



Test Review:
Preschool Language Scales- Fifth Edition (PLS-5) Spanish

Version: 5th Edition

Copyright date: 2012

Grade or Age Range: Birth through 7;11 years

Author: Irla Lee Zimmerman, Ph.D.; Violette G. Steiner, BS & Roberta Evatt Pond, MA

Publisher: Pearson

Table of Contents

<i>Section</i>	<i>Page Number</i>
1. Purpose	2
2. Description	2-3
3. Standardization Sample	3
4. Validity	3-7
a. Content	3-4
b. Construct	4-7
1. Reference Standard	4-5
2. Sensitivity and Specificity	5-6
3. Likelihood Ratio	6
c. Concurrent	6-7
5. Reliability	7-9
a. Test-Retest Reliability	8
b. Inter-examiner Reliability	8
c. Inter-item Consistency	8-9
6. Standard Error of Measurement	9-10
7. Bias	10-15
a. Linguistic Bias	10-12
1. Bilingual Speakers	10-11
2. Dialectal Variations	11-12
b. Socioeconomic Status Bias	12
c. Prior Knowledge/Experience	12-13
d. Cultural Bias	13-14
e. Attention and Memory	14-15
f. Motor/Sensory Impairments	15
8. Special Alerts/Comments	15-16
9. References	17-18

1. PURPOSE

The PLS-5 Spanish is designed to determine the presence and severity of a receptive, expressive, or receptive-expressive language delay or disorder in monolingual Spanish, or bilingual Spanish-English speaking children aged birth to 7 years 11 months. The test may also help to identify a child's strengths and weaknesses which can be used to determine appropriate intervention, and may also identify the "contexts in which a bilingual Spanish-English speaking child uses one language versus the other" (Zimmerman, Steiner, & Pond, 2012, p. 4).

2. DESCRIPTION

The PLS-5 Spanish consists of two standardized scales: Auditory Comprehension (AC) and Expressive Communication (EC). A total language (TL) composite score may also be calculated. Norm referenced scores are provided for the AC, EC, and TL scales; scores are provided at three month intervals from birth through 11 months, and at 6 months intervals from 1 year through 7;11. Specific AC tasks assessed include language precursors such as attending to speakers and appropriate object play, comprehension of basic vocabulary, concepts, morphology, syntax, comparisons and inferencing, and emergent literacy. Specific EC skills include vocal development and social communication, naming, describing, expressing quantity, using specific prepositions, grammatical markers, and sentence structures, and emergent literacy skills. The PLS-5 Spanish is designed to be a dual language assessment; each item includes an identical Spanish and English task. The Spanish item is administered to monolingual Spanish speakers. For bilingual children, the Spanish item is administered first, followed by the English item if the first response was incorrect. Scores are calculated based on the total number of correct responses, regardless of the language of administration. Three optional supplemental measures are also included (Language Sample Checklist, Articulation Screener, and *Cuestionario de Comunicación en el Hogar* [Home communication Questionnaire]). It should be noted that caregiver's responses to the Home Communication Questionnaire may support or credit items on the AC and EC scales from birth through 2;11. Finally, the PLS-5 Spanish includes three supplemental language measures to aid in analyzing a child's language skills, the Spanish Item Analysis Checklist, a Clinician's Worksheet, and the PLS-5 Spanish Profile.

The PLS-5 Spanish Manual del Examinador (Examiner's Manual) describes examiner qualifications for the test. If the test is being administered to monolingual Spanish speakers, it is highly recommended that the examiner be a fluent or near fluent Spanish speaker. If the test is being administered to bilingual Spanish-English speakers, the examiner should be bilingual as well. Also according to the manual, the test may be administered, scored, and interpreted by "Spanish-speaking speech pathologists, early childhood specialists, psychologists, educational diagnosticians, and other professions who have experience working with children of this age and training in individual assessment" (Zimmerman,

Steiner, & Pond, 2012, p. 8). If a qualified examiner is not available, the manual states that the test may be administered “in collaboration with a trained and qualified interpreter” (Zimmerman, Steiner, & Pond, 2012, p. 8). Lastly, the manual cautions that although Spanish-speaking paraprofessional staff may be trained to administer the PLS-5 Spanish and record the child’s responses, the results should only be interpreted by a “clinician who has training and experience in diagnostic assessment and knowledge in language development” (Zimmerman, Steiner, & Pond, 2012, p. 8). Specific information regarding using the test with interpreters is provided in chapter 2 of the manual.

3. **STANDARDIZATION SAMPLE**

The standardization sample was collected between May 2010 and March 2011. The standardization version of the PLS-5 Spanish was conducted by 111 examiners, including speech-language pathologists, psychologists, educational diagnosticians, and bilingual education teachers. Testing included 1,150 bilingual children in the United States and Puerto Rico. Participants were required to complete the PLS-5 Spanish in the standard manner without modifications, and understand and speak Spanish as their primary language. If the children were preverbal, Spanish was required to be the primary language spoken by their caregivers in their home. The normative sample was matched to the 2008 U.S. Census and reflects the population’s characteristics including age, sex, geographic region, caregiver’s education level, and country of origin.

4. **VALIDITY**

Content - Content validity is how representative the test items are of the content that is being assessed (Paul, 2007). Content validity was measured through literature reviews, user feedback, and expert consultation from speech-language pathologists, psychologists, educational diagnosticians, and bilingual education teachers to ensure that the test assesses skills for various stages of language development. Feedback was collected from clinicians who purchased the PLS-4 and/or the PLS-4 Spanish. Clinicians were asked to give feedback regarding scores, administration directions, content areas, test items, and picture stimuli. Based on the responses, test developers decided on aspects of the PLS-4 Spanish to use in the revision and aspects that needed to be changed. It should be noted that specific information regarding the background and training for the individuals involved was not provided. Expert knowledge of a variety of dialects and languages requires an enormous and sophisticated knowledge base. In some cases, the intricacies of dialectal variations are so small that even highly educated linguists find it difficult to determine cultural differences. Therefore, one cannot be confident that this feedback reflects the least biased content.

Tryout data was collected in February 2009 - July 2009 using 188 items, including items that were kept and modified from the PLS-4 Spanish and new items for each age group. Two samples of children are used: a nonclinical sample of 341 children aged birth through 7;11

and a clinical sample of 69 children diagnosed with a language disorder aged 2 through 7;11. The children included in the clinical group do not cover the entire age range of the PLS-5 Spanish. It should be noted that the manual did not provide information regarding how the children in the nonclinical sample were determined to be “typically developing.” Therefore, we cannot be certain of their true diagnostic status. The “clinical” (language disordered) children were required to be receiving language services at the time of testing, and were identified by a score of 77 (1.5 SD below the mean) or below on a standardized language test. Information regarding the tests that were used was not provided. Thus, one cannot be certain of the diagnostic accuracy of the test and therefore the diagnostic accuracy of the clinical children. Further, it is important to note that the children were identified based on an arbitrary cutoff score of 77. According to Spaulding, Plante, & Farinella (2006), arbitrary scores on standardized language tests often do not accurately discriminate between typically developing children and children with a language disorder. Thus, their true diagnostic status is unknown.

All children in both groups lived in a Spanish speaking home. The children were required to take the tryout test in a standard manner, without modifications (e.g., due to deficits in fine motor or sensory abilities). Results of the tryout test were used to develop scoring guidelines for new open-ended items and items were changed or deleted if they did not meet requirements for fairness, scoring ease, and item-level difficulty. Responses from the clinical and nonclinical sample were also compared and items were deleted if they did not differentiate between the two groups.

The content validity is considered insufficient due to potentially biased feedback from unqualified reviewers, and tryout data groups with questionable diagnostic accuracy.

Construct – Construct validity assesses if the test measures what it purports to measure (Paul, 2007). Construct validity was determined using the three measures listed below (*Reference Standard, Likelihood Ratio, Sensitivity and Specificity*). Special group studies of typically developing children and children with language disorders were conducted to determine if the test discriminated between the groups.

Reference Standard

In considering the diagnostic accuracy of an index measure such as the PLS-5 Spanish, it is important to compare the child’s diagnostic status (affected or unaffected) with their status as determined by another measure. This additional measure, which is used to determine the child’s ‘true’ diagnostic status, is often referred to as the “gold standard.” However, as Dollaghan & Horner (2011) note, it is rare to have a perfect diagnostic indicator, because diagnostic categories are constantly being refined. Thus, a *reference standard* is used. This is a measure that is widely considered to have a high degree of

accuracy in classifying individuals as being affected or unaffected by a particular disorder, even accounting for the imperfections inherent in diagnostic measures (Dollaghan & Horner, 2011).

The reference standard to identify the *sensitivity* group required that the child was at least one year of age, and had been diagnosed with a moderate to severe receptive, expressive, or combined receptive-expressive language disorder (as determined by a score of 77 [1.5 SD below the mean], or less on a standardized language test). All the children were also receiving speech therapy at the time of the study. The children were required to take the test in a standardized fashion (e.g. they could not have fine or gross motor impairments). It is important to note that arbitrary cut off scores on standardized language tests often do not accurately discriminate between typically developing children and children with a language disorder (Spaulding, Plante, & Farinella, 2006). Thus, the true diagnostic status is unknown. Further, the Examiner's Manual did not report which standardized tests were used to identify the children in the reference standard. Therefore, we cannot be sure of the test's diagnostic accuracy. The age range of the reference standard also did not cover the entire age range of the PLS-5 Spanish; the test claims it is for used with children aged birth to 7;11, however, accuracy was only determined for children from 1;0 to 7;11.

The reference standard to identify the *specificity* group was defined as a child who had not been previously diagnosed with a language disorder and was not receiving speech therapy at the time of the study. The children were required to take the test in a standardized fashion (e.g. they could not have fine or gross motor impairments). They were matched to sensitivity group based on age, sex, ethnicity, and primary caregiver's education level. According to Dollaghan (2007) performance on the reference standard cannot be assumed. As the same reference standard for the sensitivity group (a score of 77 on a standardized test) was not applied, it is important to consider spectrum bias (Dollaghan & Horner, 2011). As the children in the specificity group were not administered the same reference standard, their diagnostic status cannot be determined for certain, rendering the reference standard insufficient. Further, as the minimum age was 1 year, the children in the reference standard did not cover the entire age range of the PLS-5 Spanish (beginning at birth).

Sensitivity and Specificity

Sensitivity measures the proportion of students who have a language disorder that will be accurately identified as such on the assessment (Dollaghan, 2007). For example, sensitivity means that Johnny, an eight-year-old boy previously diagnosed with a language disorder, will achieve a score that identifies him as having a language disorder on this assessment. According to Plante & Vance (1994), validity measures above .9 are good, measures between .8 and .89 are fair, and measures below .8 are unacceptable. The PLS-5 Spanish reports the sensitivity to be .85, or "fair" as compared to the standard in

the field. However, it is important to consider the implications of these measures. A sensitivity of .85 means that 15/100 children who have a language disorder will not be identified as such by the PLS-5 Spanish, and therefore will not receive the extra academic and language support that they need.

Specificity measures the proportion of students who are typically developing who will be accurately identified as such on the assessment (Dollaghan, 2007). For example, specificity means that Peter, an eight-year-old boy with no history of a language disorder, will score within normal limits on the assessment. The PLS-5 Spanish reports the specificity to be .88, which is considered to be “fair” (Plante & Vance, 1994). It is important to consider that a specificity of .88 means that 12/100 typically developing children will be identified as having a language disorder and may be unnecessarily referred for special education services.

Likelihood Ratio

According to Dollaghan (2007), likelihood ratios are used to examine how accurate an assessment is at distinguishing individuals who have a disorder from those who do not. A positive likelihood ratio (LR+) represents the likelihood that an individual who is given a positive (disordered) score on an assessment actually has a disorder. The higher the LR+ (e.g. >10), the greater confidence the test user can have that the person who obtained the score has the target disorder. Similarly, a negative likelihood ratio (LR-) represents the likelihood that an individual who is given a negative (non-disordered) score actually does not have a disorder. The lower the LR- (e.g. < .10), the greater confidence the test user can have that the person who obtained a score within normal range is, in fact, unaffected.

The LR+ is calculated to be 7.08 indicating that a positive test score can suggest the presence of the disorder, but is insufficient for diagnosis (Dollaghan, 2007). The LR- is .17 indicating that a negative test score suggests that the student does not have the disorder, but is insufficient to rule out the disorder.

Overall, construct validity, including the reference standard, sensitivity and specificity, and likelihood ratios of the PLS-5 Spanish was determined to be insufficient. Even though the likelihood ratios were suggestive of the presence or absence of a language disorder, the sensitivity and specificity measures are considered insufficient because the reference standard against which the PLS-5 Spanish were compared had unacceptable levels of diagnostic accuracy. As a result, the likelihood ratios, sensitivity and specificity measures reported by the PLS-5 are invalid.

Concurrent - Concurrent Validity is the extent to which a test agrees with other valid tests of the same measure (Paul, 2007). According to McCauley & Swisher (1984) concurrent validity can be assessed using indirect estimates involving comparisons amongst other tests designed to measure *similar behaviors*. If both test batteries result in similar scores, the tests

“are assumed to be measuring the same thing” (McCauley & Swisher, 1984, p. 35). Concurrent validity of the PLS-5 Spanish was provided through comparison to the PLS-4 Spanish and CELF Preschool-2 Spanish.

In comparing the PLS-5 Spanish to the PLS-4 Spanish, a sample of 117 children was used, aged birth through 6;11. It is important to note that this does not cover the entire age range of the PLS-5 Spanish (up to 7;11). According to the Examiner’s Manual, adjusted correlation coefficients between the two tests were .69 for AC, .71 for EC, and .81 for TL. However, it is important to consider the psychometric data of the comparison test. Concurrent validity “requires that the comparison test be a measure that is itself valid for a particular purpose” (APA, 1985, as cited in Plante & Vance, 1994). The specificity of the PLS-4 Spanish for AC, EC, and TL was .56, .69, and .63 respectively (Zimmerman, Steiner, & Pond, 2002). According to Plante & Vance (1994), measures below .8 are considered unacceptable and therefore the PLS-4 Spanish cannot be considered a valid measure.

The PLS-5 Spanish was compared to the CELF-P2 Spanish using a sample of 94 children, aged 3 years through 6;11. This comparison test does not include the entire age range of the PLS-5 Spanish. According to the Examiner’s Manual, adjusted correlation coefficients between the PLS-5 Spanish and the CELF-P2 Spanish were .76 for AC, .70 for EC, and .68 for TL. As noted above, it is important to consider the validity of the comparison measure. The normative sample of the CELF-P2 Spanish included 450 monolingual (Spanish) and bilingual (Spanish-English) children, which decreases its validity if it claims to identify language delays/disorders in *bilingual* children.

The concurrent validity is considered unacceptable because neither comparison test included the entire age range of the PLS-5 Spanish and the validity of both comparison tests is considered insufficient based on the standards of validity in the field (Vance & Plante, 1984).

5. RELIABILITY

According to Paul (2007, p. 41), an instrument is reliable if “its measurements are consistent and accurate or near the ‘true’ value.” Reliability may be assessed using different methods, which are discussed below. It is important to note, however, a high degree of reliability alone does not ensure diagnostic accuracy. For example, consider a standard scale in the produce section of a grocery store. Say a consumer put on 3 oranges and they weighed 1 pound. If she weighed the same 3 oranges multiple times, and each time they weighed one pound, the scale would have *test-retest reliability*. If other consumers in the store put the same 3 oranges on the scale and they still weighed 1 pound, the scale would have *inter-examiner reliability*. Now say an official were to put a 1 pound calibrated weight on the scale and it weighed 2 pounds. The scale is not measuring what it purports to measure—it is not valid. Therefore, even if the reliability appears to be sufficient as compared to the standards in the field, if it is

not valid and accurate it is still not appropriate to use in assessment and diagnosis of language disorder.

Test-Retest Reliability – Test-retest reliability “refers to the stability of test scores over time” (McCauley & Swisher, 1984, p. 36). This means that despite the test being administered several times, the results are similar for the same individual. Test-retest reliability was calculated based on data collected on 193 children (ages birth through 7:11) chosen from the normative sample. Reliability coefficients were calculated for AC, EC and TL in 3 age groups (0-2;11, 3-4;11, 5-7;11), yielding 9 coefficients. The reliability coefficients were corrected for the variability of the sample using the variability correction of Allen and Yen (2002, as cited in Zimmerman, Steiner, & Pond, 2012). According to Salvia, Ysseldyke, and Bolt, (2010, as cited in Betz, Eickhoff, and Sullivan, 2013), a correlation coefficient of 0.90 is the minimum to ensure that test scores are stable over a period of time. Considering the corrected coefficients, 3 out of 9 coefficients met this standard. The TL test-retest coefficient was considered acceptable across age ranges, whereas the AC and EC coefficients fell below this standard for all ages. Therefore, the test-retest reliability is considered insufficient. It is also important to note the number of children used for each age group within the sample. The largest sample was for ages 0-0;11 which was 37 children. These samples are too small as compared to the standards in the field, which recommends sample sizes of 100 or more (APA, 1974). If a small sample is used than the norms are likely to be less stable and less representative (McCauley & Swisher, 1984).

Inter-examiner Reliability– Inter-examiner reliability is used to measure the influence of different test scorers or different test administrators on test results (McCauley & Swisher, 1984). It should be noted that the inter-examiner reliability for index measures is often calculated using specially trained examiners. When used in the field, however, the average clinician will likely not have specific training in test administration for that specific test and thus the inter-examiner reliability may be lower in reality. According to the Examiner’s Manual, “all PLS-5 Spanish standardization protocols were scored by trained scorers. Seven scorers were trained on the subjective scoring rules” (Zimmerman, Steiner, & Pond, 2012). Inter-examiner reliability was calculated using two measures. In the first, a sample of 200 tests was randomly selected from the standardization protocols from across the test’s age range. A trained scorer rescored the appropriate subjective items. In the second method, 10% of all protocols were “double scored” by two separate trained scorers. The interscorer reliabilities were .99 for Auditory Comprehension and .99 for Expressive Communication, which meet the standard in the field (Salvia, Ysseldyke, & Bolt, 2010, as cited in Betz, Eickhoff, and Sullivan, 2013). However, as noted above, adequate reliability does not ensure accuracy or appropriateness of the test.

Inter-item Consistency – Inter-item consistency assesses whether “parts of the test are measuring something similar to what is measured by the whole” (Paul, 2007, p. 41). Inter-item consistency for both the normative sample and clinical sample was calculated using the split-half method. This method uses the correlation between scores from two-halves of the test which are administered and scored separately. Coefficients were calculated for AC, EC, and TL. For the normative sample, across the age range of the test, coefficients for AC ranged from .8 to .94; coefficients from 7 out of 18 age ranges did not meet the minimum standard of .9 as recommended by Salvia, Ysseldyke, and Bolt, (2010, as cited in Betz, Eickhoff, and Sullivan, 2013). Coefficients for EC ranged from .80-.95; coefficients from 7 out of 18 age ranges did not meet the minimum standard of .9. For TL, coefficients ranged from .87-.97; coefficients from 5 out of 18 age ranges did not meet the minimum standard of .9. Therefore, the overall inter-item consistency is considered insufficient.

For children with receptive and expressive language disorders, across AC, EC, and TL, coefficients ranged from .98 to .99, suggesting that the inter-item consistency is acceptable for this population. It should be noted, however, that the children used for this study ranged in age from 1;6 to 7;11, and therefore do not represent the entire age range of the PLS-5 Spanish.

6. STANDARD ERROR OF MEASUREMENT

According to Betz, Eickhoff, and Sullivan (2013, p.135), the Standard Error of Measurement (SEM) and the related Confidence Intervals (CI), “indicate the degree of confidence that the child’s true score on a test is represented by the actual score the child received.” They yield a range of scores around the child’s score, which suggests the range in which their “true” score falls. Children’s performance on standardized assessments may vary based on their mood, health, and motivation. For example, a child may be tested one day and receive a standard score of 90. Say he was tested a second time and he was promised a reward for performing well; he may receive a score of 96. If he were to be tested a third time, he may not be feeling well on that day, and thus receive a score of 84. As children are not able to be assessed multiple times to acquire their “true” score, the SEM and CIs are calculated to account for variability that is inherent in individuals. Current assessment guidelines in New York City require that scores be presented within confidence intervals whose size is determined by the reliability of the test. This is done to better describe the student’s abilities and to acknowledge the limitations of standardized test scores (NYCDOE CSE SOPM 2008).

The PLS-5 provides CIs at the 90% and 95% confidence levels for AC, EC, and TL. The clinician chooses a confidence level (usually 90% or 95%) at which to calculate the CI. Although a larger range is yielded with a higher confidence level, the clinician can be more *confident* that the child’s ‘true’ score falls within that range. A lower confidence level will produce a smaller range of scores but the clinician will be less confident that the child’s true

score falls within that range. The wide range of scores necessary to achieve a high level of confidence, often covering two or more standard deviations, demonstrates how little information is gained by administration of a standardized test. For example, if a child aged 2;6-2;11 achieved a raw score of 22 on the EC scale, this would convert to a standard score of 73. As this score falls below 1.5 SD from the mean, they may be falsely identified as having an expressive language disorder and may be given services unnecessarily, possibly leading to serious long term consequences on the child's development and achievement. However, considering the CI, at the 95% confidence interval, the child's true score range would be between 67 and 84. Thus, all the clinician can determine from administration of the PLS-5 Spanish is that the child's true language ability (according to the test) ranges from moderately impaired to within normal limits. The wide range of the necessary CI means the scores from the PLS-5 Spanish are of little use. Even if the test were valid, reliable and accurate, when the confidence interval is applied little to no information is gained regarding the diagnostic status of the child.

7. **BIAS:**

According to Crowley (2010), IDEA 2004 regulations stress that assessment instruments must not only be "valid and reliable" but also free of "discriminat[ion] on a racial or cultural basis." In addition to being an invalid measure of language ability, the PLS-5 Spanish contains many inherent biases against culturally and linguistically diverse children.

Linguistic Bias

Bilingual Speakers

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. Many students who will be administered the PLS-5 Spanish may be students who are learning English as a second language in school. Consider, for example, a child from a Spanish speaking family who enters kindergarten. Although they only spoke Spanish until they started school at 5 years old, they may refuse to speak it once they start learning English in school. Thus, they may be referred for an evaluation in English, a language they have only been learning for about one year. Although Spanish was their first language, after a year of little to no practice using it, they may be experiencing subtractive bilingualism. This occurs when "acquisition of the majority language comes at the cost of loss of the native language" (Paradis, Genesee, & Crago, 2011, p. 49). As a child gains skills in their second language and ceases using their first language, their proficiency in the first language declines. Since language tests are cognitively demanding and require significant amounts of metalinguistic and academic language skills and vocabulary, a typically developing child experiencing subtractive bilingualism may show depressed skills in both languages. According to 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). Only a clinician with significant training and experience evaluating bilingual children and using other assessment tools (i.e. not a norm-referenced test) would be able to pick up on why a bilingual child would have delayed skills in both languages. As a dual language assessment, the PLS-5 Spanish attempts to compensate for children who are learning English as a second language by allowing for test administration in either language. Scores are calculated considering a correct response in either language. However, biases such as those mentioned above, are still present in the test.

On the PLS-5 Spanish, students learning English as a second language may be falsely identified as having a language disorder on tasks such as EC28 (*Uses past tense forms*). The examiner shows the child a pair of pictures, one depicting an action currently taking place, and one that is completed. The clinician probes the child to use the past tense to describe the picture (e.g. “The ice cream...”). According to Paul (2007), children learning English as a second language often omit the -ed ending to mark the past tense. Thus, they may falsely be identified as having a language disorder.

Dialectal Variations

A child’s performance on the PLS-5 Spanish may also be affected by the dialects of Spanish and English that are spoken in their homes and communities. The English items are presented in Standard American English (SAE), however, the manual does not provide information regarding the dialect of Spanish that is used. It is important to note that there are many different dialects of Spanish from different regions that vary significantly. In the normative sample alone, 6 different countries of origin are reported: Central America, Cuba, Dominican Republic, Mexico, Puerto Rico, and South America. It can be safely assumed that this test is administered to children who speak even more dialects of Spanish. It is important to consider the issues of the test being administered in a child’s non-native dialect of either Spanish or English. For example, imagine being asked to repeat the following sentence, written in Early Modern English: “Whether 'tis nobler in the mind to suffer The slings and arrows of outrageous fortune Or to take arms against a sea of troubles And by opposing end them” (Shakespeare, 2007). Although the content of the sentence consists of words in English, because of the unfamiliar structure and semantic meaning, it would be difficult for a speaker of SAE to repeat this sentence as compared to a similar sentence in SAE. The same would hold true for being asked to repeat a sentence in a dialect of Spanish that was different from the child’s.

Speakers of dialects other than those used in the PLS-5 Spanish (e.g. African American English [AAE], Patois, regional dialects of Spanish) face a similar challenge when asked to complete tasks such as EC45 (*Repeats Sentences*). The examiner reads a sentence aloud and the child is instructed to repeat the sentence. Although the manual indicates

that the child may repeat the sentence following either Spanish or English presentation, if the child speaks a different dialect of either language, it may be difficult for them to complete this task. Many dialects exist for both Spanish and English and if the child does not speak the “standard” dialect of either, they may have difficulty with this task.

It should be noted that for specific items in the *Protocolo*, notes are given to the examiner to where language variations may be present. For example, for item EC31 (*Uses plurals*), the following is noted, “Do not penalize for language variations, such as consistent aspiration of the /s/” in the *Protocolo* to ensure that children are not penalized for aspects of normal language variation. It is also important to consider that even a Spanish speaking test administrator would need to be aware of potential dialectal variations in the Spanish language. Understanding of such dialectal variations requires a vast knowledge base. For example, in the previously mentioned case of /s/ aspiration, the examiner would need to know that while this is typical in certain word positions, it would be unusual in other positions and perhaps indicative of a phonological process. This is best determined through comparison of the child’s dialect to that of his or her speech community. Thus, test administrators should be trained in dialect issues so they are able to accurately discriminate between a dialectal feature and an error to appropriately score the child’s responses.

Socioeconomic Status Bias

Hart & Risley (1995) found that a child’s vocabulary correlates with his/her family’s socioeconomic status; parents with low SES used fewer words per hour when speaking to their children than parents with professional skills and higher SES. Children from families with a higher SES will likely have larger vocabularies and score better on standardized tests since many items are actually a test of vocabulary or highly dependent on vocabulary. A child from a lower SES background may be falsely identified as having a language disorder on standardized language tests due to a smaller vocabulary than his higher SES peers. Certain items on the PLS-5 Spanish are biased against children from low SES backgrounds because they require a diverse vocabulary such as EC56 (*Uses synonyms*). In this task, the child is given a word (e.g. beautiful), and are asked to provide another word that has the same meaning. A child from a low SES home who is not exposed to a diverse vocabulary may have difficulty with this task.

Prior Knowledge/Experience

A child’s performance on PLS-5 (Spanish) may also be affected by their prior knowledge and experiences. For example, many questions in the PLS-5 Spanish require the child to be well versed in playing with toys and manipulating books and print items, including interacting with and manipulating a toy bear (e.g. “*Don Osito tiene sueño, acuestelo a dormir/Don Osito is sleepy, make him go to sleep.*”) According to Peña and Quinn (1997), some infants are not

exposed to books, print, take-apart toys, or puzzles. If a child did not have previous experience with toys such as this, they may not perform as well on this task as their peers and may falsely be identified as having a language disorder. Further, in item AC50 (*Identifies a picture that does not belong*), the child is required to identify one item from a field of four that doesn't belong (e.g. *screwdriver, spoon, wrench, hammer*). If the child had not previously been exposed to these items, they may have difficulty with these questions, and may be falsely identified as having a language disorder. Also, some children who have not had exposure to print items may not realize that a caricatured illustration of an object is supposed to represent a real-life object. From birth, children from mainstream higher SES backgrounds are consistently instructed and reminded that a yellow and orange, flat shape in a book represents a much larger, moving, usually brown or green, loud animal we call a duck. Finally, item number AC63 (*Understands time concepts*), the child is required to point to a picture that visually represents a specific season. If the child is from a region where they do not experience all 4 seasons, they not respond accurately to this question.

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 & Quinn (1997), 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. The PLS-5 Spanish contains various testing formats, many of which are dependent upon prior knowledge and experiences. For example, item EC24 (*Identifies photographs of familiar objects*) is a strictly labeling task. A child without previous exposure to this type of task may not perform well. Further, item numbers AC56-60 and EC49-51 require the child to complete a "known information" task. On these items, the child listens to a short story read aloud by the clinician and are then asked to respond to comprehension questions regarding the story and retell the narrative. As the examiner was present for the presentation of the task, it can be assumed that they are familiar with the story. Thus, it is a "known information question." A child who is not accustomed to an adult asking questions to which they already know the answer may fail to respond appropriately.

Further, a child's performance on the test may be affected by their prior exposure to books. According to Peña and Quinn (1997), some infants are not exposed to books, print, take-apart toys, or puzzles. The PLS-5 Spanish requires children to attend to the test book for the length of the assessment, which may be challenging for a child who has not had prior exposure with structured tasks. He or she must also realize that pictures and symbols have meaning and attend to them (print awareness); this is not an innate skill but a learned one. In addition, lack of access to books and print materials results in a lack of familiarity with and delayed pre-literacy skills including letter knowledge and phonological awareness. For example, item AC51 requires the child to identify the initial sound of a word. If they have not had the opportunity to gain pre-literacy skills, they may have difficulty.

Cultural Bias

According to Peña & Quinn (1997), tasks on language assessments often do not take into account variations in socialization practices. For example, the child's response to the type of questions that are asked (e.g. known information questions, labeling), the manner in which they are asked, and how the child is required to interact with the examiner during testing, may be affected by the child's cultural experiences and practices. During test administration, children are expected to interact with strangers. In middle class mainstream American culture, young children are expected to converse with unfamiliar adults as well as ask questions. In other cultures, however, it is customary for a child to not speak until spoken to. When he does speak, the child often will speak as little as possible or only to do what he is told. 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. Further, cultural biases are present in items EC49-51, which require the child to retell a narrative that is first orally dictated by the clinician. The child is penalized for not including elements of the story such as a proper introduction, and a logical conclusion. It is important to note, however, that styles of storytelling vary greatly between different cultures (Paul, 2007). Therefore, if the child does not use the mainstream storytelling style, they may be inappropriately penalized.

Attention and Memory

Significant attention is required during administration of standardized tests. If the child is not motivated by the test's content, or they exhibit a lack of attention or disinterest, they will not perform at their true capacity on this assessment. Further, fatigue may affect performance on later items in the test's administration. Even a child without an attention deficit may not be used to sitting in a chair looking at a picture book for an hour. A child that has never been in preschool and has spent most of his days in an unstructured environment and playing with peers and siblings may find it very challenging to sit in front of a book for extended periods of time.

According to the Manual de administración y puntuación (Administration and Scoring manual), specific modifications may be made during test administration in order to maximize a child's performance. For example, with the exception of the *Repeats sentences* task, items may be repeated *once* if the child requests repetition, if the administration was interrupted, or if the clinician determines it to be appropriate (p. 29). Further, the manual states that a child is allowed to take breaks during test administration if the clinician determines that it is necessary. It is important to allow for these modifications if necessary to ensure the child is performing as close as possible to their true capacity.

Short term memory could also falsely indicate a speech and/or language disorder. Many of the test items require the child to hold several items in short term memory at once, then compare/analyze them and come up with a right answer. In item AC61 (*Understands*

qualitative concepts), the child is asked to point to a picture representing a qualitative statement (e.g. *Which girl has the fewest balloons?*). The child must remember the target, determine the meaning of the concept, and indicate their response. A child with limited short-term memory may perform poorly on standardized assessments due to the demands of the tasks. However, he may not need speech and language therapy but rather techniques and strategies to compensate for short-term or auditory memory deficits. Further, as the sample population did not include children from this population, results of this assessment are invalid for children with attention deficits.

Motor/Sensory Impairments

In order for a child to participate in administration of this assessment, they must have a degree of fine motor and sensory (e.g. visual, auditory) abilities. If a child has deficits in any of these domains, their performance will be compromised. For example, for a child with vision deficits, if they are not using proper accommodations, they may not be able to fully see the test stimuli, and thus their performance may not reflect their true abilities. A child with motor deficits, such as a child with typical language development but living with cerebral palsy (CP), may find it much more frustrating and tiring to be pointing to/attending to pictures for an extended period of time than a typically developing non-disabled child. The child with CP may not perform at his highest capacity due to his motor impairments and would produce a lower score than he or she is actually capable of achieving. Further, as the sample population did not include children from this population, results of this assessment are invalid for children with motor and sensory impairments.

8. SPECIAL ALERTS/COMMENTS

The PLS-5 Spanish is designed to determine the presence and severity of a receptive, expressive, or receptive-expressive language delay or disorder in monolingual Spanish speakers or bilingual Spanish-English speaking children from birth to 7 years 11 months. The test may also help to identify a child’s strengths and weakness to determine appropriate intervention, and may also identify the “contexts in which a bilingual Spanish-English speaking child uses one language versus the other” (p. 4). Despite the PLS-5 Spanish’s attempt to design a comprehensive language battery, results obtained from administration are not valid due to lack of information as to how tasks and items were deemed appropriate, and an insufficient reference standard. It is also important to note that although the discriminant accuracy was considered “fair,” the PLS-5 Spanish does not discriminate typically developing children from children with a language disorder. Instead, it discriminates children who scored below an arbitrary cutoff score on the PLS-4 Spanish. In addition, the test-retest and inter-item reliability measures did not meet the standard in the field and thus are unacceptable.

According to the Manual de administración y puntuación, “For an overall evaluation of a child’s language ability, the results of the PLS-5 Spanish should be supplemented with a

complete family and academic history, primary caregiver interview, analysis of spontaneous language sample, classroom behavioral observations, observations of peer interactions, evaluations of pragmatic and interpersonal communication abilities, and the results of other linguistic and metalinguistic abilities tests” (p. 7). One may question why the PLS-5 should be administered if the manual itself states that it should be supplemented with various other measures that require clinical judgment and essentially constitute an appropriate speech and language evaluation on their own. Why spend an hour administering the standardized test at all? Although as a dual language assessment, the PLS-5 Spanish attempts to compensate for second language acquisition issues by probing the child in both Spanish and English, many biases are still inherent in the test. Due to cultural and linguistic biases (e.g. exposure to books, cultural labeling practices, communication with strangers, responses to known questions, etc.), and assumptions about past knowledge and experiences, this test should only be used to probe for information and not to identify a disorder or disability. Therefore, scores should not be calculated and used as the sole determinant of classification or referral to special education services.

REFERENCES

- 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]. Available from www.asha.org/policy.
- 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.
- Dollaghan, C. (2007). *The handbook for evidence-based practice in communication disorders*. Baltimore, MD: Paul H. Brooks Publishing Co.
- Dollaghan, C., & Horner, E. A. (2011). Bilingual language assessment: a meta-analysis of diagnostic accuracy. *Journal of Speech, Language, and Hearing Research, 54*, 1077-1088.
- Guadagnoli, E. and Velicer, W. F. (1988). Relation to sample size to the stability of component patterns. *Psychological Bulletin, 103*, 2, 265-275. doi: 10.1037/0033-2909.103.2.265
- Hart, B & Risley, T.R. (1995) *Meaningful Differences in the Everyday Experience of Young American Children*. Baltimore: Paul Brookes.
- McCauley, R. J. & Swisher, L. (1984). Psychometric review of language and articulation tests for preschool children. *Journal of Speech and Hearing Disorders, 49*(1), 34-42.
- New York City Department of Education (2009). *Standard operating procedures manual: The referral, evaluation, and placement of school-age students with disabilities*. Retrieved From <http://schools.nyc.gov/nr/rdonlyres/5f3a5562-563c-4870-871f-bb9156eee60b/0/03062009sopm.pdf>.
- Paul, R. (2007). *Language disorders from infancy through adolescence (3rd ed.)*. St. Louis, MO: Mosby Elsevier.

-
- 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.
- Paradis, J., Genesee, F., & Crago, M. B. (2011). *Dual language development & disorders: A handbook on bilingualism & second language learning (2nd ed.)*. Baltimore, MD: Paul H. Brookes.
- 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.
- Plante, E. & Vance, R. (1994). Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools*, 25, 15-24.
- Shakespeare, W. (2007). *Hamlet*. David Scott Kastan and Jeff Dolven (eds.). New York, NY: Barnes & Noble.
- Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2002). *Preschool Language Scales (4th ed.)*, (Spanish) (PLS-4 Spanish). San Antonio, TX: Psychological Corporation.
- Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2012). *Preschool Language Scales (5th ed.)*, (Spanish) (PLS-5 Spanish). Bloomington, MN: Pearson.