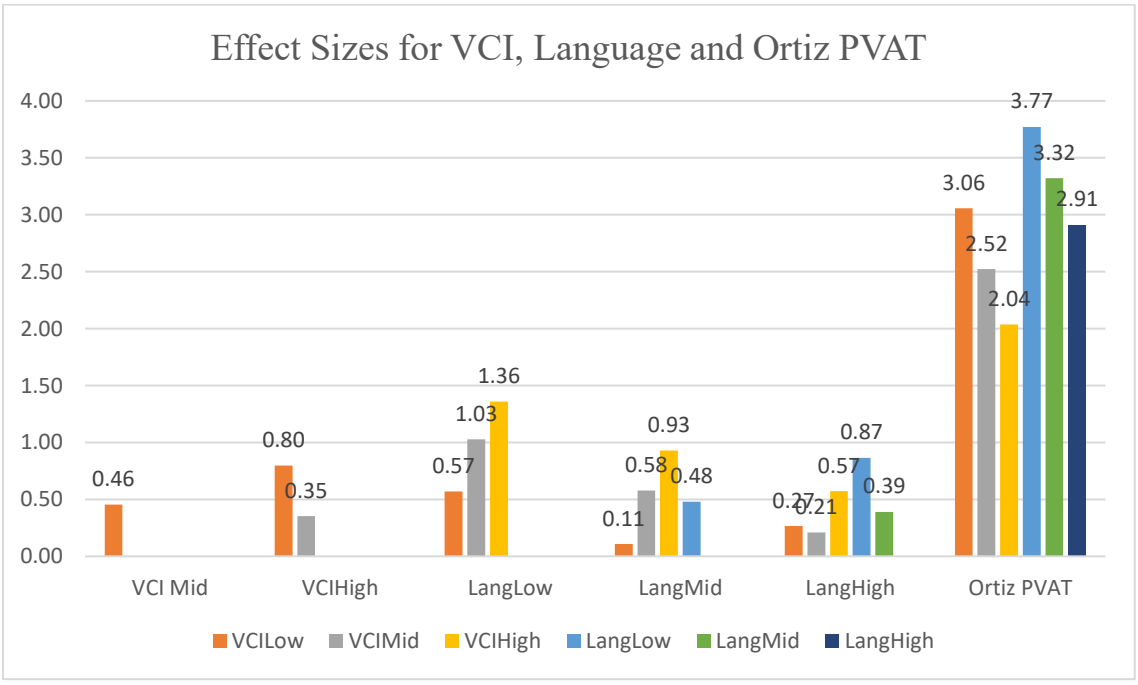

Figure 2: Effect sized of Verbal Ability (VCI) (low, mid, high ranges), Language Scores (low, mid, high ranges) and Ortiz PVAT scores.
Cohen's *d* = Small = .2; Medium = .5; Large = .8

References

Abedi, J. (2006). Psychometric issues in the ELL assessment and special education eligibility. Teachers College Record 108, 2282-2303.

Adamek, E. (2019). Concurrent validity of the WISC-V VCI and The Ortiz PVAT on English learners [ProQuest Information & Learning]. In *Dissertation Abstracts International Section A: Humanities and Social Sciences* (Vol. 81, Issue 2–A).

American Speech-Language-Hearing Association. (2004). Knowledge and Skills Needed by Speech-Language Pathologists and Audiologists to Provide Culturally and Linguistically Appropriate Services [Knowledge and Skills].

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Artiles, A. J., & Bal., A. (2008). The next generation of disproportionality research: Toward a comparative model in the study of equity in ability differences. Journal of Special Education, 42, 4-14.

Artiles, A. J., & Klingner, J. K. (2006). Forging a knowledge has on English language learners with special needs: Theoretical, population, and technical issues. Teachers College Record, 108, 2187-2194. doi: 10.1111/j. 1467-9620.2006.00778.X

Bassai, F., & Wasyliw, O. E. (2001). Review of the Expressive Vocabulary Test. In B. S. Plake & J. C. Impara (Eds.), The fourteenth mental measurements yearbook (pp. 475-478). Lincoln, NE: Buros Institute of Mental Measurements.

Betz, S. K., Eickhoff, J. R., & Sullivan, S. F. (2013). Factors influencing the selection of test for the diagnosis of specific language impairment. Language, Speech, and Hearing Services in Schools, 44, 133-146.

Bedore, L. M., Pena, E. D., Summers, C. L., Boerger, K. M., Resendiz, M. D., Greene, K., & ... Gillam, R. B. (2012). The measure matters: Language dominance profiles across measures in Spanish–English bilingual children. *Bilingualism: Language & Cognition*, *15*(3), 616-629.

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language & Cognition*, *13*(4), 525-531.

Brayboy, B. M. J., Castagno, A. E., & Maughan, E. (2007). Equality and justice for all? Examining race in education scholarship. Review of Research in Education, v31, 159-194.

Caesar, L. G., & Kohler, P. D. (2009). Tools clinicians use: A survey of language assessment procedures used by school-based speech-language pathologists. Communication Disorders Quarterly,30(4), 226–236.

Campbell, T., Dollaghan, C., Needleman, H., Janosky, J. (1997). Reducing bias in language assessment: Processing-dependent measures. Journal of Speech, Language, and Hearing Research, 40, 519-525.

Cattel, R. B. (1943). The description of personality: basic traits resolved into clusters. *The Journal of Abnormal and Social Psychology, 38*(4), 476-506.

Cheatham, G. A., & Hart Barnett, J. E. (2017). Overcoming common misunderstandings about students with disabilities who are English language learners. *Intervention in School & Clinic*, *53*(1), 58-63.

Chondrogianni, V., & Marinis, T. (2011). Differential effects of internal and external factors on the development of vocabulary, tense morphology and morpho-syntax in successive bilingual children. Linguistic Approaches to Bilingualism, 3, 318–345.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cormier, D. C., McGrew, K.S., & Ysseldyke, J. E. (2014). The influence of linguistic demand and cultural loading on cognitive test scores. *Journal of Psychoeducational Assessment, 32*(7), 610-623.

Cummins, J. C. (1984). *Bilingual and special education: Issues in assessment and pedagogy.* Austin, TX: PRO-ED.

DeNavas-Walt, C., & Proctor, B. D. (2015). Income and poverty in the United States: 2014. Washington, D.C.: U.S. Census

Denman, D., Speyer, R., Munro, N., Pearce, W. M., Chen, Y. W., & Cordier, R. (2017). Psychometric Properties of Language Assessments for Children Aged 4-12 Years: A Systematic Review. Frontiers in psychology, 8, 1515. doi:10.3389/fpsyg.2017.01515

Dollaghan, C. (2007). *The handbook for evidence-based practice in communication disorders*. Baltimore, MD: Paul H. Brooks Publishing Co.

Dollaghana, C. A., Horner, E. A., & Oetting, J. (2011). Bilingual language assessment: A meta-analysis of diagnostic accuracy. *Journal of Speech, Language & Hearing Research*, *54*(4), 1077-1088.

Dunn, L.M. & Dunn, D.M. (2007). Peabody Picture Vocabulary Test – Fourth Edition: Manual. Minneapolis, MN: Pearson Assessments.

Fernández, M.C., Pearson, B.Z., Umbel, V. M., & Molinet-Molina, M. (1992). Bilingual receptive vocabulary in Hispanic preschool children. Hispanic Journal of Behavioral Sciences, 14, 268-276.

Figueroa, R. A., & Newsome, P. (2006). The diagnosis of LD in English learners: Is it nondiscriminatory? Journal of Learning Disabilities, 39, 206-214. doi:0.1177/00222194060390030201

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2007). *Essentials of cross-battery assessment* (2nd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Flanagan, D. P., Ortiz, S. O., Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Genesee, F., Paradis, J., & Crago, M. (2004). Dual Language Development and Disorders: A Handbook on Bilingualism and Second Language Learning. Baltimore, MD: Brookes Publishing.

Genesee, F., Lindholm-Leary, K., Saunders, W, Si Christian, D. (2005). English language learners in U.S. schools: An overview of research findings. Journal of Education for Students Placed at Risk, 10, 363-385.

Gibson, T. A., Peña, E. D., & Bedore, L. M. (2014). The receptive–expressive gap in bilingual children with and without primary language impairment. *American Journal of Speech-Language Pathology*, *23*(4), 655-667. doi:10.1044/2014_AJSLP-12-0119

Gilliam, W. S., & de Mesquita, P. B. (2000). The relationship between language and cognitive development and emotional behavioral problems in financially disadvantaged preschoolers: A longitudinal investigation. Early Child Development and Care, 162(1), 9–24.

Goldenberg, C. (2008). Teaching English language learners: What the research does—and does not—say. American Educator, 32(2) pp. 8-23, 42-44.

Hakuta, K., Butler, Y. G., & Witt, D. (2000). How long does it take English learners to attain proficiency? (pp. 1-28, Rep. No. 2000-1). University of California Linguistic Minority Research Institute.

Hanushek, E., J. Kain and S. Rivkin. 2002. Inferring Program Effects for Specialized Populations: Does Special Education Raise Achievement for Students with Disabilities? Review of Economics and Statistics 84(4): 584-599.

Hart, B., & Risley, T.R. (1995). Meaningful Differences in the Everyday Experience of Young American Children. Baltimore, MD: Brookes Publishing.

Harry, B. & Klingner, J., (2006). Why are so many minority students in special education?: Understanding race and disability in schools. New York: Teachers College Press, Columbia University

Hayward, D. V., Stewart, G. E., Phillips, L. M., Norris, S. P., & Lovell, M. A. (2008). Test review: Expressive Vocabulary Test (EVT). Language, Phonological Awareness, and Reading Test Directory (pp. 1-5). Edmonton, AB: Canadian Centre for Research on Literacy. Retrieved from http://www.uofaweb.ualberta.ca/elementaryed/ccrl.cfm.

Hoffman, L., Templin, J., & Rice, M. (2012). Linking Outcomes from Peabody Picture Vocabulary Test Forms using Item Response Model. *Journal Speech Language,* 55(3): 754–763. doi:10.1044/1092-4388(2011/10-0216).

Horton-Ikard, R., & Weismer, S.E. (2007). A preliminary examination of vocabulary and word learning in African American toddlers from middle and low socioeconomic status homes. American Journal of Speech Language Pathology, 16, 381-392.

Individuals with Disabilities Education Improvement Act of 2004, 20 U.S.C. § 1400 et seq. (2004). Reauthorization of the Individuals with Disabilities Education Act of 1990.

Keller-Allen, C. (2006). English language learners with disabilities: Identification and other state policies and issues. Alexandria, VA: Project Forum.

Klingner, J. K., Artiles, A. J., & Mendez-Barletta, L. (2006). English learners who struggle with reading: Language acquisition or LD? Journal of Learning Disabilities, 39, 108-128.

Mercer, J. R. (1972). *The System of Multicultural Pluralistic Assessment: Technical Manual.* New York: The Psychological Corporation.

National Association of School Psychologists. (2015). *The provision of school psychological services to bilingual students* [Position statement]. Bethesda, MD: Author.

National Center for Education Statistics. (2016). The condition of education 2017 (NCES 2017-144). Washington, DC: U. S. Department of Education, Institute of Education Sciences. Retrieved from https://nces.ed.gov/pubs2017/2017144.pdf

National Research Council. (2002). Minority students in special and gifted education. Committee on Minority Representation in Special Education. M. Suzanne Donovan and Christopher T. Cross (Eds.), Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press

Ochoa, S. H., Rivera, B., & Ford, L. (1997). An investigation of school psychology training pertaining to bilingual psycho-educational assessment of primarily Hispanic students: Twenty-five years after Diana v. California. Journal of School Psychology, 35(4), 329-349.

Ortiz, S. O. (2017). *Contemporary Assessment of English Learners: Evidence-based evaluation and best practice* [PowerPoint slides]. Retrieved from http://casponline.org/event-handouts/fall17/W-33%20handouts.pdf

Ortiz, S. O. (2014). Best Practices in Nondiscriminatory Assessment. In P. Harrison & A. Thomas (Eds.) *Best Practices in School Psychology VI: Foundations* (pp. 61-74), Bethesda, MD: National Association of School Psychologists.

Ortiz, S. O. (2018). *Ortiz Picture Vocabulary Acquisition Test*. Toronto, Canada: Multi-Health Systems.

Paradis, J. (2005). Grammatical morphology in children learning English as a second language: Implications of similarities with Specific Language Impairment. Language, Speech and Hearing Services in the Schools, 36, 172-187

Peña, E. & Quinn, R. (1997). Task familiarity: Effects on the test performance of Puerto Rican and African American children. Language, Speech, and Hearing Services in Schools, 28, 323-332.

Peña, E., Spaulding, T & Plante, E. (2006). The composition of normative groups and diagnostic decision making: Shooting ourselves in the foot. American Journal of Speech-Language Pathology, 15, 247-254.

Rhodes, R. Ochoa, H. S., & Ortiz, S. O. (2005). *Assessing culturally and linguistically diverse students: A practical guide.* New York, NY: Guilford Press.

Rueda, R., & Windmueller, M. P. (2006). English language learners, LD, and overrepresentation: A multiple-level analysis. Journal of Learning Disabilities, 39(2), 99-107.

Salvia, J., & Ysseldyke, J. E. (1991). Assessment (5th ed.). New York: Houghton Mifflin.

Sotelo-Dynega, M., & Dixon, S. G. (2014). Cognitive assessment practices: A survey of school psychologists. *Psychology in The Schools*, *51*(10), 1031-1045.

Sullivan, A. L. (2011). Disproportionality in special education identification and placement of English language learners. Exceptional Children, 77(3), 317-334.

Thomas, W. P., & Collier, V. P. (2004). The astounding effectiveness of dual language education for all. NABE Journal of Research and Practice, 2, 1-20.

U.S. Census Bureau. (2015). New census bureau report analyzes U.S. population projections. Retrieved from http://www.census.gov/newsroom/press-releases/2015/cb15-tps16.html.

U.S. Department of Education (2010). Twenty-ninth annual report to Congress on the implementation of the Individuals with Disabilities Education Act, 2007. Washington, DC: Office of Special Education and Rehabilitative Programs.

U.S. Department of Education. (2017). *Office of English language acquisition: Profiles of English learners* [Fast Facts Data]. Retrieved from http://www.ncela.us/files/fast_facts/05-19-2017/ProfilesOfELs_FastFacts.pdf

Wechsler, D., Raiford, S. E., & Holdnack, J. A. (2014). *WISC-V Technical and Interpretive Manual.* San Antonio, TX: The Psychological Corporation.

Williams, K. T. (2007). Expressive Vocabulary Test, Second Edition.

Zhang, D., Katsiyannis, A., Ju, S., & Roberts, E. (2014). Minority representation in special education: 5-year trends. *Journal of Child & Family Studies, 23*, 118-127.

Vita

Name	*Anna C. Tello*
Baccalaureate Degree	*Bachelor of Arts, Rutgers University, New Brunswick* *Major: Political Science and Spanish*
Date Graduated	*May, 2001*
Other Degrees and Certificates	*Master of Science, St. John's University, Jamaica* *Major: School Psychology* *Master of Arts, Rutgers University, New Brunswick* *Major: Social Work* *May 2006*
Date Graduated	*May, 2018*