1-2021

Assessing Dual Language Learners of Spanish and English: Development of the QUILS: ES

Jill de Villiers
Smith College, jdevilli@smith.edu

Aquiles Iglesias
University of Delaware

Roberta Golinkoff
University of Delaware

Kathy Hirsh-Pasek
Temple University

Mary Sweig Wilson
University of Vermont

See next page for additional authors

Recommended Citation

Authors

Jill de Villiers, Aquiles Iglesias, Roberta Golinkoff, Kathy Hirsh-Pasek, Mary Sweig Wilson, and Ratna Nandakumar

Assessing Dual Language Learners of Spanish and English:

Development of the QUILS: ES.

Abstract

Introduction and Objectives: Developing a language screener for Dual Language Learners presents numerous challenges. We discuss possible solutions for theoretical and methodological problems often encountered in the development of such a test and illustrate possible solutions using a newly developed language screener for Dual Language Learners.

Materials and Methods: The process for developing, validating and norming the screener is also offered as a potential model for the development of other assessments for Dual Language Learners throughout the world. The twelve types of subtests are described with in the areas of Vocabulary, Syntax, and Process.

Results and Conclusions: Results from the Tryout and Norming phase on 362 Dual Language Learners aged 3 to 5;11 years are presented, together with the results of item selection via IRT, validity, and reliability testing. The advantage of using Best Scores is highlighted as a useful measure that helps identify children who are at risk of encountering language difficulties that will impact their academic success. Importantly, knowledge is found to be distributed across the languages.

Keywords: Screener, Process, Distributed knowledge, Best scores, Pre-school, Dual Language Learners

1.0 The difficulties of Dual Language (English/Spanish) Screening

The general need for a language screener for preschool children is based on research findings that proper instruction and intervention are likely to be more effective in younger children, and overlooked problems can have long-term consequences for children's success in academics and life (Glogowska, Roulstone, Enderby, & Peters, 2000; Law, Kot, & Barnett, 1999; Ramey & Ramey, 1998; Roberts & Kaiser, 2015; Wake et al., 2011). Even by 3 years of age, the effects of lower language competence are evident: for example, children with poor communication skills are less sought after as conversational partners and more likely to be ignored or excluded by their peers (Rice, 1993). These children then fall further behind socially and tend to develop poor self-esteem as they advance through childhood (e.g., Conti-Ramsden & Botting, 2004; Craig, 1993; Jerome, Fujiki, Brinton, & James, 2002; Lindsay & Dockrell, 2000). Even short-term gains in language ability can enhance social relationships and mitigate the negative impact of language delay on behavioral, social, and emotional development (Olswang, Rodriguez, & Timler, 1998; Paul, 1996; Robertson & Weismer, 1999). Although several screeners are available for monolingual English speakers in the US, Dual Language Learners have been largely neglected as a group, and are often mis-identified as having language problems based on testing only one language (Peña, Gillam, Bedore, & Bohman, 2011; Gillam, Peña, Bedore, Bohman, & Mendez-Perez, 2013). It is well-documented that there is over-identification of English language learners (ELLs) as having language delays (Artiles, Rueda, Salazar, & Higareda, 2005), but under-identification is also a problem, where SLPs do not trust that a test is adequate to assess a language in the process of being learned (ref). A screener is necessary to assess whether a bilingual child has a language difference or potentially a language disorder.

In the US, there is a critical need to develop linguistically appropriate and valid assessment tools for children growing up in homes where they are exposed to English and Spanish (Barrueco Barrueco, Lopez, Ong, & Lozano, 2012). Some children are primarily exposed to Spanish at home, but a large proportion will be raised in an environment in which both languages are used (Rojas, Iglesias, Bunta, Miller, Goldenberg, & Reese, 2016). Assessing the progress of dual language learning children is difficult for two reasons. First, children are arrayed along a continuum of bilingualism, from knowing mostly Spanish to knowing mostly English, with every alternative in between, thus making it hard to find norms in either language that treat all children fairly. Second, what Dual Language Learners know in each language remains obscure. It has been known for many years that vocabulary is distributed across the languages of children exposed to two languages, and not just at the very start, where children might resist having two words for one referent (Pearson & Fernandez, 1994; Pearson, 1998; Core, Hoff, Rumiche, & Señor, 2013; Mancilla-Martinez, & Vagh, 2013). There is evidence even up to college age that students have different vocabulary items in each language, with many words that do not have corresponding lexical items in the other language (Dong, Gui, & MacWhinney, 2005). What children store is distributed across the two languages. One purpose of the present report is to demonstrate that it is not just vocabulary that is distributed in young Dual Language Learners, but also syntactic development, and even the ease with which children learn new forms and words, or the process of learning. A dual-language learning child must be assessed in both of their languages to understand whether they at risk of a language delay or disorder. Thus, the QUILS: ES assesses both languages. and it also provides a metric to evaluate the child's overall langage competence.

The test-development process reported here might also serve as a schema for others looking to create dual language screeners for different language combinations, either for the US or other

countries with a significant population of children learning two languages at an early age. The principles of test construction, choice of measures and methods of sampling, reliability and validity, should transcend the particular languages involved.

2.0 Challenges and Solutions

There are specific challenges in developing an adequate language screener for Dual Language Learners, and we highlight five below, together with the solutions we have devised from the process of developing a new screener, the Quick Interactive Language Screener: English-Spanish (QUILS: ES).

2.1 First Challenge and Solution: Persistent language problems are hard to identify early.

Some children are identified as "late talkers" at age 2 or 3 years based on their low language production. However, research suggests many of these children go on to develop language within the typical range (Dollaghan, 2013; Leonard, 2014; Rescorla, 2000). Language comprehension may provide a better predictor of which children will continue to have problems (Leonard, 2014; Thal & Bates, 1988) and require intervention. Parents and teachers can spot a child who is not speaking, but not all children who are late talkers require intervention; some children who appear to have language delays can comprehend language. Comprehension measures are at the cutting edge of children's linguistic capability (Hirsh-Pasek & Golinkoff, 1996; Seidl, Hollich, & Jusczyk, 2003; Weisleder & Fernald, 2009; Friend, Smolak, Liu, Poulin-Dubois, & Zesiger, 2018). Thus, it is essential to probe children's language comprehension because it may serve as a more sensitive measure of language skill than children's language production.

Relying on language production (what children say) can be problematic because young children may have limited expressive capacities and are often reluctant to demonstrate their full expressive potential in an assessment context with an unfamiliar examiner (Brown, 1973). With comprehension measures, the burden of communication with an examiner the child does not know can be reduced. In addition, the minimal response demands of comprehension—in the case of the QUILS: ES, touching the correct picture on a screen—are much lower than those of production and do not require examiners to make judgments in the face of children's early, nonstandard pronunciations. The QUILS: ES invites children to play a game in which there are brightly colored pictures and animated scenes. It circumvents the problem of coaxing children to speak or to answer questions posed by a stranger. Children engage with the touchscreen computer or tablet in a way that is fun and yet reveals their language skill. The QUILS-ES screener presents items to children on a touchscreen, and the items are narrated automatically in the appropriate language. After a few training items that teach the child how to touch the screen, the test unfolds with a few interspersed animated gifs that congratulate the child on their efforts and encourage the child to keep going.

These advantages of a comprehension instrument accrue to young children whether they are dual language learning or not. All children picked out as being at risk by such a screening tool will also need assessment of their production skills in a more thorough clinical workup.

2.2 Second Challenge and Solution: Assessments must examine the ability to learn as well as the products of learning.

Results from research on monolingual children show that oral language skills at age 3, including syntax as well as vocabulary, contribute to reading outcomes in first grade regardless of socioeconomic status (SES; NICHD ECCRN, 2005). Likewise, vocabulary and syntactic

ability in prekindergarten are unique predictors of language variability in third grade (LARRC, 2015; Pace, Alper, Burchinal, Golinkoff, & Hirsh-Pasek, 2019). However, assessments have not incorporated more recent research that underscores the importance of evaluating the *processes* by which children learn language in addition to the products of language learning: syntax and vocabulary. That is, existing screeners and assessments measure what the child *knows* with little attention to *how* the child learns (Hirsh-Pasek, Kochanoff, Newcombe, & de Villiers, (2005). Process measures that have become popular include dynamic assessment (Peña) , and response to intervention ( ). In the current context, we assess the process of learning in a single test, not over time, by designing items that test how adequately children can learn new word meanings (a process called fast mapping), by exploiting the syntactic contexts in which new words appear, and to extend words to new contexts—all of which jointly contribute to children's skills as language learners (Fisher, 1996; Golinkoff, Jacquet, Hirsh-Pasek, & Nandakumar, 1996; Seymour, Roeper, & de Villiers, 2004). In addition to assessing vocabulary and syntax, the QUILS: ES focuses on the process – in both languages - by which children learn language; that is, their proficiency at learning new vocabulary items and generalizing syntactic information in new contexts. For example, a child may have fewer vocabulary words than peers (e.g., perhaps due to limited exposure to language models) but be in line with his or her age group in terms of vocabulary acquisition skills, such as quickly acquiring a new word after a limited number of exposures. Children who have low scores in acquired vocabulary and syntax for example, but prove capable at the process of learning new items and structures, have the machinery to learn language and perhaps only lack exposure to more high-quality language interactions. Those who are poor at language learning *and* have low levels of acquired vocabulary and syntax are more

likely to need further assessment to determine eligibility or a remediation plan to bolster their existing language skills.

Our solution consisted of creating two distinct, although parallel sections (English and Spanish) that assessed product (vocabulary and syntax that child knows) and process (child's ability to learn new vocabulary and syntactic structures). Each section (English or Spanish) of the QUILS: ES is arranged according to the three areas described below: Vocabulary, Syntax, and Process. Each area measures different types of language knowledge (e.g., prepositions) and the specific items are not the same in each of the two sections (e.g., "la muñeca *está* arriba del regalo" "the girls are *between* the motorcycles"). The screener uses animations to provide a more precise depiction of an event sequence that may be challenging for young children to glean from still pictures of actions or event sequences. Table 1 shows the subtest types under each area. The Vocabulary Area looks at words the children have already learned, that is, children's existing knowledge. The vocabulary items are grouped into 4 types: Nouns, Verbs, Prepositions, and Conjunctions. The Syntax Area examines children's knowledge of sentence structure. The items in the Syntax Area are grouped into four types: Wh-Questions, Past Tense, Prepositional Phrases, and Embedded Clauses. The Process Area items require children to quickly infer the meaning of a new words from the sentences plus pictures, learning new Nouns, Verbs and Adjectives. It also includes a subtest of extending a syntactic structure to a newly learned verb, where the child hears a new verb in an active structure and then has to understand it in the passive voice[1]. The process items, embedded throughout the assessment, are grouped into 4 types: Verb Learning, Converting Active to Passive, Noun Learning, and Adjective Learning. The final QUILS:ES has 45 items in each language. These are shown in the Appendix.

---

[1]

Table 1: Contents of the QUILS:ES screener

Areas and Subtest Types		
Vocabulary	**Syntax**	**Process**
Nouns	Wh-Questions	Noun Learning
Verbs	Past Auxiliary and Copula	Adjective Learning
Prepositions	Prepositional Phrases	Converting Active to Passive
Conjunctions	Embedded Complements	Verb Learning

Two illustrations are provided in Figure 1. These are stills of the final scene, but there is animation preceding this to allow the child to see the events unfold in time.

Figure 1 An illustration from clausal connectives (CC) in English and Spanish.
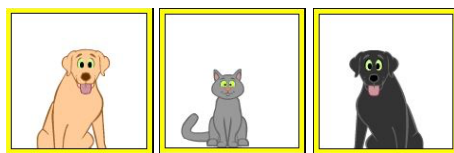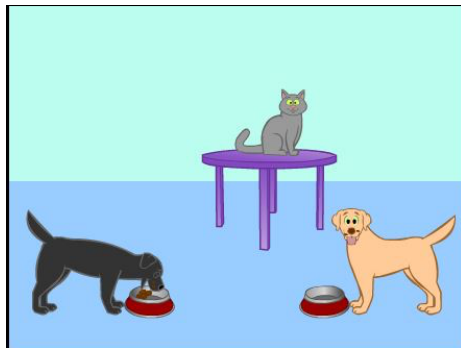
Question: Who ate the food before the cat jumped on the table?
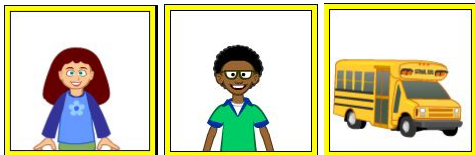
Question : ¿Quién se deslizó por el tobogán despues que llego el autobus?

"Who slid down the slide before the bus came?"

2.3 Third Challenge and Solution: Assessments must be applicable to the population assessed

The procedure by which we arrived at the final selection of items for QUILS: ES happened in multiple stages. All of the items on the QUILS: ES were chosen by experts in the science of child language development and are based on the most current research in language acquisition. During item development and creation, native English and Spanish-speaking experts evaluated each item, ensuring that the items 1) were feasible for both English-monolingual, Spanish monolingual and Spanish–English bilingual children, and 2) did not discriminate between children who spoke different dialects of English or Spanish. All items were chosen to be adaptable to English or Spanish, rather than relying on simple translation, and only words that were neutral across Spanish dialects were considered for inclusion in the screener. In addition, the use of obvious cognates, or words that overlap in form and meaning across languages such as the English *cafeteria* and Spanish *caféteria,* were avoided. This design prevents a speaker of Spanish from scoring correctly on an English item because of his or her Spanish knowledge rather than English

knowledge of the word. *Foils* (i.e., the incorrect alternative answers) all represent choices children might plausibly make if they were guessing or had a false idea about the meaning of the word or sentence. These ideas were grounded in research studies wherever possible (e.g., Golinkoff, Bailey, & Wenger, 1992).

2.3.1 Fairness across Dialects

The QUILS: ES was designed with linguistic and cultural fairness in mind by selecting items through careful testing to be culturally and dialectally neutral in both languages. The development team was attentive to racial, ethnic, and cultural differences. For example, speakers of African American English might be tested, and children might be speaking any of several varieties of Spanish. Thus, all items included in the QUILS: ES had to contain words or linguistic structures that would not be biased against speakers of African American English or different Spanish dialects. Furthermore, the characters portrayed in the QUILS: ES show a variety of ages, races, genders, and disability status.

2.3.2 Multi-step Process to Match Item Levels across Sections

To find appropriate items that would allow matching level items across the English and Spanish sections, the QUILS: ES development process occurred in four main phases over 5 years: 1) Item Development, 2) First Item Tryout, 3) Second Item Tryout, and 4) Creation of the Final Version of the QUILS: ES.

Pilot testing was completed prior to First Item Tryout. All Spanish items were tested with a sample of monolingual Spanish children recruited from a preschool program in a Head Start program in Springfield, Massachusetts. Children were from families of low-income migrant

workers from Mexico and Guatemala. The purpose of testing monolingual Spanish children was to determine whether the test items were appropriate for the targeted age range and showed developmental trends. A sample of 27 children, ages 3- ($n = 8$), 4- ($n = 10$), and 5- ($n = 9$) year-olds were tested. Results revealed that the Spanish items captured growth well in linguistic ability of monolingual Spanish speakers over this age range. After ensuring that the Spanish version of the assessment was appropriate for monolingual Spanish speakers, First Item Tryout of the bilingual version was carried out.

Once the items had been created and piloted individually in laboratory settings, the development team used all items that withstood their scrutiny for First Item Tryout. Following conventional evidence-based practice in psychometrics (Schmeiser & Welch, 2006), the development team tried out twice the number of items to appear in the final version of the QUILS: ES. For First Item Tryout, we began with 96 English items and their equivalents in Spanish. Given the length of each test, it was not feasible to test a Dual Language Learner on each version of an item, but there was no way in advance to assess which item sets would be equivalent in English and Spanish. Our solution was to divide the 96 items in each language into set A (48 items) and set B (48 items). A given child received either Spanish A and English B, or Spanish B and English A, that is, 96 items each. In this way, each child received equivalent items but never the same item in English and Spanish. The two tests were given in counterbalanced order. All parents completed the Parent Questionnaire and only children who scored between 1.5 (Mostly English) and 4.5 (Mostly Spanish) took the test. Seventy-six children aged 3 to 6 in day care centers throughout the North East participated in this phase. Children were randomly assigned to receive Form A or B.

First Item Tryouts on the bilingual test guided our assignment of items to each language. By examining performance on items against general child ability level across all of the items, we assessed whether an item behaved well or not. The rule was that an item "behaved well" if the more able children passed it, and the less able children failed it. We examined each item to see if the children who passed it had a total score that exceeded the total score of the children who chose one of the foils. By this means, we selected the items with the best discrimination between ability levels in each language, and chose which items were more successfully discriminating in English than Spanish or vice versa.

After this process was complete, a new 96-item version of the screener was developed, with 48 items chosen to be presented in Spanish and 48 items chosen to be presented in English, balancing the numbers across subtests with 16 in each area (Vocabulary, Syntax, and Process), and 4 in each type within the areas. Again, we counterbalanced the order of the tests. These two versions were administered to children in preschools, child care centers, and Head Start programs in Massachusetts, Pennsylvania, Delaware, Florida, and Nebraska. There were a total of 578 children tested in the Second Item Tryout, balanced by gender. The majority of the children tested were from low-SES families (84.8%), and the remaining children were from mid-SES families (14.3%) with 5 being unreported.

About a third of the children tested during the bilingual Second Item Tryout also were randomly assigned to receive one validity or reliability measure: 49 were tested on the Preschool Language Scale, PLS-5 (Zimmer. Steimer, & Pond, 2012), 48 on the Bilingual English Spanish Oral Screener, the BESOS (Peña, Bedore, Gutierrez-Vlellen, Iglesias, & Goldstein, unpublished; Lugo-Neris, Peña, Bedore, & Gillam, 2015Wrong font), and 51 on the QUILS:ES retest. Both the PLS-5 and the BESOS evaluate Spanish as well as English, but in different ways than the

QUILS:ES , detailed below. An additional 20 children received the English version of the Peabody Picture Vocabulary test, the PPVT (Dunn & Dunn, 1997).

After completion of the Second Item Tryout, the data were reduced to include only children who received all the items on both versions of the test. Using the responses from these 446 children, Rasch analyses (See also Tucci, Plante, Vance, & Ogilvie, 2019) were conducted to remove problematic items that were either redundant or non-discriminating. The final QUILS: ES then was created using the best 45 items in each language.

2.3.3 Inclusion Criteria for the Normative Sample

The normative sample for the QUILS: ES included children 3 (3;0) through 5 (5;11) years old with no reported visual or hearing difficulties who were screened in their child care centers, preschools, kindergartens, and Head Start programs in Massachusetts, Pennsylvania, Delaware, Florida, and Nebraska. Children who spoke a language other than English or Spanish were not included in the sample. A Language Questionnaire completed by parents (ref) or school-supplied information was used to determine the degree to which English or Spanish were used. Since the normative sample was designed to be representative of dual language learning Spanish-English children in this age range in the United States, it likely includes some children who had language disorders.

Table 2 shows the final normative sample for the QUILS: ES. For the children in the norming sample, information on socioeconomic status (SES) was provided either in the form of mothers' self-reported educational attainment or by enrollment in a low-income child care center. The majority of the children tested were from low SES families (79.4%), and 20.6% of the children were from mid-SES families. (see Table 2). The percentage of mid-SES families approximates the percentage reported in the 2014 U.S. census data for Hispanic females. A more

precise determination is difficult to achieve. Hispanic females with children under 18 having an education level of an associate's degree and above was 26.1% in 2015 (NCES 2015). However, that figure includes women who achieve a degree later in life. If one looks at rate of completion of bachelor's degrees or higher among Hispanic females in the years from 2006 to 2016 the rate is between 12.9 and 16.6% (U. S. Census Bureau, 2016).

Table 2

Composition of the norming sample for the QUILS: ES

	Final norming sample
Total N	362
Age	
3-year-olds: n (%)	69 (26.3)
4-year-olds: n (%)	159 (43.9)
5-year-olds: n (%)	134 (37.0)
Mean age (years): M (SD)	4;8 (0;9)
Gender	
Male: n (%)	177 (48.9)
Female: n (%)	184 (50.8)
No Reported: n (%)	1 (.2)
SES	

Low: *n* (%)	288 (79.5)
Mid: *n* (%)	74 (20.4)

Key: SES, socioeconomic status; *SD*, standard deviation.

Demographic data for race were available for 66.6% of the final bilingual sample: 55.8% were White, 6.6% were Black/African American, 1.4% were multiracial, 0% were Asian, and 1.9% were other races. Additionally, 82% of parents reported whether or not their child was of Hispanic origin; of those who reported on it, 91.2% self-identified as being of Hispanic origin.

2.4 Fourth Challenge and Solution: Knowledge is Distributed across Languages.

A crucial decision in the design of the new screener for Spanish-English Dual Language Learners was to assess both languages in an equivalent way, so as to assess what a child knew in each language, and also overall. Our approach to capturing the child's overall language uses their *best score* in each of the language areas assessed, and compares their performance to other Dual Language Learners. It would not be appropriate to compare these children's language skill to monolingual English or monolingual Spanish speakers who have only heard a single language. Therefore, screening bilingual children in both of their languages, and using their best score provides us with information about whether children are developing language at an appropriate rate for their age.

Why are the Best Scores important for assessment of dual language learners? First, because they make it possible to develop peer group comparisons for children who vary in whether they are stronger in English than Spanish or vice versa, namely, across the broad

continuum of types of dual language learner. Second, because Best Scores consider that a child may know one feature in one language – let's say negation - and another feature in another language, hence be disadvantaged if only one language is assessed. With Best scores, we see whether they have controlled that language feature generally. Third, the point here is not to emphasize how strong the skills are overall, despite the word Best Scores. A child whose Best Scores lie outside the range of his peers - even peers along this varied continuum - reveals a deficit that is of clinical concern, because he does not show understanding in either language.

2.4.1 **Distributed Knowledge**

The performance across the various subtest types provides useful information about what a given child knows already, though based on a very small sample of items. Nevertheless, for our purposes the patterns of responses reveal the important fact that a child's knowledge is distributed across the languages. As Figure 1 reveals, these two sample children show quite different profiles of which subtests they find easy and hard in Spanish versus English. It is not just knowledge of particular lexical items that is distributed in a Dual Language Learner, but also syntax and process.

Figure 2 The profiles of two different children across the subtests of QUILS:ES. The Y axis refers to the proportion correct, and the abbreviations on the X-axis refer to the subtests in Table 1.

Example 1

Example 2
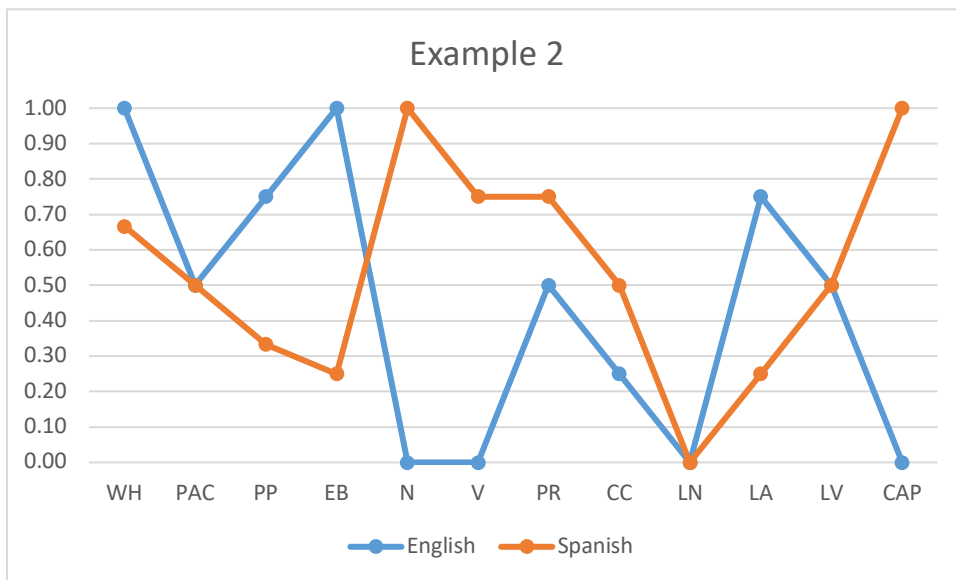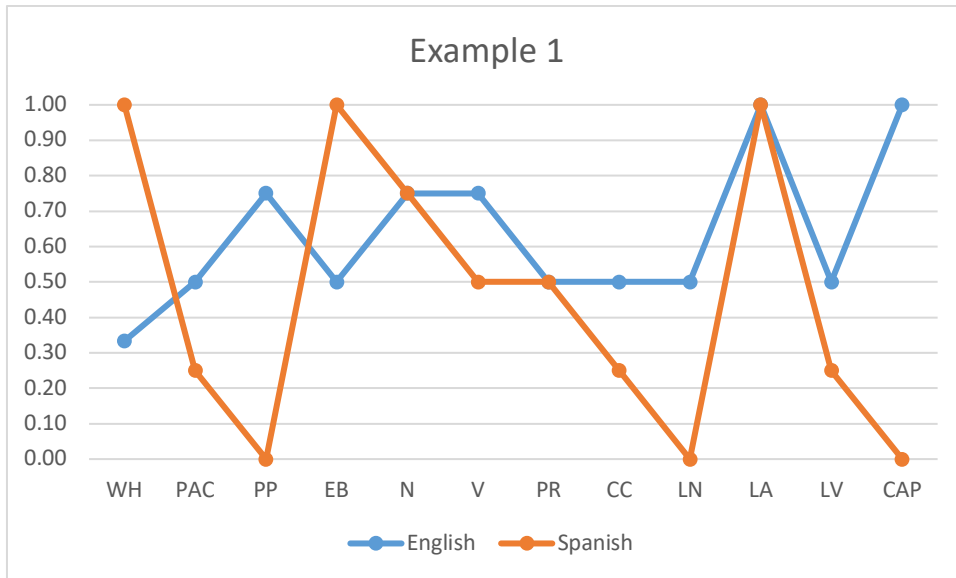
2.4.2 Best Scores across English and Spanish

The Best Score uses the maximum score on each subtest type from each language to get an overall view of the child's functioning (Peña, Gutiérrez-Clellen, Iglesias, Goldstein, & Bedore, 2018; de Villiers, 2015). Best Scores capture the fact that a bilingual child's knowledge can be distributed between their two languages (Peña, Bedore, & Zlatic-Giuta, 2002). Thus, Best Scores were computed from the types of language items tested in each language: Wh- questions, Noun Learning, and so forth. For each pair of types (English-Spanish), the maximum score

achieved in a language was included in the child's total score. The comparison was of proportions correct as the numbers of items in each area varied. These total scores provide Best *area* scores (e.g. Best Process, Best Vocabulary, Best Syntax) and Best *Totalscores*.

It is evident that the two children presented in Figure 2 differ in what they find easy or hard in each language. But is every case unique, or are there similarities across the group? One troubling question in a comparison of this sort is how we could match the level of sophistication of items in Spanish to those in English. For example, despite the piloting and first Tryout work, we might have accidentally chosen a harder set of verbs in English, or a more difficult set of scenarios for conjunctions in Spanish. If that were true, then the Best Scores would give the pattern away, because there would be uniformity as to which language the children did better in for a given subtest. On the other hand, if this varies, then the pattern must be due to something other than the difficulty of the items chosen.

To answer this question, we derived *difference* scores on each subtest, i.e., English minus Spanish. Then we added the subtests together for each general area : Vocabulary, Syntax and Process. A positive score means English was superior to Spanish for that skill, and a negative score means Spanish was better. The differences across the whole sample were tested using a one sample t-test where, hypothetically, the expected value is zero if the children as a group knew both languages equally. In fact, there are significant differences across the subtests, with four favoring Spanish (verbs, prepositions, wh-questions, and fast mapping adjectives) and the remaining eight favoring English. However, the differences in general are very close to zero (mean=.02, or 2% difference) and with a large standard deviation (.36).

To the extent that a subtest changes valence across time, it must be that the child is acquiring knowledge that allows them to score higher in the other language. For almost all

subtests, there is a significant drift towards English skills being better than Spanish skills from age 3;0 to 5;11. This is in keeping with the children's attendance at largely English speaking day-cares and preschools. Taking a wider lens, difference scores for the summed subtests in Vocabulary, Syntax, and Process show a broader pattern in which Spanish Vocabulary (though only verbs and prepositions) dominates, whereas Syntax and Process shift earlier to an English preference. A repeated measures ANOVA with the three area scores as the dependent variable and age and gender as the independent variables revealed a significant difference across the different areas ($F(1,357) = 43.97$, $p<.001$, $\eta_p^2 = .12$), and a small but significant interaction with age ($F(1,357= 3.14$, $p<.05$, $\eta_p^2=.18$). Vocabulary is different in profile than the other two areas since children do better on the Spanish items, but all show the same movement across age towards English. Figure 2 shows the change across age in which general area children do better in Spanish or English.

Figure 2: Mean difference scores by area across languages across age groups.

Error bars: 95% CI
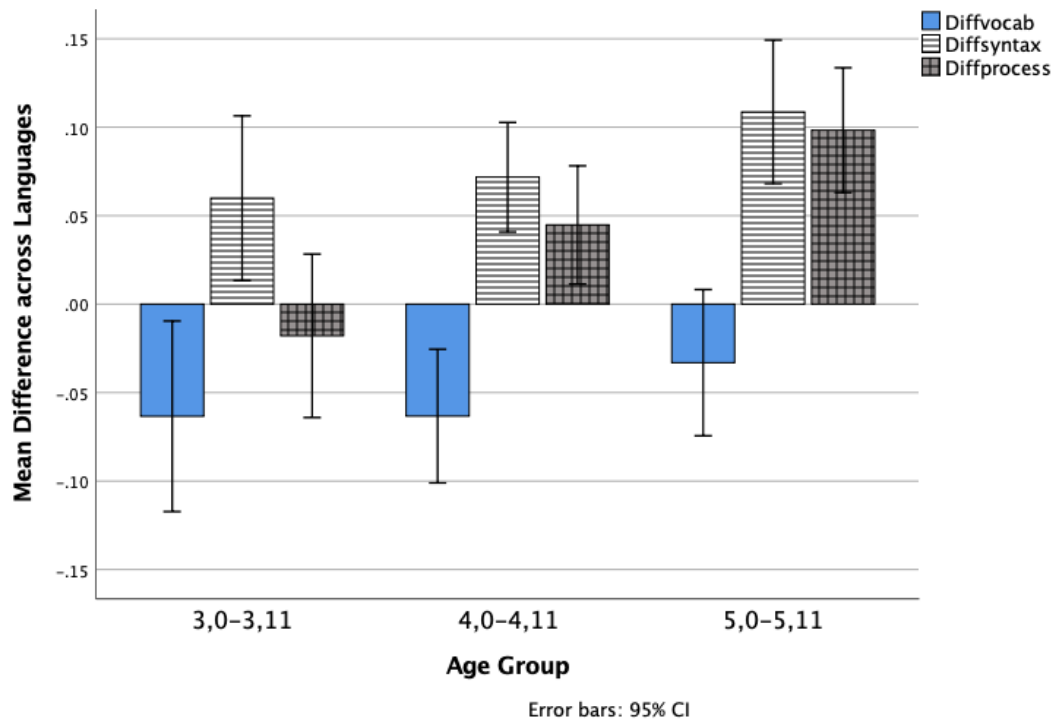
Are there other factors controlling this shift in strength of abilities by language? A second analysis relates to the caregivers' reports of language used in daily life. Children's daily experience varied from mostly English to mostly Spanish, by caregiver report. Though we only had this information on a subset of 109 children, as a rough index we divided the group at the midpoint of the 5-point scale to choose children who heard more English (64) and those who heard more Spanish (45). This was used in a further repeated measures ANOVA looking at the difference measures on the area types. There is a significant effect of area type $(F (1,107) = 43.1$, $p<001$, $\eta_p^2 =.29)$ that varies by the child's experience of language use in the home $(F(1,107)= 10.36, p= .002, \eta_p^2 =.09)$. Those children being raised with more English do better on English across areas compared to the children raised with more Spanish, who show the opposite

preference in vocabulary. It is clear that some of the variability is predicted by age and experience in English versus Spanish.

2.5 Fifth Challenge and Solution: Assessments must be psychometrically sound

Screening instruments have to pass certain psychometric standards to be useful for practitioners, and these include establishing that they have sufficient validity and reliability.

2.5.1 Construct Validity

Validity of an instrument is examined to ensure a test is actually measuring what it claims to measure. That is, do the items on the QUILS: ES form a coherent set (*construct validity*)? A screener must be based on phenomena that expert researchers, teachers, and other educators regard as linguistically significant and educationally meaningful for children in the age range being examined. Without adequate theoretical and empirical backing to establish construct validity, no screener or test can be considered adequate.

2.5.2 Concurrent/Convergent Validity

The QUILS: ES was also assessed for concurrent or *convergent validity*: does children's performance on the QUILS: ES correlate with their results on other established language assessments of Spanish and English for Dual Language Learners of those languages?

We compared children's performance on the QUILS: ES, correlating English and Spanish scores separately against a second, existing test for bilingual children. Standardized language measures were administered within four weeks of QUILS: ES testing to establish concurrent validity.

The PLS-5 was chosen to check concurrent validity for the QUILS: ES as it also provides both a Spanish and an English score. A subgroup of 44 children tested on QUILS: ES completed the English Preschool Language Scales-5 (PLS; Zimmerman, Steiner, & Pond, 2011) and the Spanish PLS-5 (Zimmerman, Steiner, & Pond, 2012) – including the Expressive Communication and Auditory Comprehension portions of the test[2]. This group completed the full PLS administered in English and in Spanish in counterbalanced order.

The PLS-5 has two components: expressive competence (EC) and receptive competence (AC) and provides a total score in each language. To prepare the data for the validity analyses, a total score was derived for the QUILS: ES by adding together the 45-item scores in each language. To compare with the standard scores of the PLS and PPVT, these totals were then converted to standard (Z) scores by age group. Bivariate correlations between the QUILS: ES in English and the PLS-total English reveal a moderately high correlation ($r(44)=.693$, $p<.001$). Bivariate correlations between the QUILS: ES total Spanish scores and the PLS-total in Spanish reveal a smaller but still highly significant correlation with the ($r(44))=.449$, $p<.002$).

As part of the concurrent validity testing, 44 other children completed the QUILS: ES and the BESOS: the Bilingual English/Spanish Oral Screener. This test designed for ages 4 to 7 contains Morphosyntax (BESOS-MS) and Semantics (BESOS-S) subtests in both English and Spanish (Lugo-Neris, Peña, Bedore, & Gillam, 2015). We looked at the inter-correlations between the Spanish and English BESOS with the Spanish and English QUILS: ES, shown in Table 9.7. Since the BESOS has only been normed for ages 4 and up, we only included in the analyses the 29 children (out of 44 total) who were older than 4.

The correlations reported for the relation between QUILS and BESOS were only modest, though for all but the semantics area are still statistically significant (at p<.05). However, what is tested on each screener is quite distinct. The semantics tests on the BESOS have to do with conceptual categories and relationships of nouns, whereas QUILS taps knowledge of nouns, verbs, prepositions, and conjunctions. The BESOS-MS focuses on morphological markers that are specific to SLI in Spanish or in English. QUILS: ES does not test morphology except for the past auxiliary and copula, as we tried to avoid areas of specific difficulty for African American English learners. Finally, the BESOS was specifically designed to screen children who might have a language impairment, whereas QUILS: ES was designed as a quick screener for all children who are low language performers, including children with language impairment. For all these reasons the two tests are likely to be complementary in the picture they paint of language abilities.

For an additional test of concurrent validity only with the English half of the QUILS: ES, the PPVT in English was used with a small sample of 20. The total English QUILS: ES score correlated very well with the English PPVT ($r(20)=.727$, $p<.001$).

2.5.3 Internal Reliability

A test must also have internal integrity. The items on the test must form a coherent set that inter-correlate even though the items may vary in difficulty. To ensure this for the QUILS: ES, an analysis called Rasch modeling was used (Rasch, 1960; Wright & Stone, 1979). In seeking internal integrity, the goal is to identify which items serve the intended purpose and which items are poor at doing so, or are redundant because other items test the same thing. *Item response*

theory, using Rasch modeling, provides a way to evaluate the worth of the individual items to the test as a whole.

Rasch analyses were undertaken separately for English and Spanish. Rasch results were highly promising on the 48-item versions of each test. In the Rasch analyses with 48 items and 446 children, three items were removed from each test on the basis of their misfit values, resulting in final tests of 45 items in each language.

The results of Rasch analyses for the QUILS: ES sample, for both Spanish and English sections had Infit values for items within the expected range (0.8 and 1.3), denoting good fit of these items to the scale. The person mean and item mean are close to each other denoting a good match between items and persons. Item-maps for the English and Spanish sections show that these tests discriminate well between children of varying abilities. For both English and Spanish sections, the screener captured abilities ranging from under two standard deviations below the mean to over two standard deviations above the mean. Moreover, items have satisfactory spread throughout the scale.

Demonstrating that a test's items have internal consistency is another metric of reliability. Cronbach's (1951) coefficient alpha is used to calculate this. Coefficient alpha provides a lower bound value of test reliability and is considered to be a conservative estimate of a test's reliability (Allen & Yen, 1979; Carmines & Zeller, 1979; Reynolds, Livingston, & Willson, 2009). The Cronbach Alphas for English was .89, and for Spanish it was .85, while the components ranged from .65 to .82. These good-to-high coefficient values demonstrate that items are coherent in measuring the unidimensional construct underlying each area of the screener and also each language of QUILS: ES. DIF analysis was also performed with respect to gender. Although some individual items show DIF in favor of one or another gender, it can be argued that since on balance,

the DIFs cancel out, neither of the groups is disadvantaged by including these items (Nandakumar, 1993).

2.5.4 Test–Retest Reliability

A second session of QUILS: ES testing was administered four to six weeks after the initial QUILS: ES testing to establish test-retest reliability. Children received both English and Spanish portions of the QUILS: ES after their initial QUILS: ES session, in the same order in which the initial QUILS: ES was administered. As with the initial QUILS: ES session(s), for the retest, the two language portions of the QUILS: ES were given within two weeks of each other. Using Best Scores as the measure, the test-retest reliability was high (.89).

The instrument has good internal reliability, test-retest reliability, and validity against other accepted measures like the BESOS, the PLS, and the PPVT(English). QUILS: ES has not yet been fully tested on a clinical population of children with language delays, though that work is underway in two clinics and the results are promising. We need to establish the specificity and sensitivity of the test for clinical use, but its use as a screener in educational settings is not precluded and should provide useful information.

At the completion of both sections of the text, the QUILS:ES provides several kinds of automatic reports designed for parent, teacher and school in different levels of specificity and formality, of the child's individual language scores, their norms, percentiles, and an evaluation of risk status based on their overall performance. A sample "Student Brief Report" is provided in Appendix B.

3.0 Discussion

In this paper we have addressed five significant issues that need to be tackled by designers of a screening instrument for Dual Language Learners. We argue that construct validity is essential: SLPs, linguists, psychologists, and experts on language acquisition and disorder need to collaborate to choose appropriate areas of assessment. These areas should reflect linguistic properties that are diagnostic of the stages of development in early childhood, but also their use in everyday life, for example in preparation for the demands of schooling. That is why we emphasize assessment of how children can learn new things, not just a sample of what they already know. It is important that the sample match the group for whom it is designed, both in terms of adequate representation across SES and, on the screener, that the items are neutral with respect to culture and dialect. Given the way dual languages are represented in the mind, we emphasize that the scores take into account distributed knowledge.

We addressed each in turn and presented the solutions we adopted in the making of a new screener for Spanish-English bilinguals in the US. The results demonstrate that a touchscreen screener for bilingual Spanish-English learners is a viable option for fair testing of children aged 3-6 years in the US. It is a self-contained test, where narration and scoring are automatic, making it broadly useful even in areas where the number of bilingual SLPs is low relative to the population of children in need of screening. We would like to test it in wider arenas such as Latin America, where the Spanish section might prove useful even with monolingual learners of Spanish.

The screener emphasizes the use of the Best Score as a fair index of a Dual Language Learner's competence with language development. There are three new findings here. First, we demonstrate that there is distributed knowledge in Dual Language Learners not just in vocabulary but in syntax and process indices too. Second, differences between subtests begin to

switch over the course of the preschool years towards English, though Spanish retains strength in the areas of Vocabulary. Third, these changes are predictable from parental reports of the proportional use of the different languages in the home.

Finally, we recognize that this is a screener with potential extension to even younger children. We recently completed work on a touch screen assessment (BabyQUILS) with simpler language subtests that includes vocabulary, syntax, and process items, and is normed on US two-year-olds (N=440) who are monolingual in English (de Villiers et al., 2019). In the process of collecting data, some children (N=83) were tested whose exposure in the home was to other languages as well, and many of their scores approached the normal range for English, especially by 30-36 months. This gives us confidence that a screener is a future possibility for two-year-old Dual Language Learners, who could reveal their full linguistic knowledge distributed across different versions of the assessment. The existing work on this younger age range focuses heavily on vocabulary, so a broader assessment that included grammar and process would be a valuable addition to the research base.

References

Allen, M. K., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, CA:
Brooks/Cole Pub. Co.

Artiles, A. J., Rueda, R., Salazar, J., & Higareda, I. (2005). Within-group diversity in minority
disproportionate representation: English Language Learners in urban school districts.
Exceptional Children, 71, 283-300.

Barrueco, S., Lopez, M., Ong, C. A., & Lozano, P. (2012). *Assessing young children within and
across two languages*. Baltimore, MD: Paul H. Brookes Publishing Co.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Carmine, E. G., & Zeller, R. A. (1979). *Reliability and validity.* Thousand Oaks, CA: Sage

Conti-Ramsden, G., & Botting, N. (2004). Social difficulties and victimization in children with SLI
at 11 years of age. *Journal of Speech, Language, and Hearing Research*, 47, 145–161.

Core, C., Hoff, E., Rumiche, R., & Señor, M. (2013). Total and Conceptual Vocabulary in Spanish–
English Bilinguals From 22 to 30 Months: Implications for Assessment. *Journal of Speech,
Language, and Hearing Research*, 56 • 1637–1649.

Craig, H. K., (1993). Social skills of children with specific language impairment: Peer relationships.
Language, Speech, and Hearing Services in Schools, 24, 206–215.

de Villiers, J. (2015). Taking account of both languages in the assessment of dual language learners.
In *Seminars in Speech and Language*, 36(2), 120–132.

de Villiers, J., Golinkoff, R. M., Hirsh-Pasek, K., Iglesias, A., Jackson, E., Levine, D., Odean,
R., McCollum, R., Puttre, H., & Weaver, H. (2019, November). Measuring receptive
language competence in two-year-olds: Technological innovations & validity outcomes.

Poster presented at the American Speech and Hearing Association Annual Convention, Orlando, FL.

Dollaghan, C. (2013). Late talker as a clinical category: A critical evaluation. In L. A. Rescorla & P. S. Dale (Eds.), *Late talkers: Language development, interventions, and outcomes* (pp. 91–112). Baltimore, MD: Paul H. Brookes Publishing Co.

Dong, Y., Gui, S., MacWhinney, B (2005). Shared and separate meanings in the bilingual mental lexicon. *Bilingualism: Language and Cognition*, 8, 221-238.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test— Third Edition*. Circle Pines, MN: American Guidance Service.

Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) Early Child Care Research Network (ECCRN). (2005). *Child care and child development: Results from the NICHD study of early child care and youth development.* New York, NY: Guilford Press.

Fisher, C. (1996). Structural limits on verb mapping: The role of analogy in children's interpretation of sentences. *Cognitive Psychology*, 31, 41–81.

Friend, M., Smolak, E., Liu, Y., Poulin-Dubois, D., & Zesiger, P. (2018). A cross-language study of decontextualized vocabulary comprehension in toddlerhood and kindergarten readiness. *Developmental psychology*, *54*(7), 1317.

Gillam, R. B., Peña, E.D., Bedore, L.M., Bohman, T.B., Mendez-Perez, A. (2013). Identification of Specific Language Impairment in bilingual children: I. Assessment in English. *Journal of Speech, Language, and Hearing Research*, 561813-1823.

Glogowska, M., Roulstone, S., Enderby, P., & Peters, T. J. (2000). Randomized controlled trial of

    community based speech and language therapy in preschool children. *British Medical*

    Journal, 321(7266), 923–928.

Golinkoff, R. M., Hirsh-Pasek, K., Bailey, L., & Wenger, N. (1992). Young children and adults

    use lexical principles to learn new nouns. *Developmental Psychology, 28,* 99-108.

Golinkoff, R. M., Jacquet, R. C., Hirsh-Pasek, K., & Nandakumar, R. (1996). Lexical principles

    may underlie the learning of verbs. *Child Development*, 67, 3101–3119.

Hirsh-Pasek, K., Golinkoff, R. M. (Eds.). (1996a). *The origins of grammar: Evidence from early*

    language comprehension. Cambridge, MA: MIT Press.

Hirsh-Pasek, K., Kochanoff, A., Newcombe, N. S., & de Villiers, J. (2005). Using scientific

    knowledge to inform preschool assessment: Making the case for "empirical validity."

    Social Policy Report, 19.

Jerome, A. C., Fujiki, M., Brinton, B., & James, S. L. (2002). Self-esteem in children with specific

    language impairment. *Journal of Speech, Language, and Hearing Research*, 45, 700–714.

Language and Reading Research Consortium (LARRC). (2015). The dimensionality of language

    ability in young children. *Child Development*, 86, 1948–1965.

Law, J., Kot, A., & Barnett, G. (1999). A comparison of two methods for providing intervention to

    three year old children with expressive/receptive language impairment. London, United

    Kingdom: City University London.

Leonard, L. B. (2014). *Children with specific language impairment.* Cambridge, MA: MIT Press.

Lindsay, G., & Dockrell, J. (2000). The behaviour and self-esteem of children with specific speech

    and language difficulties. *British Journal of Educational Psychology*, 70, 583–601.

Lugo-Neris, M. J., Peña, E. D., Bedore, L. M., & Gillam, R. B. (2015). Utility of a language screening measure for predicting risk for language impairment in bilinguals. American *Journal of Speech-Language Pathology*, 24(3), 426–437. doi:10.1044/2015_AJSLP-14-0061

Maier, M. F., Bohlmann, N. L., Palacios, N. (2016). Cross-Language Associations in the Development of Preschoolers' Receptive and Expressive Vocabulary. *Early Childhood Research Quarterly*, 36, 49-63.

Melby-Lervåg, M. & Lervåg, A. (2011). Cross-linguistic transfer of oral language, decoding, phonological awareness and reading comprehension: a meta-analysis of the correlational evidence. *Journal of Research in Reading*, 34, 114–135.

Nandakumar, R. (1993). Simultaneous DIF amplification and cancellation: Shealy-Stout's test for DIF. *Journal of Educational Measurement*, 30, 293–311.

National Center for Education Statistics (2015). *Digest of Education Statistics*. Retrieved from (https://nces.ed.gov/programs/digest/d15/tables/dt15_104.40.asp?current=yes)

Olswang, L. B., Rodriguez, B., & Timler, G. (1998). Recommending intervention for toddlers with specific language learning difficulties: We may not have all the answers, but we know a lot. *American Journal of Speech-Language Pathology*, 7, 23–32.

Pace, A., Alper, R., Burchinal, M. R., Golinkoff, R. M., & Hirsh-Pasek, K. (2019). Measuring success: Within and cross-domain predictors of academic and social trajectories in elementary school. *Early Childhood Research Quarterly*, *46*, 112-125.

Paradis, J., Rusk, B., Duncan, T.SD., Govndarajan, K. (2017). Children's Second Language Acquisition of English Complex Syntax: The Role of Age, Input, and Cognitive Factors. *Annual review of Applied Linguistics*, 37, 148-167.

Paul, R. (1996). Clinical implications of the natural history of slow expressive language development. *American Journal of Speech-Language Pathology*, 5(2), 5–30.

Pearson, B. Z. (1998). Assessing lexical development in bilingual babies and toddlers. The *International Journal of Bilingualism*, 2, 347-372.

Pearson, B. Z., & Fernández, S. (1994). Patterns of interaction in the lexical growth in two languages of bilingual infants and toddlers. *Language Learning*, 44, 617-653.

Peña, E. D., Bedore, L. M., Iglesias, A., Gutiérrez-Clellen, V. F., & Goldstein, B. A. (2008). *Bilingual English Spanish Oral Screener– Experimental Version (BESOS).* Unpublished instrument.

Peña, E.D., Bedore L., &; Zlatic-Giuta, R. (2002). Use of a category generation task to assess vocabulary skills of Spanish–English bilinguals. *Journal of Speech, Language, & Hearing Research*, 46, 938–947.

Peña, E. D., Gillam, R. B., Bedore, L. M., & Bohman, T. M. (2011). Risk for poor performance on a language screening measure of bilingual preschoolers and kindergarteners. *American Journal of Speech-Language Pathology*, 20, 302–314. Doi:1058-0360-2011-1010-0020

Peña, E.D., Gutiérrez-Clellen, V. F., Iglesias, A., Goldstein, B. A., & Bedore, L. M. (2018). *Bilingual English–Spanish Assessment*™ (BESA™). Baltimore, MD: Paul H. Brookes Publishing Co.

Ramey, C. T., & Ramey, S. L. (1998). Early intervention and early experience. *American Psychologist*, 53, 109–120.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Institute of Educational Research.

Rescorla, L. (2000). Do Late talking toddlers turn out to have reading difficulties a decade later? *Annals of Dyslexia*, 50, 87-102

Reynolds, C. R., Livingston, R. B., & Wilson, V. (2009). *Measurement and assessment in education*. New York: Pearson.

Rice, M. L. (1993). Social consequences of specific language impairment. *Language acquisition problems and reading disorders: Aspects of diagnosis and intervention*, 111-128.

Roberts, M. Y., & Kaiser, A. P. (2015). Early intervention for toddlers with language delays: A randomized controlled trial. *Pediatrics*, 135, 686–693.

Robertson, S. B., & Weismer, S. E. (1999). Effects of treatment on linguistic and social skills in toddlers with delayed language development. *Journal of Speech, Language, and Hearing Research,* 42, 1234–1248.

Rojas, R., Iglesias, A., Bunta, F., Miller, J., Goldenberg, C., & Reese, L. (2016). Interlocutor differential effects on the expressive language skills of Spanish-speaking English language learners. *International Journal of Language & Communication Disorders*, 18, 166–177.

Schmeiser, C. B., & Welch, C. J. (2006). Test development. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 307–353). Westport, CT: Praeger Publishers.

Seidl, A., Hollich, G., & Jusczyk, P. W. (2003). Early understanding of subject and object wh-questions*. Infancy,* 4, 423–436.

Seymour, H. N., Roeper, T., & de Villiers, J. G. (2004). Conclusions, future directions, and implications for remediation. *Seminars in Speech and Language*, 25, 113–115.

Thal, D., & Bates, E. (1988). Language and gesture in late talkers. *Journal of Speech and Hearing Research,* 31, 115–123.

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research,* 40, 1245–1260.

Tucci, A., Plante, E., Vance, R. & Oglivie, T. (2019) Data-driven item selection for the Shirts and Shoes Test. *Journal of Communication Disorders*, 78 (2019) 46–56

U.S. Census Bureau, (2016). *Educational Attainment in the United States: 2016*. Retrieved from (https://www.census.gov/data/tables/2016/demo/education-attainment/cps-detailed-tables.html).

Wake, M., Tobin, S., Girolametto, L., Ukoumunne, O. C., Gold, L., Levickis, P., . . . Reilly, S. (2011). Outcomes of population based language promotion for slow to talk toddlers at ages 2 and 3 years: Let's Learn Language cluster randomized controlled trial. *British Medical Journal,* 343, d4741.

Weisleder, A., & Fernald, A. (2009). Real-time processing of postnominal adjectives by Latino children learning Spanish as a first language. In *Proceedings of the 33rd annual Boston University Conference on Language Development* (pp. 611–621). Somerville, MA: Cascadilla Press.

Zimmerman, I. L., Steiner, V. G., &; Pond, R. E. (2011). *Preschool Language Scales–Fifth Edition (PLS-5)*. San Antonio, TX: Pearson.

Appendix A

Area	Item Type	Item	
		English	Spanish
Vocabulary	Nouns	Find, the fireworks	Enséñame, el marinero.
	Verbs	Who is lugging something?	¿Quién está regresando?
	Prepositions	Show me the chickens behind the clock	Enséñame, la muñeca está arriba del regalo.
	Conjunctions	Who ate the food before the cat jumped on the table?	¿Quié recogió las hojas antes de que la niña subiera al árbol?
Syntax	Wh-Questions	Who is kissing the baby?	¿Qué le está cayendo encima a la niñita?
	Past tense	Where was the flower?	¿Dónde estaba el sombrero?
	Prepositional Phrase	Show me, the mouse with the yellow hat	Enséñame, la bandera debajo de un mono sucio.
	Embedded Clauses	What did Cowboy Bob tell Mia to do?	¿Dónde le dijo Sofia a Mauricio que estaba Javier?
Process	Verb Learning	Find, the boy is meeging	Encuentra, alguien está braleando a alguien
	Noun Learning	Show me, the merf.	Enséñame la teña
	Adjective Learning	What else is dorbish?	¿Qué otra cosa es quefosa?
	Converting Active to Passive	Which one got koobed?	¿Qué fue braleado?

Appendix B

QUILS:ES
QUICK INTERACTIVE LANGUAGE SCREENER™: ENGLISH–SPANISH

Student Brief Report

●●

Student Information

Student Name: Debra F
Student ID: 7272015

Date of Birth: 07/27/2015
Date of Screening: 04/03/2020
Age at Screening: 4 years

Strong language skills are essential for every student's success in school and in life. The Quick Interactive Language Screener™: English —Spanish (QUILS™: ES) measures a student's emerging abilities in Vocabulary, Syntax, and Process and offers an overall score. These results are expressed as standard scores and percentile ranks. (See Chapter 9 of the QUILS: ES User's Manual for more information.)

Debra F's Performance

On 04/03/2020, Debra F's language skills were screened using the QUILS: ES. The overall percentile rank of 83.8 means that Debra F scored as well as or better than 83.8% of 4-year-olds in the standardization sample. In addition to the overall results, standard scores and percentile ranks for each of the three areas of the QUILS: ES were calculated. Based on these calculations, a recommendation is listed below for Debra F's results.

— The 25th percentile rank (see criteria for recommendation below)
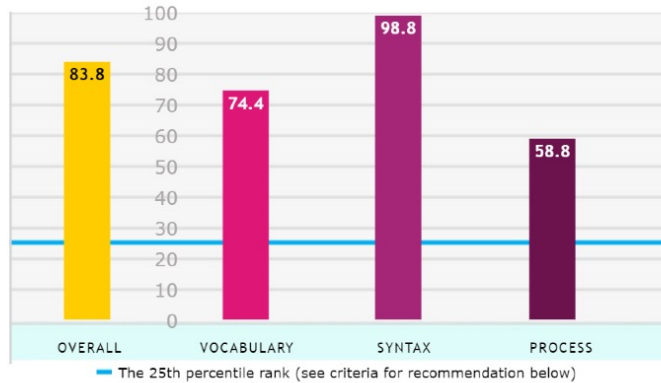
Overall		Vocabulary		Syntax		Process	
Standard Score	**Percentile Rank**	**Standard Score**	**Percentile Rank**	**Standard Score**	**Percentile Rank**	**Standard Score**	**Percentile Rank**
114	83.8	110	74.4	128	98.8	104	58.8

Vocabulary area asks about words students use or understand, including ordinary things (nouns), actions (verbs), prepositions, and conjunctions.
Syntax area asks about structure of sentences including *wh*-questions, tense markers (past tense), prepositional phrases, and embedded clauses.
Process area asks about learning new words (verbs, nouns, and adjectives) and about how children use syntax, such as converting active sentences to passive sentences.

Recommendation:

Based on Debra F's performance on the QUILS: ES, Debra F's language comprehension appears to be within the typical range relative to age, and no follow-up evaluation is recommended at this time.

> **Criteria for recommendation:**
> - Students with an overall percentile rank below 25 should be referred for follow-up evaluation.
> - Students with a percentile rank below 25 in the Process area (regardless of the Vocabulary and Syntax scores) should be referred for follow-up evaluation.
> - Students with percentile ranks below 25 in both Vocabulary and Syntax (regardless of the Process score) should be referred for follow-up evaluation.

●●

BROOKES