Parents Can Test Preschool Children's Language: the Parent-Administered Language (PAL) Test[1]

Karin Stromswold,* Ellyn Sheffield,† Debra Truit,† and Diane Molnar*

*Department of Psychology & Center for Cognitive Science
Rutgers University – New Brunswick

†Department of Psychology, Salisbury University

ABSTRACT

This paper investigates the concurrent validity of the Parent Administered Language (PAL) test, a 10-minute test of preschool-aged children's articulation, lexical access, receptive vocabulary, and syntactic abilities. Parents of 122 racially and socio-economically (SES) diverse children administered the PAL test to their child and an experimenter administered the Denver Articulation Screening Examination and the three core subtests of the Clinical Evaluation of Language Fundamentals- Preschool 2 (CELF). These data suggest that PAL test scores are valid measures of the linguistic abilities of boys, girls, African American children, non-Hispanic white children, low SES children, and high SES children (composite PAL- CELF score r's $\geq$ .70) for all group). PAL test scores were also excellent at discriminating between children who were and were not language impaired (area under Receiver Operating Characteristic curve > .90). Thus, the PAL test may fill an important niche in the language assessment tools available to researchers and clinicians.

[1] Portions of this work were supported by grants from the National Science Foundation (BCS-0446838) the Bamford-Lahey Children's Foundation, and an unrestricted grant from Rutgers University. We are grateful to Mahvish Haq, Avneet Madan, Ali Shafget, and Ravali Tammareddy and for their assistance coding the data and to the parents and children who participated in this study.

Parents Can Test Preschool Children's Language: the Parent-Administered Language (PAL) Test

Parent-completed questionnaires of children's language are popular because they allow researchers and clinicians to quickly and inexpensively quantify the linguistic abilities of large numbers of children. The MacArthur-Bates Communicative Development Inventory (CDI) checklists are the most frequently used parent-completed language questionnaires, with over 200 published papers using the CDI checklists. The CDI-I assesses 8- to 15- month old children's receptive and expressive vocabulary by having parents check off which words their child understands and produces (Fenson et al., 1993a). The CDI-II is designed for children ages 16 to 29 months and contains expressive vocabulary and sentence production checklists (Fenson et al., 1993b). The CDI–III is much like the CDI-II, but is meant for children who are 30 to 36 months of age (Dale, 2001). Although some researchers and clinicians have expressed reservations about using the CDI checklists as research or clinical tools (e.g., Feldman et al., 2000), several studies indicate that CDI scores are reasonably good measures of toddlers' language skills and that CDI scores are fairly good at identifying toddlers whose language development may be delayed (e.g., Dale, Price, Bishop, & Plomin, 2003; Feldman et al., 2005; Heilmann, Weismer, Evans, & Hollar, 2005).

Rescorla's (1989) Language Development Survey (LDS) is another popular parent-completed language questionnaire, with over 20 published papers reporting using the LDS. In the LDS, parents of children ages 18 to 35 months report whether their child spontaneously says 310 words, whether their child produces multi-word utterances and, if so, give examples 3 of their child's longest sentences. Like the CDI, studies have shown that the LDS is a reliable test and is useful as a screening test for developmental language impairments (e.g., Rescorla, 1989; Klee et al., 1998; Rescorla & Alley, 2001; Rescorla & Achenbach, 2002). The LDS differs from the CDI insomuch as the CDI was designed primarily as a tool for measuring the language of typically-developing children, whereas the LDS was designed primarily as a screening test for identifying toddlers with language delays.

A third popular parent-completed language questionnaire is the Children's Communication Checklist-2 (CCC-2, Bishop, 2006), which has been used in over 24 published papers. The CCC-2 is a language-screening test for 4- to 16-year old children. The CCC-2 asks parents about their child's semantic/pragmatic and discourse (e.g., initiation of conversations, choice of words, understanding of non-literal language, discourse coherence, etc.), phonological, lexical, and syntactic skills.

Finally, the Communication and Symbolic Behavior Scales Developmental Profile (CSBS, Wetherby & Prizant, 2002) includes parent-completed language assessments. In the one-page CSBS checklist and the four-page CSBS questionnaire, parents answer questions about their 6- to 24- month old child's articulation, understanding and production of gestures, words and sentences. Like the LDS and CDI, studies suggest the CSBS has good concurrent and predictive validity (Wetherby, Allen, Cleary, & Kublin, 2002; Wetherby, Goldstein, Cleary, Allen, & Kublin, 2003).

Notice, that although there are parent-completed language assessment tools for children who are less than 37 months old and more than 47 months old, there is no parent-completed language assessment tool for children who are three year olds.[2] Notice also that the CDI, LDS, CCC-2 and CSBS are all questionnaires. A frequently-voiced concern about estimates of children's linguistic abilities that are derived from parent-completed questionnaires is that parents may have biased or faulty memories and judgments with respect to their child's linguistic accomplishments. Whereas completing a vocabulary checklist probably requires no special skills, completing a morphological, syntactic or phonological checklist might. Particularly worrisome is the possibility that parents' beliefs about the size of their child's vocabulary may influence their assessment of their child's abilities in other aspects of language (see Stromswold, 2001). Usually, the gold standard for measuring preschool children's language is a well-designed, valid language test administered by a trained professional under optimal conditions. The reason for this is that standardized test scores are generally more reliable, objective and quantitative measures of

[2] Bricker and Squire's (1999) parent-completed Ages and Stages Questionnaire (ASQ) is designed for children up to 60 months of age. However, because it asks only 6 questions about children's communication ability, it provides a somewhat crude measure of children's communicative skills.

children's language than other measures that are feasible to collect from large numbers of children. [3]

Because, unlike parent-completed questionnaires, parent-administered tests of language do not require parents to remember what their child has done in the past or judge what their child can do now, scores on parent-administered language tests may provide a more reliable and valid measure of children's non-lexical abilities than scores on parent-completed language tests. Thus, what is needed to fill the gap in the arsenal of language assessment tools is a quick -- yet comprehensive – *test* of preschool children's spoken language which is simple enough for parents to administer and provides a valid quantitative measure of children's linguistic abilities. Such a test would make it feasible to screen large numbers of preschool children and identify children who should be evaluated by speech-language pathologists. In addition, such a test would provide researchers with a valid, quantitative measure of the linguistic abilities of the preschool children who participate in their studies. This would be particularly useful for researchers whose studies involve many participants (e.g., genetic studies, epidemiological studies, large intervention studies, etc.), and for researchers whose primary research focus is not language development.

The ideal screening test would not only identify children who suffer from language delays, but identify what aspects of their language may be delayed. Similarly, the ideal research test would provide researchers with quantitative measures of their participants' phonological, lexical and syntactic abilities, and not just a measure of the children's overall linguistic abilities. Unfortunately, no such test exists: a comprehensive review of the literature and of existing speech-language tests, queries to the CHILDES and ASHA electronic mailing lists, and a recent report evaluating existing tests of preschool children's language abilities (Nelson, Nyren, Walker, & Panosha, 2006; US Preventive Services Task Force, 2006) failed to uncover a parent-administered test of preschool children's language abilities.

We developed a 10-minute, parent-administered test of preschool children's language, the **P**arent **A**dministered **L**anguage (PAL) test to fill this gap. The PAL test assesses children's abilities in the areas of spoken language most frequently assessed in standardized language tests (articulation, receptive and expressive vocabulary and syntax) using tasks that 1) are simple and fast to administer; 2) require no special equipment; and 3) elicit relatively unambiguous responses that are easy to observe and record.

PAL Articulation Test. The PAL Articulation test uses a word repetition task. Because it is easiest to detect mispronunciations at the beginnings of words (onsets), the PAL only tests whether children pronounce onsets correctly. In this test, parents ask their child to repeat 12 monosyllabic words, and they check off whether the child correctly says the onset of each word. If the child fails to respond, the parent checks "no response." If the child mispronounces an onset, the parent records what the child said. For example, if the target word is ***r****at* and the child says ***w****at*, the parent writes *wat*. If the target word has a consonant cluster as an onset, the parents reports whether the child says the entire consonant cluster correctly. For example, the child only gets credit for correctly pronouncing the onset of the word *split* if he or she correctly says the /s/, /p/ and /l/ in that order. In order to minimize the number of items on the test and prevent ceiling or floor effects, 3-, 4- and 5-year old children repeat different words, and for each age group, the wordlists include some words with onsets that children of that age typically have mastered and some words with onsets that children typically have not mastered (Sanders, 1972; Vihman, 1996). For example, at 4 years of age, children repeat the words *rat*, *lip*, *ship*, *cheek*, *zip*, *jeep*, *that*, *thin*, *trick*, *clock*, *frog* and *split*. Appendix 1 gives all PAL test items.

PAL Receptive Vocabulary Test Like most tests of children's receptive vocabularies, the PAL test uses a picture-pointing task. However, rather than the usual task of choosing the correct picture from a small set of pictures with the set of pictures being different for each word, in the PAL Receptive Vocabulary test, all of the pictures that correspond to all of the words appear on a single page. So, when the parent says the word *mittens*, the child must choose the correct picture from a set of 12 pictures. Because the child chooses from a larger number of pictures, for each trial, the probability of randomly

[3] Because what is appropriate from a discourse and semantic/pragmatic perspective critically depends on the context in which a speech act occurs, discourse/pragmatic skills are difficult to assess in non-natural situations. For this reason, few standardized language tests assess these skills and questionnaires may provide as good or better estimates of children's pragmatic skills (Bishop, 1998).

choosing the correct picture is less than on traditional picture-pointing vocabulary tests. Thus, the range of possible scores on the PAL Receptive Vocabulary test is greater than on traditional picture-matching tests of the same length. This means that the PAL test can have fewer words and, hence, be faster to give.

The words on the PAL Receptive Vocabulary test are 8 easily depicted nouns. Because the same words are used for 3-, 4- and 5-year old children, and scores must be informative for all children (i.e., no ceiling or floor effects for any age group), 4 of the words are fairly common, high frequency words (*nurse, dentist, mittens, helicopter*) and 4 are uncommon, low frequency words (*canoe, kayak, trumpet*, *saxophone*). Three sources were used to determine the frequency/commonness of words: the frequency with which adults and children said the words in English CHILDES corpora, the number of web pages that contained the words (as determined by Google searches) and the CDI age of acquisition percentiles.

In theory, children could keep track of which pictures they have pointed to and avoid pointing to them again. If they did this and the test had an equal number of words and pictures, children could get the last words right by the process of elimination. To reduce this possibility, the PAL Receptive Vocabulary test includes 4 distracter pictures that are never named. Distracter pictures are semantically similar to two of the other words (*gloves* for *mittens* and *sandals*; *astronaut* for *nurse* and *doctor*; *canoe* for *helicopter* and *kayak*; and *guitar* for *trumpet* and *saxophone*). Four target words are more similar to one another than either is to the distracter (i.e., *nurse* and *doctor* are more similar to one another than either is to the distracter *astronaut*; *trumpet* and *saxophone* are more similar to one another than either is to *guitar*), and for the other 4 words, the distracter is more similar to one of the target words than the target words are to each other (i.e., *gloves* is more similar to *mitten* than *mittens* is to *sandals*; *canoe* is more similar to *kayak* than *kayak* is to *helicopter*). Two distracter nouns are frequent (*guitar* and *gloves*) and two are infrequent (*astronaut* and *canoe*). All pictures are gray-scale and appear on a single sheet. (See Appendix 2.)

PAL Lexical Access Test. The PAL test uses verbal fluency tasks to assess children's lexical retrieval skills (and, secondarily, their expressive vocabularies). There are two reasons for using verbal fluency tasks. First, typically-developing children's performance on verbal fluency tasks increases with age (e.g., Riva, Nichelli, & Devoti, 2000; Koren, Kofman, & Berger, 2005). Second, performance on verbal fluency tasks has been shown to be a sensitive measure of the lexical access abilities of typically developing children (e.g., Riva, Nichelli, & Devoti, 2000; Koren, Kofman, & Berger, 2005-), spoken language impaired children (e.g., Weckerly, Wulfeck, & Reilly, 2001; Messer & Dockrell, 2006), and dyslexic children (e.g., Levin, 1990; Cohen, Morgan, Vaughn, Riccio, & Hall, 1999), with typically-developing children performing better than children with spoken or written language impairments. The PAL Lexical Access Test consists of two verbal fluency tasks. In the first task, children name as many animals as they can in 30 seconds (Benton & Hamsher, 1977). This test is widely used to assess lexical access in typically-developing children and children with a wide range of neuropsychological deficits (see Messer & Dockrell, 2006 and references therein). In the second task, children have 30 seconds to give an example of a word that fits the verbal description of 10 items (e.g., name a vegetable, name something round, etc.). Children's lexical access score are the average of their scores on the two tasks.

PAL Syntax Test. In the PAL Syntax test, parents say 12 sentences and ask their child to point to which of two pictures matches the sentence. Sentence-picture matching comprehension tasks are widely used in research and clinical settings, yield relatively unambiguous responses that are easy to observe and record, and are arguably the easiest syntactic test to administer to children (see Gerken & Shady, 1996). One limitation of the sentence-picture matching task is that one can only test syntactic constructions whose propositional content are easy to depict clearly and unambiguously.[4]

The PAL Syntax test includes active sentences with reflexive and non-reflexive pronouns because some studies have suggested preschool children who are linguistically normal (e. g., Chien & Wexler,

[4] This means that the sentence-picture matching task is not well-suited for testing children's knowledge of inflectional morphology, an area of language that has long been argued to be a key feature of grammatical development (Brown, 1973), and more recently has been argued to be a core grammatical deficit in children with specific language impairments (see for example, Rice, Wexler, & Cleave, 1995; Rice & Wexler, 1996; Rice, Wexler, & Hershberger, 1997; Leonard, 1998).

1990) and older children with specific language impairment (SLI, e.g., van der Lely & Stollwerck, 1997) sometimes interpret sentences with non-reflexive pronouns as if they had reflexive pronouns. The PAL Syntax test includes passive sentences because passive sentences are harder to understand and produce than active sentences for typically-developing preschool children (see O'Grady, 1997) and older SLI children (e.g., van der Lely & Dewart, 1986; van der Lely, 1996; Leonard, Wong, Deevy, Stokes, & Fletcher, 2006). To minimize the number of items on the test and prevent ceiling or floor effects, children are tested on different sentences depending on their age. Specifically, 3- and 4- year olds receive 4 *by* passive sentences, 4 active sentences with 2 lexical NPs, 2 active sentences with reflexive pronouns, and 2 active sentences with non-reflexive pronouns. Five-year olds receive 6 passive sentences (3 *by* passives, 3 truncated passives), 1 active sentence with 2 lexical NPs, 2 active sentences with non-reflexive pronouns, 2 active sentences with reflexive pronouns, and 1 active sentence without an overt object NP.

Because many studies have shown that children can use non-syntactic cues to interpret sentences (see O'Grady, 1997) and the PAL Syntax test is meant to selectively measure children's syntactic abilities (and not other linguistic or nonlinguistic abilities), care was taken to eliminate non-syntactic cues and confounds. For example, all of the sentences are semantically reversible, insomuch as the agent and the patient can be switched and the resulting sentence is still semantically plausible (e.g., *The pig was kissed by the sheep* and *The sheep was kissed by the pig*). For this reason, all sentences have verbs that are felicitous in active sentences and in verbal passive sentences with animate patients and overt animate agents. All sentences also contain noun phrases that refer to animals, with animals paired such that either animal is equally plausible as the agent of the sentence (e.g., *pig* and *sheep*). The pairs of pictures contain no cues as to which picture in a pair matches a sentence. Specifically, the animals in the (colored) pictures are all drawn in the same cartoon style, and pairs of pictures differ only in which animal is the agent and which is the patient. (See Appendix 3.) Over the course of the test, each animal in each pair is the agent and the patient equally often, the animal that is the agent appears on the left and the right of the patient equally often, and the correct picture is the left and right picture equally often.

The PAL test could fill an important niche in assessment tools available to researchers and clinicians. However, in order for the PAL test to a useful tool, it must provide a valid measure of preschool children's linguistic abilities and it must be a sensitive and specific screening test for language impairments. To determine whether the PAL test exhibits these properties, in the following experiment, children's PAL test scores were compared with their standardized language test scores.

EXPERIMENT

METHOD

Participants. One hundred twenty-four monolingual children between the ages of 36 and 71 months participated in the study. Children were recruited from 7 day care centers and preschools located on the Eastern Shore of Maryland. Two children were eliminated because they did not complete the standardized tests. The remaining 122 children had a mean age of 52.4 months (SD = 10.1 months). Children's ages were adjusted for prematurity, and gestational age (GA)-adjusted ages were used to determine which tests to administer. The children's GA-adjusted age was 52.2 months (SD = 10.3).

All of the children spoke Standard American English and 60 were male (49%) and 62 were female (51%). According to the Center for Disease Control (Martin et al., 2005), 12% of babies born in the US in 2003 were premature (GA < 37 weeks), 8% had low birth weights (< 2500 grams) and 3% were twins. The children in this study reflected these statistics (mean GA = 39.0 +/- 2.0 weeks, mean BW = 3324 +/- 600 grams; 3% twins). Between 7 to 10% of children in the United States have a speech or language impairment (e.g., Tomblin, Smith, & Zhang, 1997; Shriberg, Tomblin, & McSweeny, 1999; National Information Center for Children and Youth with Disabilities, 2000), Thus, the parent-reported incidence of spoken language impairment among the children in this study (4%) was somewhat low.

According to the US Census Bureau, the US population is 69% non-Hispanic white, 13% Hispanic, 12% African American, 4% Asian and 2% Native American, with a language other than English being regularly spoken in 18% of households. Because only monolingual English-speaking children were included, the children in this study reflected the ethnic and racial diversity of monolingual English-speaking children with 94 children (77%) being non-Hispanic white, 20 children (16%) being black, 7 children (6%) being Hispanic and 1 child (< 1%) being Asian. In 2003, the median 4-person family income was $65,000. Thus, economically-speaking, the children in the current study were representative of the US (16% < $25,000; 20% between $25,000 and $49,999; 23% between $50,000 and $74,999; 41% ≥ $75,000). In the US, 85% of people over 25 have graduated from high school and 28% have a bachelor's degree. Thus, the parents in this study were better educated than the US norm with 53% of mothers having a bachelor's degree and 37% of the fathers having a bachelor's degree.

Standardized Tests. In addition to taking the PAL tests, children took two standardized tests, Children's articulation was assessed using the word repetition Denver Articulation Screening Exam (DASE, Drumwright, 1971). The DASE assesses 2.5 to 7 year old children's ability to correctly pronounce 30 consonant or consonant clusters in word initial, medial and final position. Children's syntactic, morphological and lexical abilities were assessed using the 3 core subtests of the second edition of the Clinical Evaluation of Language Fundamentals - Preschool (CELF, Wiig, Secord, & Semel, 2004). The CELF's Expressive Vocabulary subtest is a test of children's expressive vocabulary. In this subtest, children say the words that correspond to objects and actions shown in pictures (e.g., *what is this?* answer: flag; *what is this girl doing?* answer: riding). The Sentence Structure subtest tests children's receptive morphosyntactic abilities by having children point to the picture that matches spoken sentences. The sentences in the Sentence Structure subtest vary in syntactic complexity and structure (e. g., *The boy is sleepy*, *the man who is sitting under the tree is wearing a hat*). The Word Structure subtest uses .a cloze procedure in which children complete a sentence that contains a targeted closed class morpheme (e.g., the *–ing* in *the girl is sleeping*, the pronoun *her* in *he is waving at her*, etc.). For some Word Structure items, the child only has to provide the correct grammatical morpheme (e.g., *he is waving at her*) and in some items, the child must generate both the correct grammatical morpheme and the correct lexical morpheme (e.g., *this is her bike*, *horse-s*).

Testing Procedures. Parents administered the PAL test in their homes. One to 3 days later, a trained experimenter administered the standardized tests individually to each child in his or her daycare center or school, in a quiet room away from other children. All children took the DASE before the CELF, and the person who administered the standardized tests did not know children's PAL test scores.az

Data Treatment. A research assistant hand-scored each PAL test, and entered the scores into a computer database. A second research assistant hand-scored the standardized tests and entered them into a database. The two research assistants then scored and entered the data for the other type of test. Neither research assistant knew which child the data came from, or how the child did on the other type of test. For the "Name Animals" task, repetitions and incorrect responses were excluded, and the number of correct responses was entered into the databases. The inter-coder reliability for this task was 98%, with disagreements being about whether responses were repetitions (e.g., *kitty* and *kitten*) and whether a particular response counted (e.g., *monster*). For all other tasks, the inter-coder reliability was 99% or greater. A third person served as a tiebreaker when the first two coders' scores differed.

In addition to determining each child's scores on the three CELF subtests, following the procedures used by Wiig, Secord and Semel (2004), we calculated each child's composite CELF Core Language Scores (CELF CLS). We also calculated 2 composite PAL scores. The first composite score (PAL Oral 1 score) was the sum of children's scores on all 4 PAL tests. The second composite score (PAL Oral 2 score) was the sum of children's PAL Articulation, Receptive Vocabulary and Syntax scores. PAL Oral 2 scores were calculated because most standardized language tests (including the CELF) do not include a lexical access task and, thus, PAL Oral 2 scores are more similar to composite standardized test scores. Second, because the PAL Lexical Access test is a timed test, it is the most difficult PAL test to administer. Third, the Lexical Access test is the hardest PAL test to score (see Data Treatment).

RESULTS

Correlations between PAL Test Scores and Standardized Test Scores: All children

Table 1 gives children's mean PAL test scores (with SDs). To compensate for the number of correlations performed, the p level for significance was set at $p \leq .01$. When data from all children were combined, 28 of the 30 PAL scores-standardized test scores correlations were significant, with the mean for all 30 correlation coefficients being .48 (median r = .47, see Table 2). Following Cohen (1988), correlation coefficients between .10 and .29 were considered small, correlation coefficients between .30 and .49 were considered medium-sized, and correlation coefficients of .50 or greater were considered large. Using these guidelines, one of the significant PAL test-standardized test correlations was small, 17 were medium-sized and 10 were large.

Table 1: Mean PAL Scores (with Standard Deviations)

	PAL SCORES					
	Articulation	Vocabulary	Lexical Access	Syntax	Oral 1	Oral 2
All Children	10.28 (1.86)	5.63 (1.74)	5.84 (2.28)	8.96 (2.36)	30.45 (5.64)	24.80 (4.12)
Male	10.10 (2.01)	5.40(1.76)	5.56 (2.31)	9.05 (1.84)	30.13 (5.43)	24.61 (4.11)
Female	10.45 (1.69)	5.66 (2.00)	5.80 (2.55)	8.87 (1.77)	30.76 (5.86)	24.98 (4.15)
White	10.21 (1.85)	5.49 (1.79)	5.51 (2.40)	9.02 (1.76)	30.30 (5.33)	24.79 (3.92)
Black	10.37 (1.80)	5.79 (2.10)	6.45 (2.47)	8.89 (1.94)	31.83 (6.14)	25.33 (4.45)
Low SES	10.42 (1.83)	5.47 (1.70)	5.95 (2.20)	8.49 (1.84)**	30.22 (5.67)	24.32 (4.33)
High SES	10.36 (1.77)	5.80 (1.89)	5.57 (2.86)	9.54 (1.62)**	31.29 (5.53)	25.65 (3.68)

** $p < .01$

As shown in Table 2, children's PAL Articulation scores were significantly and moderately correlated with all 5 standardized test scores (mean r = .43, median r = .42), with the PAL Articulation-CELF Sentence Structure correlation (the one non-expressive test) being the smallest (r= .38) and the PAL Articulation-CELF Word Structure and PAL Articulation-CELF CLS correlations being the largest (both r's = .47). PAL Receptive Vocabulary scores were significantly correlated with all CELF scores

Table 2: Correlations Among PAL Test Scores and Standardized Test Scores (All Children)

	PAL SCORES					
	Articulation	Vocabulary	Lexical Access	Syntax	Oral 1	Oral 2
DASE	.42****	.17	.14	.23**	.36****	.39****
CELF Vocabulary	.40****	.55****	.41****	.42****	.66****	.64****
CELF Sentence	.38****	.47****	.45****	.46****	.66****	.62****
CELF Word	.47****	.46****	.42****	.42****	.67****	.65****
CELF CLS	.47****	.57****	.48****	.47****	.74****	.71****
Mean r	.43	.44	.38	.40	.62	.60
Median r	.42	.47	.42	.42	.66	.64

** $p \leq .01$
*** $p < .001$
**** $p < .0001$

(mean r = .51) but not DASE scores. Children's PAL Lexical Access scores were also significantly

correlated with all CELF scores (mean r = .44), but not DASE scores. PAL Syntax scores were significantly correlated with all 5 standardized test scores (mean r = .40. median r = .42, range .23 - .47), with the 4 PAL Syntax- CELF score correlations being medium-sized (mean r = .44) and the PAL Syntax-DASE correlation being small. Lastly, composite PAL Oral scores were significantly correlated with all standardized test scores. PAL Oral 1 scores were highly correlated with the 4 CELF scores (mean r = .68) and PAL Oral 1 scores were moderately correlated with DASE scores. The PAL Oral 1-CELF CLS correlation (i.e., the correlation between composite PAL scores and composite CELF scores) was very high (r = .74). PAL Oral 2 correlations were virtually identical to PAL Oral 1 correlations.

Correlations between PAL Scores and Standardized Test Scores: Boys and Girls

We next analyzed the data for boys and girls separately. Our reasons for doing so were two-fold. First, results of some studies suggest that boys develop language more slowly than girls (see Bornstein, Hahn, & Haynes, 2004 and references therein) and boys are more likely to suffer from language impairments than girls (see Stromswold, 1998 and references therein). Second, studies have generally shown that girls have significantly higher scores than boys on the CDI I, II and III (Fenson et al. 1994, (Feldman et al., 2000; Feldman et al., 2005) and the LDS (Rescorla & Alley, 2001).

As shown in Table 1, boys' and girls' PAL scores did not differ significantly. For boys, 25 of the 30 PAL–standardized test correlations were significant, with the mean for all 30 correlation coefficients being .49 (median r = .50, see Table 3a). Of the 25 significant correlations, 10 were medium-sized and 15 were large. Boys' PAL Articulation scores were significantly correlated with all 5 standardized test scores (mean r = .47, median r = .48), with the PAL Articulation-Word Structure correlation being the lowest (r = .33) and the PAL Articulation-DASE correlation being the highest (r = .57). Boys' PAL Receptive Vocabulary scores were significantly correlated with all CELF scores (mean r = .49), but not DASE scores, and Lexical Access scores were significantly correlated with CELF CLS scores (r = .34), but not CELF subtest or DASE scores. Boys' PAL Syntax scores were significantly correlated with all 5 standardized test scores (mean and median r = .46), with the PAL Syntax-DASE correlation being the lowest (r = .35) and the PAL Syntax- CELF Sentence Structure correlation being the highest (r = .53). Lastly, boys' composite PAL scores were highly correlated with all standardized test scores, with PAL Oral-DASE correlations (r = .52) being lower than PAL Oral-CELF subtest correlations (all r's between .63 and .70), which in turn were lower than PAL Oral – CELF CLS correlations (r's = .73 and .74).

Table 3a: Correlations Among PAL Test Scores and Standardized Test Scores (Boys)

	PAL SCORES					
	Articulation	Vocabulary	Lexical Access	Syntax	Oral 1	Oral 2
DASE	.57****	.30	.18	.35**	.52****	.52****
CELF Vocabulary	.46***	.53****	.32	.44***	.69****	.70****
CELF Sentence	.48****	.45***	.31	.53****	.63****	.63****
CELF Word	.33**	.43***	.30	.46***	.66****	.66****
CELF CLS	.49****	.54****	.34**	.50****	.73****	.74****
Mean r	.47	.45	.29	.46	.65	.65
Median r	.48	.45	.31	.46	.66	.66

** $p \leq .01$
*** $p < .001$
**** $p < .0001$

For the girls, 23 of the 30 PAL – standardized test score correlations were significant, with the mean for all 30 correlation coefficients being .47 (median r = .49). Of the 23 significant correlations, 9 were medium-sized and 14 were large (see Table 3b). In addition to the girls' mean and median correlation coefficients being similar to the boys', the girls' correlations between composite PAL scores

and composite CELF scores (PAL Oral 1-CELF CLS r = .76, PAL Oral 2-CELF CLS r = .68) were very similar in size to the boys'. Furthermore, individual PAL-standardized test correlation coefficients were generally similar for girls and boys. However, in 5 cases (PAL Articulation-DASE, PAL Articulation-CELF Expressive Vocabulary, PAL Syntax-DASE, and PAL Oral 1-DASE, PAL Oral 2-DASE), PAL test - standardized test correlations were statistically significant for boys but not girls, and in 4 cases (PAL Lexical Access-CELF Expressive Vocabulary, PAL Lexical Access-CELF Sentence Structure, PAL Lexical Access-CELF Word Structure, PAL Lexical Access-CELF CLS), girls' correlations were statistically significant and boys' correlations were not.

Table 3b: Correlations Among PAL Test Scores and Standardized Test Scores (Girls)

	PAL SCORES					
	Articulation	Vocabulary	Lexical Access	Syntax	Oral 1	Oral 2
DASE	.24	.01	.11	.14	.23	.27
CELF Vocabulary	.27	.56****	.53****	.43***	.64****	.57****
CELF Sentence	.45***	.49****	.63****	.38**	.70****	.63****
CELF Word	.47****	.49****	.54****	.38**	.68****	.63****
CELF CLS	.43***	.60****	.66****	.45***	.76****	.68****
Mean r	.37	.43	.49	.36	.60	.56
Median r	.43	.49	.54	.38	.68	.63

** $p \leq .01$
*** $p < .001$
**** $p < .0001$

Correlations between PAL and Standardized Test Scores: African American and White Children

Some studies suggest that parents' race has a significant effect on the scores they give their children on parent-completed language questionnaires. For example, Rescorla and Achenbach (2002) found that African American children had lower LDS vocabulary and phrase length scores than non-Hispanic white children even when the effect of SES was partialled out, Roberts, Burchinal, and Durham (1999) found that the mean CDI expressive vocabulary score for 30 month old African American children was lower than expected (mean score = 31st percentile), and Feldman et al. (2000) found that, although parents' race didn't affect children's CDI II expressive vocabulary scores, African American parents gave their children lower scores on sentence measures and higher over-regularization scores than white parents.

The observation that parents' race may affect the scores they give their children on language questionnaires, with the size and/or direction of the effect varying according to the questionnaire used and the aspect of language studied, raises the concern that the PAL test might not be equally valid for children with different racial backgrounds, even among children who speak standard American English (as was the case for all of the children in this study). For this reason, we analyzed the data from non-Hispanic white children (N = 96) and African American children (N = 19) separately. As shown in Table 1, African American and non-Hispanic white children's PAL scores did not differ significantly.

For the non-Hispanic white children, 27 of the 30 PAL - standardized test correlations were significant, with the mean for all 30 correlation coefficients being .41 (median r = .39). Of the significant correlations, one was small, 18 were medium-sized and 8 were large (see Table 4a). All of the correlations between PAL Articulation scores and standardized test scores were significant, with *r's* ranging from .32 to .39. White children's PAL Receptive Vocabulary scores were moderately and significantly correlated with all CELF scores (mean r = .43), but not DASE scores. PAL Lexical Access scores were also moderately and significantly correlated with all CELF scores (mean r = .42), but not DASE scores. Similarly, white children's PAL Syntax scores were moderately and significantly correlated with all CELF scores (mean r = .36) but not DASE scores. PAL Oral 1 scores were significantly correlated with all standardized test scores, with all CELF correlations being large (mean r =

.63) and the DASE correlation being small (r = .28). Similarly, white children's PAL Oral 2 scores were significantly correlated with their scores on all standardized tests, with all CELF correlations being large (mean r = .57) and the DASE correlation being small (r = .30).

Table 4a: Correlations Among PAL and Standardized Test Scores (Non-Hispanic White Children)

	PAL SCORES					
	Articulation	Vocabulary	Lexical Access	Syntax	Oral 1	Oral 2
DASE	.37***	.06	.07	.16	.28**	.30**
CELF Vocabulary	.33**	.44****	.35***	.34***	.56****	.52****
CELF Sentence	.32**	.39****	.46****	.36***	.64****	.56****
CELF Word	.38****	.39****	.41****	.36***	.64****	.58****
CELF CLS	.39****	.48****	.46****	.39****	.69****	.62****
Mean r	.36	.35	.35	.32	.56	.52
Median r	.37	.39	.41	.36	.64	.56

** $p \leq .01$
*** $p < .001$
**** $p < .0001$

Overall, the African American children's PAL test-standardized test correlation coefficients were 1.5 times larger than the white children's correlation coefficients (African American mean r = .62, median r = .65). Despite the African American children's correlation coefficients being larger, the fact that there were 5 times as many white children as African American children meant that fewer correlations were significant for African American children (17 out of 30), with all 17 significant correlations being .60 or greater. With respect to the number of correlations that were significant, the most extreme difference was that 4 out of 5 of the white children's PAL Lexical Access-standardized test correlations were significant, whereas none of the African American's PAL Lexical Access-standardized test correlations were significant. With respect to the size of correlations, the most extreme difference between white and African American children was that the PAL Receptive Vocabulary – standardized test correlations were about twice as great for the African American children as the white children.

Table 4b: Correlations Among PAL and Standardized Test Scores (African American Children)

	PAL SCORES					
	Articulation	Vocabulary	Lexical Access	Syntax	Oral 1	Oral 2
DASE	.60**	.39	.15	.24	.44	.49
CELF Vocabulary	.63***	.85****	.48	.53	.79****	.79****
CELF Sentence	.49	.73***	.42	.67**	.73***	.74***
CELF Word	.76****	.77****	.47	.47	.80****	.81****
CELF CLS	.71***	.87****	.48	.57	.84****	.85****
Mean r	.64	.72	.40	.50	.72	.74
Median r	.63	.77	.47	.53	.79	.79

** $p \leq .01$
*** $p < .001$
**** $p < .0001$

Correlations between PAL and Standardized Test Scores: High and Low SES Children

Some studies suggest that high socio-economic status (SES, Largo, Pfister, Molinari, Kundu, & al., 1989; Vohr, Garcia-Coll, & Oh, 1989; Landry, Smith, & Swank, 2002; Hoff, 2003) is associated with faster language development. Consistent with this, in some studies, high SES parents gave higher

expressive vocabulary scores than low SES mothers on the LDS (Rescorla, 1989) and the CDI (Arriaga, Fenson, Cronan, & Pethick, 1998). In addition, in some studies, high SES parents gave their children higher sentence scores on the CDI II and III (Feldman et al., 2000; Feldman et al., 2005). However, some studies have reported that lower SES parents gave their children higher vocabulary scores on the CDI I (Fenson et al., 1994; (Feldman et al., 2000) and CDI II (Feldman et al., 2005),

If, as these studies suggest, SES affects parents' reports of their child's language, the PAL test might not be valid for some SES children. To determine whether this was the case, we divided the children into high and low SES groups, and analyzed each group's data separately. SES was calculated by summing the mother's educational level (on a 5 point scale: not a high school graduate, high school graduate, some post-secondary technical training or education, college degree, post-BA education), father's educational level (also on a 5 point scale) and family income (on a 4 point scale, < \$25K, \$25K - \$50K, \$50K-75K, ≥ \$75K). This yielded SES scores that ranged from a low of 3 to a high of 14. Some parents declined to provide information about maternal education, paternal education, and/or family income, and these children were excluded from SES analyses. The remaining children were classified as low SES if their SES scores were 8 or lower (N = 38) and high SES if their SES scores were 9 or higher (N = 45). High and low SES children's PAL scores were only significantly different on the PAL Syntax test, with high SES children's scores being significantly greater than low SES children's scores ($F(1, 78) = 7.35$, $p = .008$, see Table 1). This result is consistent with studies that have found that higher SES mothers gave their children higher CDI sentence scores (Feldman et al., 2000; Feldman et al., 2005). Although the SES effect on PAL Syntax scores could merely reflect parental biases, the fact that CELF Sentence Structure scores were the only standardized test scores for which there was a significant effect of SES (11.90 for low SES children, 15.00 for high SES children, $F(1, 81) = 7.96$, $p = .006$) suggests that the effect of SES on syntax may be real.

For the low SES children, 27 of the 30 PAL - standardized test score correlations were significant, with the mean and median correlation coefficient for all 30 correlations being .50 (see Table 5a). Low SES children's PAL Articulation scores were significantly correlated with all standardized test scores, with all but the PAL Articulation – CELF Sentence Structure correlation being large. Low SES children's PAL Receptive Vocabulary scores were significantly correlated with all CELF scores (mean $r = .46$), but not DASE scores. PAL Lexical Access scores were significantly correlated with CELF Sentence Structure, Word Structure and CLS scores, but not CELF Expressive Vocabulary or DASE scores. Low SES children's PAL Syntax scores were significantly correlated with all CELF scores (mean $r = .44$), but not DASE scores. PAL Oral 1 scores were significantly correlated with all standardized test scores, with all but the PAL Oral 1 – DASE score correlation being large. Lastly, low SES children's PAL Oral 2 scores were highly correlated with all standardized test scores.

Table 5a: Correlations between PAL Standardized Test Scores (Low SES Children)

	PAL SCORES					
	Articulation	Vocabulary	Lexical Access	Syntax	Oral 1	Oral 2
DASE	.51***	.32	.13	.34	.47**	.53***
CELF Vocabulary	.51***	.53***	.30	.43***	.67****	.66****
CELF Sentence	.43***	.37**	.49***	.43***	.61****	.52***
CELF Word	.60****	.40**	.41**	.43***	.65****	.62****
CELF CLS	.61****	.54***	.43***	.47***	.74****	.70****
Mean *r*	.53	.43	.35	.42	.63	.61
Median *r*	.50	.40	.41	.43	.65	.62

** $p \leq .01$
*** $p < .0001$
**** $p < .0001$

For the high SES children, 21 of the 30 PAL – standardized test correlations were significant, with

the mean correlation coefficient for all 30 correlations being .43 (median r = .46). Of the 21 significant correlations, 9 were medium-sized and 12 were large (see Table 5b). High SES children's correlations were generally lower than low SES children's, with the difference being particularly pronounced for PAL-DASE correlations and PAL Syntax-standardized test score correlations, where low SES children's r were generally 2 to 3 times higher than high SES children's. With respect to which correlations were significant, high SES children differed from low SES children in three ways. First, the correlations between DASE scores and PAL Articulation, Oral 1 and Oral 2 scores were only significant for low SES children. Second, the correlations between PAL Syntax scores and CELF Expressive Vocabulary, Sentence Structure and CLS scores were only significant for low SES children. Third, the PAL Lexical Access- CELF Expressive Vocabulary correlation was only significant for high SES children.

Table 5b: Correlations between PAL and Standardized Test Scores (High SES Children)

	PAL SCORES					
	Articulation	Vocabulary	Lexical Access	Syntax	Oral 1	Oral 2
DASE	.14	-.04	.01	.12	.11	.13
CELF Vocabulary	.36**	.50***	.39**	.21	.62****	.60****
CELF Sentence	.44***	.59****	.48***	.27	.72****	.67****
CELF Word	.39**	.58****	.53***	.36**	.76****	.69****
CELF CLS	.42**	.59****	.49***	.28	.73****	.69****
Mean r	.35	.44	.38	.25	.59	.56
Median r	.39	.58	.48	.27	.72	.67

** $p \leq .01$
*** $p < .0001$
**** $p < .0001$

The PAL Tests as a Screening Test for Speech and Language Impairments

To assess whether a language screening test is clinically useful, one needs to determine whether the children who score poorly on the test are the ones who are actually language impaired. If a test is a perfect screening test, all and only the children who score poorly on the test will be diagnosed as language impaired. In other words, the test will have no false positives (normal children whose low test scores erroneously result in them being classified as language impaired) and no false negatives (language impaired children whose high test scores erroneously result in them being classified as not impaired). In performing analyses to assess the validity of a new screening test, it is important to set the level of clinical impairment correctly because false negative rates will be artificially elevated if too few children are labeled language impaired and false positive rates will be artificially elevated if too few children are labeled language impaired. In order to test the validity of a language screening test, one needs to know which children are language impaired and which are not. In practice, it is not possible to know this with certainty and one must rely on a measure that – while not perfect – is considered a 'gold standard.'

PAL Oral 1 scores. With respect to language impairments, the gold standard is a valid, reliable language test administered by a speech-language pathologist under optimal conditions. For this study, we used CELF CLS scores as our gold standard for language impairment. Children were classified as language impaired if their CELF CLS scores were in the bottom 10% of scores obtained by the children in the study. We chose this threshold for 3 reasons. First, in epidemiological studies, 10% is the upper bound for the rate of language impairment among US children. Second, in clinical settings, over-referral is generally more desirable than under-referral, especially when a screening test is inexpensive and safe. Third, the bottom 10% of children had CELF CLS scores that were $\geq$ 1.5 SD below the study's mean.

We converted children's continuous scores on our measure of language impairment (CELF CLS) into the binary values of language impaired ($\leq 10^{th}$ percentile) and unimpaired ($> 10^{th}$ percentile). We then calculated the rate of true positives (the percent of language impaired children who failed the PAL, i.e., the sensitivity of the PAL) and the rate of true negatives (the percent of unimpaired children who

passed the PAL, i.e., the specificity of the PAL), the Positive Predictive Value (PPV, the percent of children who failed the PAL who were language impaired), and the Negative Predictive Value (NPV, the percent of children who passed the PAL who were unimpaired) for different PAL Oral 1 scores cutoffs.

When children with PAL Oral 1 scores in the bottom 2.5th percentile were considered language impaired, the sensitivity of PAL Oral 1 scores was 25% and the specificity was 100% (PPV = 100%, NPV = 97%). When children with PAL Oral 1 scores in the bottom 5th percentile were considered language impaired, the sensitivity of PAL Oral 1 scores was 42% and the specificity was 99% (PPV = 83%, NPV = 99%). When children with PAL Oral 1 scores in the bottom 10th percentile were considered language impaired, the sensitivity (and PPV) of PAL Oral 1 scores was 67% and the specificity (and NPV) was 96%. When children with PAL Oral 1scores in the bottom 15th percentile were considered language impaired, the sensitivity of PAL Oral 1 scores was 75.0% and the specificity was 91.3% (PPV = 50.0%. NPV = 96.9%). Finally, all of the language-impaired children had PAL Oral 1 scores below the 25th percentile (sensitivity = 100%, specificity = 86.5%, PPV = 46.2%, NPV = 100%).

As shown in Figure 1a, a Receiver Operating Characteristic (ROC) curve was generated by plotting PAL Oral 1 scores' true positive rate (i.e. its sensitivity) against its false positive rate (i.e., 1 – specificity). The area under a ROC curve (AUC) provides a measure of the overall goodness of a test. Following Hosmer and Lemeshow (2000), an AUC between 0.70 and 0.80 was considered good and an AUC of 0.80 or greater was considered excellent. The AUC for PAL Oral 1 scores was .96, indicating that PAL Oral 1 scores did an excellent job discriminating between children who were and were not language impaired.

PAL Oral 2 scores. We repeated the procedure, this time using PAL Oral 2 scores. When children with PAL Oral 2 scores in the bottom 2.5th percentile were considered language impaired, the sensitivity of PAL Oral 2 scores was 17% and the specificity was 99% (PPV = 67%, NPV = 91%). When children with PAL Oral 2 scores in the bottom 5th percentile were considered language impaired, the sensitivity of PAL Oral 2 scores was 33% and the specificity was 98% (PPV = 67%, NPV = 93%). When children with PAL Oral 2 scores in the bottom 10th percentile were considered language impaired, the sensitivity (and PPV) of PAL Oral 2 scores was 42% and the specificity (and NPV) was 93%. When children with PAL Oral 2 scores in the bottom 15th percentile were considered language impaired, the sensitivity of PAL Oral 2 scores was 75% and the specificity was 91% (PPV = 50%, NPV = 97%). Eight-three percent of language-impaired children had PAL Oral 2 scores below the 25th percentile (sensitivity = 83%, specificity = 81%, PPV = 33%, NPV = 98%), and 92% had PAL Oral 2 scores below the 30th percentile (sensitivity = 92%, specificity = 76%, PPV = 31%, NPV = 99%). The AUC for PAL Oral 2 scores was .92, indicating that, like PAL Oral 1 scores, PAL Oral 2 scores were excellent at distinguishing between children who did and did not suffer from language impairments (see Figure 1b).

PAL Receptive Vocabulary and Syntax. Given that CELF CLS scores don't include a measure of children's articulation, one might predict that children's combined scores on the PAL Receptive Vocabulary and PAL Syntax tests would do a better job than PAL Oral scores at distinguishing between children with normal and low CELF CLS scores. To test this prediction, we repeated the procedure outlined above, this time using children's combined scores on the PAL Receptive Vocabulary and PAL Syntax tests. Again children with CELF CLS scores in the bottom 10th percentile were considered language impaired. When children with combined PAL Expressive Vocabulary and Syntax scores in the bottom 2.5th percentile were considered language impaired, the sensitivity of combined PAL scores was 8% and the specificity was 98% (PPV = 33%, NPV = 90%). When children with combined scores in the bottom 5th percentile were considered language impaired, the sensitivity was 17% and the specificity was 96% (PPV = 33%, NPV = 90%). When children with combined PAL scores in the bottom 10th percentile were considered language impaired, the sensitivity (and PPV) of combined PAL scores was 33% and the specificity (and NPV) was 92%. When children with combined PAL scores in the bottom 15th percentile were considered language impaired, the sensitivity was 67 % and the specificity was 90% (PPV = 44%, NPV = 96%). As was the case with PAL Oral 2 scores, 83% of language-impaired children had combined PAL Receptive Vocabulary and Syntax scores in the bottom 25th percentile (sensitivity = 83%, specificity = 81%, PPV = 33%, NPV = 98%) and 92% had combined PAL scores that were below the 30th percentile (sensitivity = 92%, specificity = 76%, PPV = 31%, NPV = 99%).

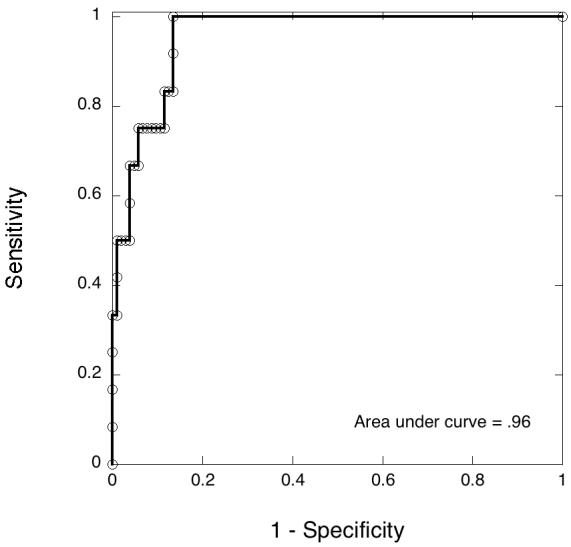


Figure 1a: PAL Oral 1 Scores

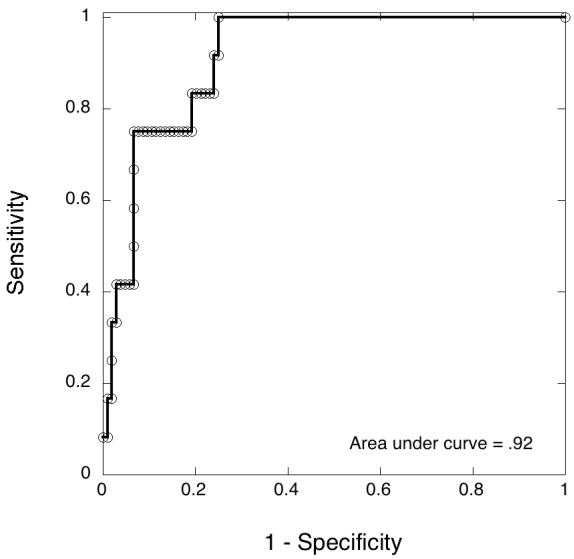


Figure 1b: PAL Oral 2 Scores

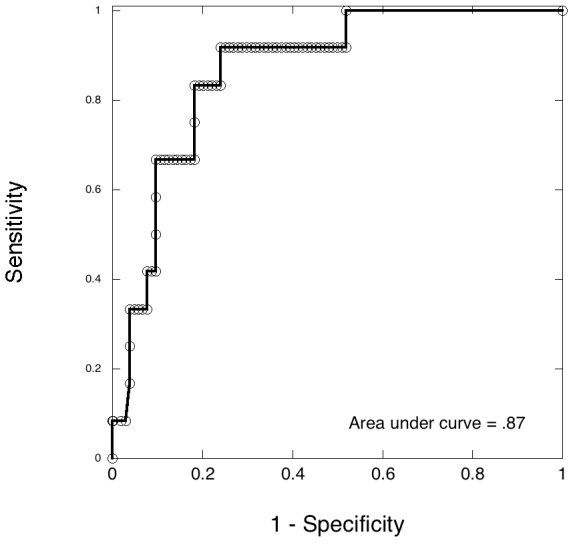


Figure 1c: PAL Expressive Vocabulary + Syntax Scores

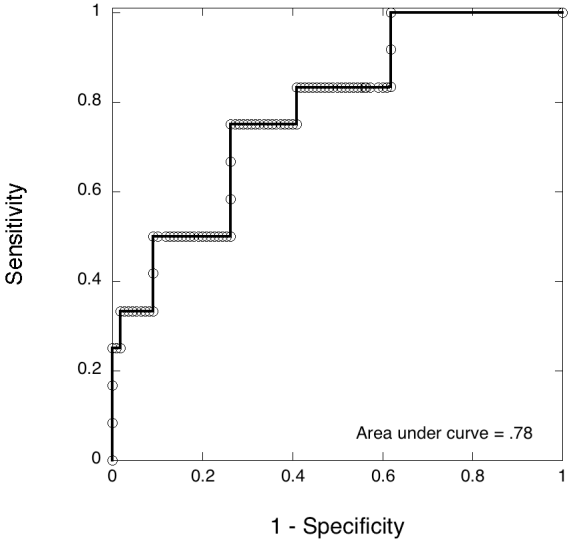


Figure 1d: PAL Articulation Scores

Figure 1: Receiver Operating Characteric (ROC) Curves for PAL Test Scores

The AUC for combined PAL Receptive Vocabulary and Syntax scores was .87 (see Figure 1c), indicating that combined PAL Receptive Vocabulary and Syntax scores were excellent at identifying children with low CELF CLS scores. At first blush, it is somewhat surprising that the AUC for combined PAL Receptive Vocabulary and Syntax scores was slightly less than the AUC's for PAL Oral 1 and 2 scores. One plausible reason is that two of the three CELF subtests are expressive (i.e., they have an articulatory component), whereas the PAL Receptive Vocabulary and Syntax tests are both receptive. Thus, because PAL Oral scores include PAL Articulation scores, PAL Oral scores might have been better at identifying children with low CELF CLS scores than combined PAL Vocabulary and Syntax scores.

PAL Articulation Scores. To investigate whether the PAL Articulation test was a good screening

test for speech impairments, we repeated the procedures outlined above. Children were classified as speech-impaired if their DASE scores were in the bottom 10% of the scores obtained by the children in the study. Very low PAL Articulation scores were an very specific indicator that a child was speech impaired (bottom 2.5th percentile sensitivity = 25%, specificity = 100%, PPV = 100%, NPV = 92%; bottom 5th percentile sensitivity = 33%, specificity = 98%, PPV = 67%, NPV = 93%). For PAL Articulation scores between the 5th and 25th percentile, increases in sensitivity were coupled with proportional decreases in specificity (bottom 10th percentile sensitivity = 33%, specificity = 93%; bottom 15th percentile sensitivity = 50%, specificity = 89%; bottom 25th percentile sensitivity = 50%, specificity = 78%). Seventy-five percent of speech-impaired children had PAL Articulation scores that were in the bottom third (sensitivity = 75%, specificity = 74%, PPV = 24%, NPV = 96%) and 83% had PAL Articulation scores in the bottom half (sensitivity = 83%, specificity = 55%, PPV = 16%, NPV = 97%). The AUC for the PAL Articulation ROC curve was .78 indicating that, overall, PAL Articulation scores were good - but not excellent - at identifying children with speech impairments (see Figure 1d).

Specificity of the PAL and Standardized Tests

The PAL test was designed with the intent that each (sub)test would selectively measure a specific linguistic skill. If PAL Articulation, Receptive Vocabulary, Lexical Access and Syntax scores are highly intercorrelated, this could indicate that the linguistic skills assessed by these tests are highly interdependent, that the PAL tests do not selectively assess specific linguistic abilities, and/or that parents' beliefs about their child's overall linguistic ability (consciously or unconsciously) result in them giving similar scores on the 4 tests. Inspection of Table 6a reveals the correlations between PAL tests – although significant – were small to medium-sized (mean r = .32, median r = .31), suggesting that these possibilities have modest effect. The lowest correlation was between PAL Articulation and Lexical Access scores, which suggests that the skills tapped by these tests have relatively little overlap. The highest correlation was between PAL Receptive Vocabulary and Lexical Access scores, which is consistent with these tests both measuring (different aspects of) children's lexical abilities.

Table 6a: Correlations Among Scores on PAL Subtests

	PAL Receptive Vocabulary	PAL Lexical Access	PAL Syntax
PAL Articulation	.35****	.24**	.31****
PAL Receptive Vocabulary	–	.38****	.29***
PAL Lexical Access	-	–	.32****

** $p \leq .01$
*** $p < .001$
**** $p < .0001$

Table 6b: Correlations Among Standardized Test Scores

	CELF Expressive Vocabulary	CELF Sentence Structure	CELF Word Structure
DASE	.28**	.34****	.26**
CELF Expressive Vocabulary	-	.68****	.70****
CELF Sentence Structure	-	-	.72****

** $p \leq .01$
*** $p < .001$
**** $p < .0001$

Like the PAL tests, the DASE and the CELF subtests are meant to assess different aspects of

children's linguistic development, with the DASE providing a measure of children's articulation, the CELF Expressive Vocabulary subtest a measure of their lexical abilities, the CELF Word Structure subtest a measure of their morphological abilities, and the CELF Sentence Structure subtest a measure of their syntactic abilities. As shown in Table 6b, the correlations between children's DASE scores and CELF subtest scores were small to medium-sized (mean r = .29, median r = .24), suggesting that the DASE test provides a fairly specific measure of children's articulation. In contrast, children's scores on the CELF subtests were all highly inter-correlated (mean and median r = .70). This could indicate that children's lexical, morphological and syntactic abilities are highly interdependent. However, the fact that the correlations among PAL test scores were small to medium-sized suggests this isn't the explanation. Although the experimenter might have unconsciously given children similar scores on their CELF subtests, the fact that DASE – CELF correlations ranged from .26 to .34 argues against this explanation.

A third possibility is that each CELF subtest places substantial demands on the linguistic abilities that the other two subtests are designed to measure. Consistent with each of the CELF subtests tapping multiple linguistic skills, all CELF subtest scores were most highly correlated with PAL composite scores (i.e., PAL Oral 1 and 2 scores), with all of the correlation coefficients being .62 or greater. What shared skills might account for the high correlations among CELF subtest scores? The high correlation between CELF Expressive Vocabulary and CELF Word Structure scores could reflect that both subtests tap children's articulation abilities (both are expressive) and lexical abilities (both require that children determine the noun or verb depicted in drawings). The high correlation between CELF Word Structure and Sentence Structure scores probably reflects that both subtests tap children's morphosyntactic abilities (with the need to determine the nouns and verbs depicted in drawings playing a lesser role).

Articulation. To the extent that the PAL and standardized (sub)tests do, in fact, selectively measure different aspects of children's linguistic abilities, children's scores on each PAL test should be most highly correlated with their scores on the standardized test that taps the most similar linguistic abilities, and children's scores on each standardized subtest should be most highly correlated with their scores on the PAL test that taps the most similar linguistic abilities. Therefore, we predicted that the children's PAL Articulation scores would be more highly correlated with their DASE scores than their CELF subtest scores, and children's DASE scores would be more highly correlated with their PAL Articulation scores than their scores on other PAL subtests. Although children's PAL Articulation scores were, indeed, significantly correlated with their DASE scores (r = .42), PAL Articulation scores were also significantly correlated with all 3 CELF subtest scores (all r's between .38 to .47). The correlations between PAL Articulation and CELF Word Structure and between PAL Articulation and CELF Expressive Vocabulary likely reflect, at least in part, that all subtests require children to say words. That the PAL Articulation-CELF Word Structure correlation is somewhat higher than the PAL Articulation- CELF Expressive Vocabulary correlation (r = .47 and .40, respectively) could reflect that, in the Word Structure subtest, children must correctly say consonants and unstressed syllables, whereas in the Expressive Vocabulary they say phonologically salient free-standing words. It seems unlikely that the correlation between PAL Articulation and CELF Sentence Structure scores reflects specific shared task demands of the two subtests. Plausibly it could reflect characteristics about the child (e.g., the child's willingness to participate in an experiment) or that different linguistic (and nonlinguistic) skills are not completely independent. We predicted children's DASE scores would be most highly correlated with their scores on the PAL Articulation test. Consistent with this, children's DASE scores were only correlated with scores on two of the 4 PAL tests, with the DASE-PAL Articulation correlation coefficient being about twice as large as the DASE-PAL Syntax correlation coefficient.

Vocabulary. Given that the CELF Expressive Vocabulary and the PAL Receptive Vocabulary tests assess the size of children's vocabularies, we predicted that children's PAL Receptive Vocabulary scores would be most highly correlated with their CELF Expressive Vocabulary scores, and that children's CELF Expressive Vocabulary scores would be more highly correlated with their PAL Receptive Vocabulary scores. Both predictions were borne out with the PAL Receptive Vocabulary-CELF Expressive Vocabulary correlation coefficient being .55. Plausibly, the significant (but smaller) correlations between scores on the PAL Receptive Vocabulary test and the other two CELF subtests

reflect that all CELF subtests require that children determine what nouns and verbs are depicted in drawings. The smaller (but significant) correlations between scores on the CELF Expressive Vocabulary subtest and the other 3 PAL tests could reflect that, like the CELF Expressive Vocabulary test, the PAL test is expressive (PAL Articulation, PAL Lexical Access), the PAL test has a substantial lexical component (PAL Lexical Access), or that the PAL test involves pictures (PAL Syntax).

Lexical Access. Contrary to our prediction that PAL Lexical Access scores would be more highly correlated with CELF Expressive Vocabulary scores than scores on the other two CELF subtests, the correlations between scores on the PAL Lexical Access test and the three CELF subtests were all about the same (all *r*'s between .41 and .45), and about 2.5 times greater than the PAL Lexical Access-DASE correlation. Plausibly, this could reflect that, like the PAL Lexical Access test, the CELF subtests all have a strong lexical component, whereas the DASE test does not. Given that the PAL Lexical Access test requires children to say words items, one might expect that that it would place high demands on children's articulatory abilities and, therefore, that children's PAL Lexical Access and DASE scores would be highly correlated. The fact that the PAL Lexical Access-DASE correlation was not significant suggests that, whatever other flaws it may have, the PAL Lexical Access test provides a measure of children's lexical access abilities that is not 'contaminated' by their articulatory abilities.

Morphosyntax. The correlations between scores on the PAL Syntax test and the three CELF subtests were also all about the same (all *r*'s between .42 and .46), and about twice as high as the PAL Syntax-DASE correlation. That the PAL Syntax-CELF Sentence Structure and PAL Syntax-CELF Word Structure correlations were about the same probably reflects that all 3 tests assess children's morphosyntactic abilities. One might have expected that the PAL Syntax-CELF Sentence Structure correlation would be higher than the PAL Syntax-CELF Word Structure correlation because the PAL Syntax test and CELF Sentence Structure subtest both use a picture-matching task, whereas the CELF Word Structure subtest does not. However, the effect of sharing a task may be mitigated by the fact that there is greater morphosyntactic overlap for the PAL Syntax and CELF Word Structure tests (4 out of 12 PAL Syntax items assess children's knowledge of pronouns, as do 7 out of 20 CELF Word Structure items), than the PAL Syntax and CELF Sentence Structure tests (depending on the age of the child, 4 or 6 PAL Syntax sentences are passives, and 2 out of 22 CELF Sentence Structure sentence are passives). It is puzzling that the correlation between PAL Syntax and CELF Expressive Vocabulary scores is as high as the correlations for the other two CELF subtests. This could reflect that both the PAL Syntax and CELF Expressive Vocabulary tests require that children scan drawing and determine the lexical items depicted in the drawings. It could also reflect that CELF Expressive Vocabulary scores are more reliable (test-retest $r = .88$) than CELF Word Structure scores (test-retest $r = .80$) or CELF Sentence Structure scores (test-retest $r = .78$). In other words, the correlations between PAL Syntax scores and CELF Word and Sentence Structure scores could be depressed by the fact that 40% of the variance in CELF Sentence Structure scores is essentially noise ($1- (.78)^2 = .40$) and 36% of the variance in CELF Word Structure scores is noise (as compared to 22% for CELF Expressive Vocabulary scores).

DISCUSSION

Limitations of the PAL Test or Limitations of the Standardized Tests?

Taken as a whole, the results presented above suggest that composite PAL scores are accurate measures of children's general linguistic abilities, with PAL Oral 1 scores accounting for 55% of the variance in CELF CLS scores and PAL Oral 2 scores accounting for 50% of the variance in CELF CLS scores. These figures are particularly impressive because the CELF CLS test-retest *r* is .90. This means that 19% of the variance in CELF CLS scores is noise (.9 x .9 = .81). Thus, one could argue that PAL Oral 1 scores account for 68% of the 'real' variance in CELF CLS scores (.55/.81) and PAL Oral 2 scores account for 62% of the real variance in CELF CLS scores. The results presented above indicate that PAL Oral 1 and 2 scores are remarkably good at distinguishing between children with high and low CELF CLS scores and, thus, that the PAL test may be an excellent screening test for preschool language impairments.

Given that scores on each of the PAL tests were significantly correlated with scores on each of the CELF subtests, one concern is that scores on individual PAL tests may not be measures of specific linguistic abilities. As discussed above, this probably has more to do with the CELF than the PAL test. First, the high intercorrelations among CELF subtest scores coupled with close examination of the CELF subtests indicate that each of the CELF subtests taps multiple, overlapping linguistic skills. Second, as discussed above, the test-retest reliability of CELF subtests vary, and this may have caused some PAL-CELF subtest correlations to be depressed relative to others.

A second, more serious, concern has to do with the validity of PAL Articulation scores as measures of children's articulation abilities and as a way of identifying children with speech impairments. Recall that, although highly significant (and fairly large by the standards of behavioral studies), the correlation between PAL Articulation scores and DASE scores was only .42, even though the procedures used in the two tests were very similar. It could be that the PAL Articulation subtest is only a moderately good measure of children's articulation. A second possibility is that the PAL Articulation – DASE correlation coefficient is depressed because the two tests assess different aspects of children's articulation. Specifically, the DASE tests children's articulation of initial, medial and final consonants and consonant clusters in monosyllabic and bisyllabic words, whereas the PAL Articulation test only assesses children's articulation of onsets in monosyllabic words. The specific consonants and clusters tested also differ with 8 of the 12 sounds that appear on the PAL 3 also appearing on the DASE, 5 out of 12 sounds on the PAL 4 also appearing on the DASE, and 3 out of 12 sounds on the PAL 5 also appearing on the DASE. It is unlikely, however, that differences in the stimuli are the complete explanation because a post hoc analysis revealed that the PAL 5 Articulation – DASE correlation was highly significant (r = .72).

Although either reason is possible, we believe that the reason that the PAL Articulation – DASE correlation isn't higher has to do with limitations of DASE scores as measures of the articulatory abilities of non-minority, high SES children who do not have speech impairments. The apparent validity of a test will be depressed if it is compared with another test that is flawed. Consider the hypothetical case in which PAL Articulation scores are perfectly accurate measures of all children's articulation abilities and DASE scores are poor measures of some groups of children's articulation abilities. In such a case, the correlations between PAL Articulation and DASE scores will be low, even though PAL scores measure children's articulation perfectly. Similarly, the apparent validity of the PAL Articulation test as a screening test for articulation disorders will be depressed to the extent that the DASE misclassifies speech-impaired children as unimpaired or unimpaired children as speech impaired.

The way the DASE was designed may mean that DASE scores are not good measures of typically-developing, high SES, and non-minority children's articulatory abilities. The DASE was specifically designed to "reliably differentiate between normal and abnormal development in a disadvantaged population of children" (Drumwright, Van Natta, Camp, Frankenburg, & Drexler, 1973, p. 5). To this end, Drumwright et al. (1973) tested 1,455 very low SES children (one third non-Hispanic white, one third Hispanic and one third African American) and selected 30 sounds in 22 words that had "regular maturational patterns" for each of these groups and that were said correctly by 70% of children in each group by age 6 (Drumwright, Van Natta, Camp, Frankenburg, & Drexler, 1973, p. 8). In other words, the DASE was specifically designed as a screening test for low SES children, with particular attention paid to African American and Hispanic children. In contrast, the children in our study were economically and racially representative of the US. Consistent with the DASE being a better test for children who are lower SES or African American, the PAL Articulation – DASE correlation coefficients were substantially larger for low SES children (r = .51) than high SES children (r = .14) and for African American children (r = .60) than non-Hispanic white children (r = .37). Consistent with the DASE being a screening test for speech impairments (and not a way of precisely quantifying children's articulatory abilities), overall, the PAL Articulation test proved to be a good screening test for DASE-defined speech problems (AUC = .78), with the PPV being excellent for very low PAL Articulation scores (scores < 10th percentile) and the NPV being very good for higher PAL Articulation scores (scores > 30th percentile).

Comparing the Validity of PAL Test and Parent-completed Language Questionnaires

Validity of PAL test scores as measures of children's linguistic abilities. How does the PAL test stack up against existing parent-completed language questionnaires? Unfortunately, making direct comparisons of the validity of scores on the PAL test and parent-completed language questionnaires is difficult because the ages of the children and/or aspects of language assessed by the PAL are different from those assessed by existing parent-completed questionnaires. With this caveat in mind, we can begin to compare the validity of the PAL test and parent-completed questionnaires by determining the extent to which PAL test scores correlate with standardized test scores versus the extent to which parent-completed questionnaire scores correlate with standardized test scores. In order for such comparisons to be meaningful, the PAL test score – standardized test score comparisons must be fairly similar to the questionnaire score – standardized test score comparisons.

The validity of PAL test scores compares favorably with that of CDI scores. Arguably, PAL Oral scores, CELF CLS scores, CDI Using Language scores and McCarthy Verbal scores all measure children's overall linguistic abilities. In a study of 100 three-year old children, Feldman et al. (2005) found that the CDI III Using Language - McCarthy Verbal correlation coefficient was .47. The fact that the PAL Oral 1- CELF CLS and PAL Oral 2 – CELF CLS correlation coefficients in the current study were approximately 1.5 times as great as Feldman et al.'s (2005) CDI Using Language-McCarthy Verbal correlation suggests that PAL Oral scores are better at measuring preschool children's overall linguistic abilities than CDI Using Language scores are at measuring toddlers' overall linguistic abilities. For the same group of children, Feldman et al. (2005) also found that the correlation coefficient for CDI III Expressive Vocabulary scores and Peabody Picture Vocabulary scores (Dunn, Dunn, & Dunn, 1997) was .41, whereas, in the current study, the correlation between PAL Receptive Vocabulary scores and CELF Expressive Vocabulary scores was .55. This suggests that the PAL Receptive Vocabulary scores may be somewhat better measures of children's vocabulary than CDI III Expressive Vocabulary scores.[5]

Let us now compare the PAL with the LDS. When compared to standardized vocabulary test scores, LDS vocabulary scores have excellent validity, with Rescorla (1989) reporting that children's LDS vocabulary scores were highly correlated (all r's $\geq$ .75) with their Bayley Mental Development Scale object and picture naming scores (Bayley, 1969), Reynell object and picture naming scores (Reynell, 1977), and their Preschool Language Scale picture scores (Zimmerman, Steiner, & Evatt, 1969), and Rescorla and Alley (2001) reporting that children's LDS scores were highly correlated with the number of Bayley objects and Stanford-Binet (Thorndike, Hagen, & Sattler, 1986) pictures they named (r = .69 and .74, respectively). PAL Receptive Vocabulary – CELF Expressive Vocabulary scores were also highly correlated (r = .55), but to a somewhat lesser degree than the LDS correlations. We cannot compare the validity of PAL Syntax and LDS sentence scores because no published studies have compared LDS sentence scores with standardized syntax test scores.

To date, no published studies have compared the concurrent validity of the CSBS checklist scores against standardized test scores. However, Wetherby et al. (2002) compared 33 children's composite speech scores on the CSBS at 23 months with their receptive and expressive language scores three months later on either the Mullen Scales of Early Learning (Mullen, 1995) or the Preschool Language Scales-3 (Zimmerman, Steiner, & Pond, 1992). The correlation between the CSBS checklist and receptive language scores was .38 and the correlation between CSBS checklist scores and expressive language scores was .50 (both p's < .01). For the CSBS 4-page questionnaire, the correlation was .47 for receptive language and .71 for expressive language (both p's < .01). Thus, the correlations between composite PAL Oral scores and CELF subtest and CELF CLS scores (*r's* ranging from .62 to .74, median r = .67) were higher than the CSBS checklist expressive and receptive language correlations and the CSBS

[5] One could argue that the CDI III and LDS fare worse than the PAL test because the children who took the CDI III and LDS were younger than those who took the PAL and older children's standardized test scores are more valid than younger children's test scores. This concern is mitigated somewhat by the fact the 3-year olds' PAL – Standardized test score correlations were as high as those for the children taken as a group (PAL Oral 1 – CELF CLS r = .73, p < .0001; PAL Oral 2 – CLS r = .70, p < .0001; PAL Receptive Vocabulary – CELF Expressive Vocabulary r = .54, p = .0002).

questionnaire receptive language correlation, and about the same as the CSBS questionnaire expressive language correlation. We cannot compare the validity of the PAL test and Bishop's CCC-2 because no studies have compared children's scores on the CCC-2 with their scores on standardized language tests.

The validity of the PAL test as a screening test for language impairments. As demonstrated by the area under the ROC curves (AUC), PAL Oral 1 and 2 scores were excellent at distinguishing between children who do and do not have language impairments (AUC = .96 for PAL Oral 1, AUC = .92 for PAL Oral 2). In a study with 38 children who were late talkers and 62 children who were not, Heilmann et al. (2005) found that low CDI Expressive Vocabulary scores were reasonably effective at distinguishing between typically-developing children and language-impaired children (defined as PLS Expressive Language scores and MLUs that were 1 SD below the mean). CDI Expressive Vocabulary scores that were $< 11^{th}$ percentile had a sensitivity of .68 and a specificity of .98; CDI scores that were $\leq 19^{th}$ percentile had a sensitivity of .81 and a specificity of .79, and CDI scores in the 49^{th} percentile had a sensitivity of 1.00 and a specificity of .44. The sensitivity and specificity of PAL Oral 1 and 2 scores compare favorably with Heilmann et al.'s (2005) CDI figures. PAL Oral 1 scores that were $< 11^{th}$ percentile had a sensitivity of .67 and a specificity of .96; PAL Oral 1 scores that were $\leq 19^{th}$ percentile had a sensitivity of .83 and a specificity of .88, and all of the language-impaired children had PAL Oral 1 scores below the 25^{th} percentile (sensitivity = 1.00, specificity = .87). PAL Oral 2 scores that were $< 11^{th}$ percentile had a sensitivity of .42 and a specificity of .93, PAL Oral 2 scores that were $\leq 19^{th}$ percentile had a sensitivity of .75 and a specificity of .87, and all of the language-impaired children had PAL Oral 2 scores at or below the 31^{st} percentile (sensitivity = 1.0, specificity = .75).

PAL Oral scores also compare favorable with CCC-2 General Communication Composite (GCC) scores with respect to correctly classifying children who do and do not have non-pragmatic language impairments. On the CCC-2, GCC scores that were 2 standard deviations (SD) below the mean had a sensitivity of .31, a specificity of .96, and, when the rate of language impairment is 10%, a PPV of .49 and NPV of .93, when compared against speech-pathologists' diagnoses of specific language impairment {Bishop, 2006 #4599}. When the rate of language impairment is set at 10%, PAL Oral 1 scores that were 2 SDs below the mean (bottom 5^{th} percentile) have a sensitivity of .42, a specificity of .99, a PPV of .83, and a NPV of .99, and PAL Oral 2 scores that were 2 SDs below the mean (also bottom 5^{th} percentile) had a sensitivity of .33, a specificity of .98, a PPV of .67 and a NPV of .93.

Turning to the LDS, using Rescorla's (1989) most liberal definition of language impairment (fewer than 50 words or no word combinations on the LDS), Klee et al. (1998) found that the LDS had a sensitivity of .91 and a specificity of .87 when 'true' language impairment was determined by clinical evaluation. Using the same LDS definition of impairment (which resulted in 10% of children being impaired), Rescorla and Alley (2001) found that the LDS had a sensitivity of .80, a specificity of .94, a PPV of .39 and a NPV of .99 when language impairment was defined as failing to name any objects on the Bayley Scales of Infant Development, and the LDS has a sensitivity of .64, a specificity of .94, a PPV of .39, and a NPV of .98 when impairment was defined as failing to name any picture on the Stanford-Binet Intelligence Scale. PAL Oral 1 scores were comparable to LDS scores in their power to discriminate between language-impaired and non-impaired children: when children with PAL Oral 1 scores in the bottom 10^{th} percentile were considered impaired, the sensitivity of PAL Oral 1 scores was .67 and the specificity was .96 (PPV = .67, NPV = .96) in identifying which children had CELF CLS scores in the bottom 10^{th} percentile. For the bottom 10^{th} percentile of PAL Oral 2 scores, the sensitivity was lower (.42), but the specificity (.93), PPV (.42) and NPV (.93) were comparable to the LDS. With respect to the CSBS checklist's validity, one study has investigated its validity as a screening test for a wide range of impairments (social, cognitive and linguistic), but no study has investigated its validity as a specific screening test for language impairment (Wetherby, Goldstein, Cleary, Allen, & Kublin, 2003).

In summary, the results of the experiment reported in this paper suggest that PAL test scores are valid measures of children's linguistic abilities, and that the PAL test is an excellent screening test for identifying children whose linguistic abilities warrant formal evaluation. Indeed, it is somewhat remarkable that a short (and cheap) test administered by parents compares so favorably with long (and expensive) tests administered by an experimenter. Three limitations of the current study exist. First, we

do not yet know the extent to which children's PAL test scores predict their future performance on standardized language tests or whether they will be diagnosed with a language disorder in the future. Second, while the current study is large compared to most of the early validation studies of the CDI, LDS, CCC-2 or CSBS, the data collected thus far are not sufficient to establish norms for different subgroups of children. The third concern is that the educational level of the parents who participated in the current study was somewhat higher than the national average, and while these parents had no difficulty administering the PAL test, poorly educated parents might. It should be noted that the educational level of parents who participated in many of the early studies of parent-administered language questionnaires was also higher than the US norm. In addition, if some parents were unable to administer the PAL test, paraprofessionals could, as is sometime done with the CDI and LDS. These limitations notwithstanding, the results of the current study are auspicious that the PAL test may fill an important gap in arsenal of assessment tools available to researchers and clinicians.

REFERENCES

Arriaga, R. I., Fenson, L., Cronan, T., & Pethick, S. (1998). Scores on the MacArthur Communicative Development Inventory of children from low and middle income levels. *Applied Psycholinguistics, 19*, 209-223.

Bayley, N. (1969). *Bayley Scales of Infant Development*. New York City: Psychological Corporation.

Benton, A. L., & Hamsher, K. (1977). *Multilingual aphasia examination*. Iowa City, Iowa: U. of Iowa

Bishop, D. V. M. (1998). Development of the Children's Communication Checklist (CCC): A method for assessing qualitative aspects of communication impairment in children. *Journal of Child Psychology and Psychiatry, 39*, 879-891.

Bishop, D. V. M. (2006). *The Children's Communication Checklist-2, United States Edition*. San Antonio, TX: Harcourt Assessment, Inc.

Bornstein, M. H., Hahn, C.-S., & Haynes, O. M. (2004). Specific and general language performance across early childhood: Stability and gender considerations. *First Language, 24*(3), 267-304.

Bricker, D., & Squires, J. (1999). *Ages & Stages Questionnaire: A parent-completed, child-monitoring system, second edition*: Paul H. Brookes Publishing Company.

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

Chien, Y. C., & Wexler, K. (1990). Children's knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics. *Language Acquisition, 1(3)*, 225-295.

Cohen, M. J., Morgan, A. M., Vaughn, M., Riccio, C. A., & Hall, J. (1999). Verbal fluency in children: Developmental issues and differential validity in distinguishing children with attention-deficit hyperactivity disorder and two subtypes of dyslexia. *Archives of Clinical Neuropsychology, 14*(5), 433-443.

Dale, P. S. (2001). *The MacArthur Communication Developmental Inventory: Level III*. Columbia, MO: University of Missouri-Columbia,.

Dale, P. S., Price, T. S., Bishop, D. V. M., & Plomin, R. (2003). Outcomes of early language delay: I. Predicting persistent and transient language difficulties at 3 and 4 years. *Journal of Speech, Language, & Hearing Research, 46*(3), 544-560.

Drumwright, A. (1971). *Denver Articulation Screening Examination*. Denver, CO: University of Colorado Medical Center.

Drumwright, A., Van Natta, P., Camp, B., Frankenburg, W., & Drexler, H. (1973). The Denver Articulation Screening Exam. *Journal of Speech and Hearing Disorders, 38*(1), 3-14.

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

Feldman, H. M., Dale, P. S., Campbell, T. F., Colborn, D. K., Kurs-Lasky, M., Rockette, H. E., et al. (2005). Concurrent and predictive validity of parent reports of child language at ages 2 and 3 years. *Child Development, 76*(4), 858-868.

Feldman, H. M., Dollaghan, C. A., Campbell, T. F., Kurs-Lasky, M., Janosky, J. E., & Paradise, J. L. (2000). Measurement properties of the MacArthur Communicative Development Inventories at ages one and two years. *Child Development, 71*(2), 310-322.

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., et al. (1993a). *The MacArthur-Bates Communicative Development Inventories (CDI I): Words and Gestures*. Baltimore, MD: Paul Brookes Publishing.

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., et al. (1993b). *The MacArthur-Bates Communicative Development Inventories (CDI II): Words and Sentences*. Baltimore, MD: Paul Brookes Publishing.

Gerken, L., & Shady, M. E. (1996). The picture selection task. In D. McDaniel, C. McKee & H. Cairns (Eds.), *Methods of Assessing Children's Syntax*. Cambridge, MA: MIT Press.

Heilmann, J., Weismer, S. E., Evans, J., & Hollar, C. (2005). Utility of the MacArthur—Bates Communicative Development Inventory in Identifying Language Abilities of Late-Talking and Typically Developing Toddlers. *American Journal of Speech-Language Pathology, 14*, 40-51.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development, 74*(5), 1368-1378.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression,* 2nd ed.). New York: John Wiley.

Klee, T., Carson, D. K., Gavin, W. J., Hall, L., Kent, A., & Reece, S. (1998). Concurrent and predictive validity of an early language screening program. *Journal of Speech, Language and Hearing Research, 41*(2), 627-641.

Koren, R., Kofman, O., & Berger, A. (2005). Analysis of word clustering in verbal fluency of school-aged children. *Archives of Clinical Neuropsychology, 20*(8), 1087-1104.

Landry, S. H., Smith, K. E., & Swank, P. R. (2002). Environmental effects on language development in normal and high-risk child populations. *Seminars in Pediatric Neurology, 9*, 192-200.

Largo, R. H., Pfister, D., Molinari, L., Kundu, S., & al., e. (1989). Significance of prenatal, perinatal and postnatal factors in the development of AGA preterm infants at five to seven years. *Developmental Medicine & Child Neurology, 31*(4), 440-456.

Leonard, L. B. (1998). *Children with Specific Language Impairment*. Cambridge, MA: MIT Press.

Leonard, L. B., Wong, A., Deevy, P., Stokes, S. F., & Fletcher, P. (2006). The production of passives by children with specific language impairment: Acquiring English or Cantonese. *Applied Psycholinguistic, 27*(2), 267-299.

Levin, B. E. (1990). Organizational deficits in dyslexia: Possible frontal lobe dysfunction. *Developmental Neuropsychology, 6*(2), 95-110.

Martin, J. A., Hamilton, B. E., Sutton, P. D., Ventura, S. J., Menacker, F., & Munson, M. L. (2005). *Births: Final Data for 2003*. Hyattsville, MD: National Center for Health Statistics.

Messer, D., & Dockrell, J. E. (2006). Children's Naming and Word-Finding Difficulties: Descriptions and Explanations. *Journal of Speech, Language, and Hearing Research, 49*, 309-324.

Mullen, E. M. (1995). *Mullen Scales of Early Learning*. Circle Pines, MN: American Guidance Service.

National Information Center for Children and Youth with Disabilities. (2000). Information about Speech and Language Disorders [Electronic Version]. *Fact Sheet*, 11. Retrieved November 26, 2006.

Nelson, H. D., Nyren, P., Walker, M., & Panosha, R. (2006). Screening for speech and language delay in preschool children: Systematic evidence review for the US Preventive Services Task Force. *Pediatrics, 117*(2), 298-319.

O'Grady, W. (1997). *Syntactic development*. Chicago: University of Chicago Press.

Rescorla, L. A. (1989). The Language Development Survey: A screening tool for delayed language in toddlers. *Journal of Speech and Hearing Disorders, 54*, 587-599.

Rescorla, L. A., & Achenbach, T. M. (2002). Use of the Language Development Survey (LDS) in a national probability sample of children 18 to 35 months old. *Journal of Speech, Language and Hearing Research, 45*(4), 733-743.

Rescorla, L. A., & Alley, A. (2001). Validation of the Language Development Survey (LDS): A parent report tool for identifying language delay in toddlers. *Journal of Speech & Hearing Research, 44*(2), 434-445.

Reynell, J. (1977). *Reynell Developmental Language Scales*. Windsor: NFER -Nelson Publishing.

Rice, M. L., & Wexler, K. (1996). Toward tense as a clinical marker of Specific Language Impairment in English-speaking children. *Journal of Speech and Hearing Research, 39*, 1239-1257.

Rice, M. L., Wexler, K., & Cleave, P. (1995). Specific language impairment as a period of Extended Optional Infinitive. *Journal of Speech and Hearing Research, 38*, 850-863.

Rice, M. L., Wexler, K., & Hershberger, S. (1997). Tense over time: The longitudinal course of tense acquisition in children with specific language impairment. *Journal of Speech, Language and Hearing Research, 41*, 1412-1431.

Riva, D., Nichelli, F., & Devoti, M. (2000). Developmental aspects of verbal fluency and confrontation naming in children. *Brain and Language, 71*(2), 267-284.

Roberts, J. E., Burchinal, M., & Durham, M. (1999). Parents' report of vocabulary and grammatical development of African American preschoolers: Child and environmental associations. *Child Development, 70*(1), 92-106.

Sanders, L. J. (1972). *Procedure guides for evaluation of speech and language disorders in children*. Danville, IL: Interstate Printers and Publishers.

Shriberg, L. D., Tomblin, J. B., & McSweeny, J. L. (1999). Prevalence of speech delay in 6-year-old children and comorbidity with language impairment. *Journal of Speech Language & Hearing Research, 42*(6), 1461-1481.

Stromswold, K. (1998). The genetics of spoken language disorders. *Human Biology, 70*, 297-324.

Stromswold, K. (2001). The heritability of language: A review and meta-analysis of twin, adoption and linkage studies. *Language, 77*(4), 647-723.

Thorndike, R. L., Hagen, E. P., & Sattler, J. (1986). *Stanford-Binet Intelligence Scale, 4th Edition*. Riverside, CA: Riverside Press.

Tomblin, J. B., Smith, E., & Zhang, X. (1997). Epidemiology of specific language impairment: Prenatal and perinatal risk factors. *Journal of Communication Disorders, 30*, 325-344.

US Preventive Services Task Force. (2006). Screening for Speech and Language Delay in Preschool Children: Recommendation Statement. *Pediatrics, 117*(2), 497-501.

van der Lely, H. K. J. (1996). Specifically Language Impaired and Normally Developing Children: Verbal Passive vs. Adjectival Passive Sentence Interpretation. *Lingua, 98*(4), 243-272.

van der Lely, H. K. J., & Dewart, H. (1986). Sentence Comprehension Strategies in Specifically Language Impaired Children. *The British Journal of Disorders of Communication, 21*(3), 291-306.

van der Lely, H. K. J., & Stollwerck, L. (1997). Binding theory and grammatical specific language impairment in children. *Cognition, 62*(1), 245-290.

Vihman, M. (1996). *Phonological development: The origins of language in the child*. Cambridge, MA: Blackwell Publishers.

Vohr, B. R., Garcia-Coll, C., & Oh, W. (1989). Language and neurodevelopmental outcome of low-birthweight infants at three years. *Developmental Medicine & Child Neurology, 31*(5), 582-590.

Weckerly, J., Wulfeck, B. B., & Reilly, J. (2001). Verbal fluency deficits in children with specific language impairment: Slow rapid naming or slow to name? *Child Neuropsychology, 7*(3), 142-152.

Wetherby, A. M., Allen, L., Cleary, J., & Kublin, K. (2002). Validity and Reliability of the Communication and Symbolic Behavior Scales Developmental Profile With Very Young Children. *Journal of Speech, Language and Hearing Research, 45*, 1202-1218.

Wetherby, A. M., Goldstein, H., Cleary, J., Allen, L., & Kublin, K. (2003). Early Identification of Children with Communication Disorders: Concurrent and Predictive Validity of the CSBS Developmental Profile. *Infants and Young Children, 16*(2), 161-174.

Wetherby, A. M., & Prizant, B. M. (2002). *Communication and Symbolic Behavior Scales Developmental Profile (CSBS DP)*: Paul H. Brookes Publishing Company.

Wiig, E. H., Secord, W. A., & Semel, E. (2004). *Clinical Evaluation of Language Fundamentals - Preschool, Second Edition*. San Antonio, TX: Harcourt Assessment, Inc.

Zimmerman, I. L, Steiner, V. G, & Evatt, R. (1969). *Preschool Language Scales*. Columbus, OH: Merrill.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (1992). *Preschool Language Scale - 3*. San Antonio, Texas: The Psychological Corporation, Harcourt Brace Jovanovich, Inc.

APPENDIX 1: PAL Test Items

PAL 3 (for 3-year old children)

Articulation test: *fat*, *soap*, *yuck*, *van*, *rat*, *lip*, *ship*, *cheek*, *zip*, *jeep*, *that*, *thin*
Receptive vocabulary test: *nurse, dentist, mittens, helicopter*, *canoe, kayak, trumpet*, and *saxophone*
Verbal Fluency Task 1: "Name an animal"
Verbal Fluency Task 2: "Name a thing" Name: a part of a face, a vegetable, a number, a drink, something round, a part of a car, a piece of clothing, something red, a toy, something big
Syntax test:

The lion combed the fox
The mouse bit him
The frog hid himself
The dog was licked by the bear
The bunny patted the duck
The sheep was kissed by the pig
The fox tickled the lion
The monkey splashed himself
The bear was slapped by the dog
The duck washed the bunny
The pig was scrubbed by the sheep
The cat scratched him

PAL 4 (for 4-year old children)

Articulation test: *rat*, *lip*, *ship*, *cheek*, *zip*, *jeep*, *that*, *thin*, *split*, *trick*, *clock*, *frog*
Receptive vocabulary test: same as PAL 3
Verbal fluency test: Same as PAL 3
Syntax test:

The dog licked the bear
The cat scratched himself
The fox was tickled by the lion
The pig scrubbed the sheep
The bear slapped the dog
The frog hid him
The bunny was patted by the duck
The mouse bit himself
The sheep kissed the pig
The duck was washed by the bunny
The monkey splashed him
The lion was combed by the fox

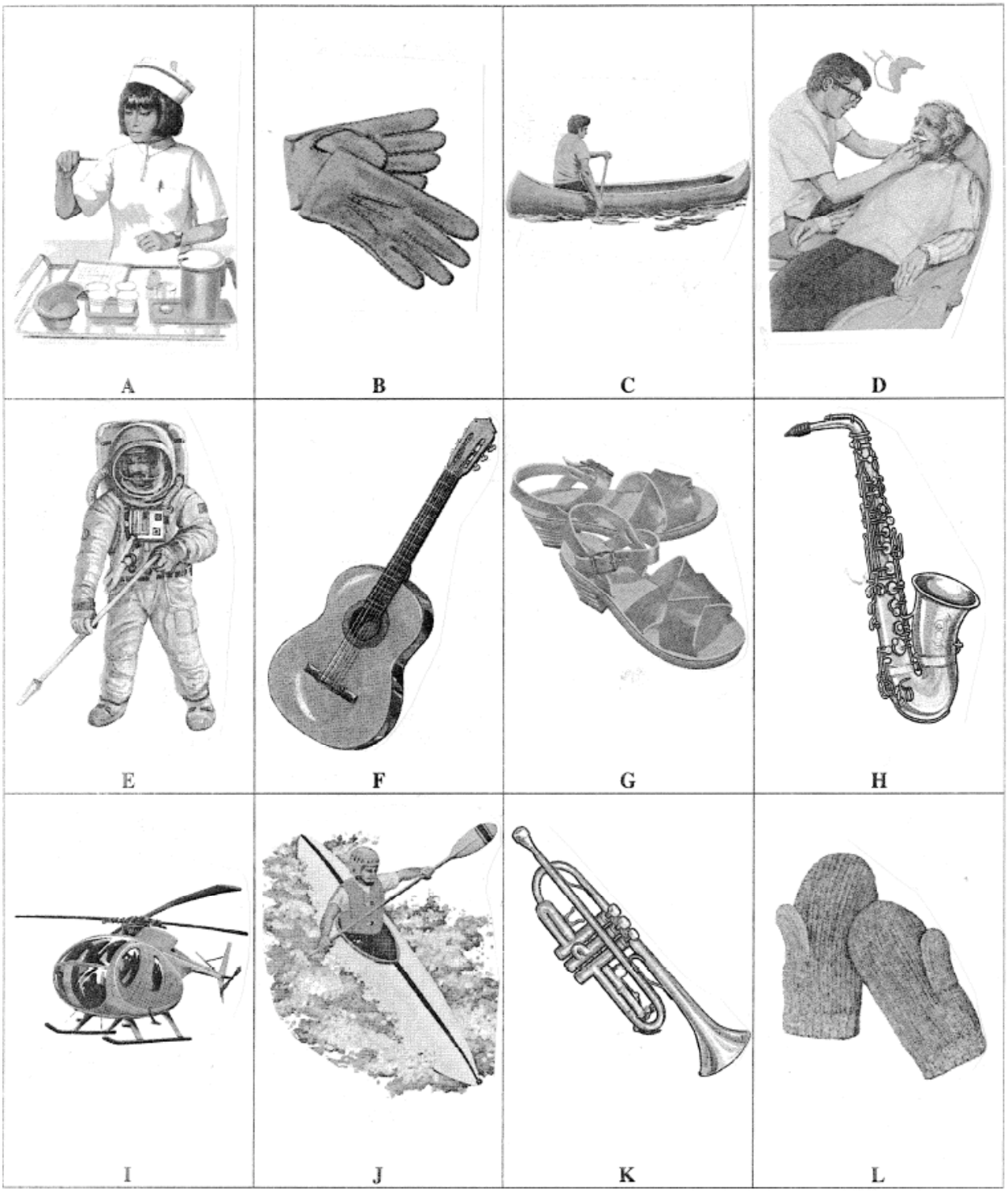
PAL 5 (for 5-year old children)

Articulation test: *split*, *trick*, *clock*, *frog*, *three*, *shrink*, *brake*, *flat*, *twin*, *street*, *scrub*, *squat*
Receptive vocabulary test: same as PAL 3
Verbal fluency test: Same as PAL 3
Syntax task:

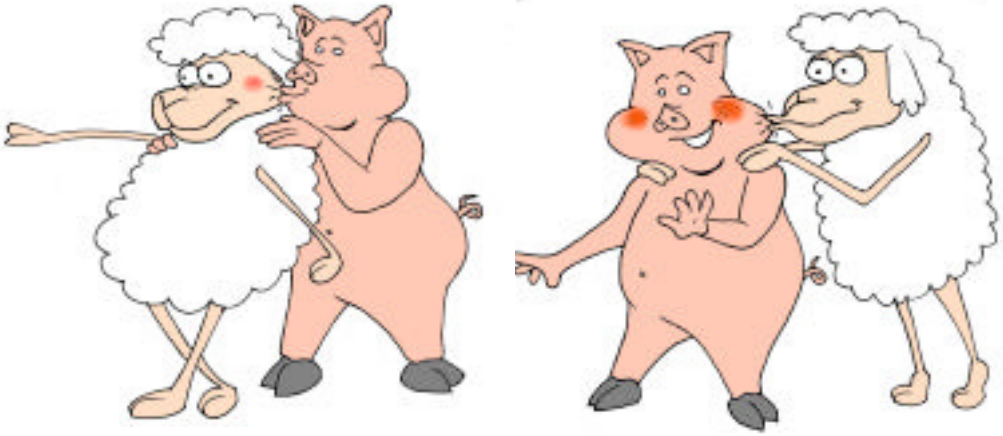
The mouse was scratching
The bear was licked
The fox was tickling himself
The bunny patted him
The monkey was splashed by the frog

The lion was combing himself
The dog was slapped by the bear
The duck was washing him
The pig was kissed by the sheep
The frog was hidden
The sheep was scrubbed
The cat was biting the mouse

APPENDIX 2: PAL Receptive Vocabulary Test Stimuli Sheet



APPENDIX 3: Picture Pair from the PAL Syntax Test



CAPTION: This pair of pictures was used with the sentences *The sheep was kissed by the pig* (PAL 3), *The sheep kissed the pig* (PAL 4), and *The pig was kissed by the sheep* (PAL 5).