The Validity of the Parent Administered Language (PAL) Test for Preschool-Aged Children *

Karin Stromswold

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

* Portions of the work reported in this paper were supported by grants from the National Science Foundation (BCS-9875168, BCS-0002010, BCS-0042561, BCS-0124095, and BCS-0446838), the National Institutes of Health (HD37818), the Busch Biomedical Research Fund, and the Bamford-Lahey Children's Foundation. I am indebted to Diane Molnar, Ellyn Sheffield, Medha Tare, a small army of undergraduate assistants, and the daycare centers, parents and children who participated in these studies.
Correspondence may be sent to Karin Stromswold (karin@ruccs.rutgers.edu)

The Validity of the Parent Administered Language (PAL) Test for Preschool-Aged Children

ABSTRACT

This paper investigates the validity of the Parent Administered Language (PAL) test, a 10-minute test of preschool children's articulation, lexical access, and syntactic abilities. In Experiment 1, parents of 528 preschool children administered the PAL tests and answered questions about their children's development. PAL scores were significantly correlated with most linguistic measures and children's initial PAL scores were significantly correlated with PAL scores one year later. Furthermore, PAL scores discriminated between children who were and were not language impaired. In Experiment 2, 20 preschool children took the PAL, Denver Articulation Screening Examination (DASE) and Clinical Evaluation of Language Fundamentals-Preschool (CELF-P) tests. PAL Articulation and DASE scores were highly correlated ($r = .66$) as were composite PAL scores and composite CELF scores ($r = 70$). These results suggest that PAL test scores are valid measures of preschool children's current linguistic abilities and good predictors of their future linguistic abilities.

THE VALIDITY OF THE PARENT-ADMINISTERED LANGUAGE (PAL) TEST FOR PRESCHOOL-AGED CHILDREN

INTRODUCTION

Over the past 50 years, thousands of researchers have devoted countless hours investigating how typically and atypically developing children learn to speak and understand English. For example, much research has centered on why children acquire some aspects of English before others and why some children learn English more quickly than others. Some studies have investigated interrelationships in the development of linguistic and nonlinguistic abilities. Other researchers have sought to determine whether the pattern of language development for special populations of children is different from that of typically developing children and, if so, how and why. Some studies have attempted to determine what biological and social factors raise the risk of language delays and impairments, and what the efficacy is for different types of clinical interventions. Genetic studies have investigated the extent to which the genetic factors that affect language are specific to language, and how interactions among genetic and environmental factors affect language development. Because definitively answering these questions often requires studying hundreds or thousands of children, the answers to many of these questions remain unclear. What is needed to better answer these questions is an inexpensive, easy-to administer language test that provides valid, quantitative measures of preschool children's linguistic abilities.

One of the most important skills preschool children must develop by the time they enter elementary school is the ability to speak so others can understand them, and to understand what other people say to them. Although by the time they are 5 or 6 years old, most children have mastered English, producing and understanding with ease phonologically and syntactically complex sentences that contain a large variety of lexical items, 5% to 10% of preschool children in the United States suffer from spoken language impairments or delays (see for example, Tomblin et al., 1997). Follow-up studies generally indicate that 40% to 60% of these children continue to have difficulties in later years (Hall & Tomblin, 1978; Aram & Nation, 1980; King, C., & Lasky, 1982). Not only is adequate spoken language important in and of itself, prospective and retrospective studies have shown that children with poor spoken language skills have more difficulty learning to read, write and spell than children with good spoken language skills (for a review, see Stromswold, 1997).

Some children with spoken language impairments have social or biological risk factors for language impairments (see, for example, Stanton-Chapman, Chapman, Bainbridge, & Scott, 2002), but many do not. Because many children with language impairments would not be considered "high risk," if we are to identify all (or even most) children with developmental language impairments, it is necessary to evaluate <u>all</u> preschool children's spoken language. Although many standardized tests of preschool children's language exist, these tests are designed to be administered by specially trained professionals such as speech-language pathologists and psychologists. Because it is not feasible to formally evaluate every preschool child's language skills, what is needed is a screening test that can be administered by people with no special training.

Parent-completed questionnaires of children's language have become popular in recent years because they allow language researchers and clinicians to quickly and inexpensively quantify the linguistic abilities of large numbers of children. The MacArthur-Bates Communicative Development Inventories (CDI) checklists are the most commonly used parent-completed language questionnaires, with over 200 published papers reporting using the CDI. 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 35 months of age (Dale, 2001). Three large studies have recently evaluated the validity of the CDI. In the first study, Feldman et al. (2000) compared the CDI-I and CDI-II scores of 2,156 children from diverse backgrounds. They found that children's scores on the CDI-I and CDI-II were significantly correlated with one another, indicating that CDI scores reflect children's language development. However, given that the correlations between CDI-I and the CDI-II scores ranged from .18 to .39, Feldman et al (2000) urged researchers and clinicians to "exercise caution" in using the CDI I to identify children with language impairments. In a subsequent paper, Feldman et al. (2005) compared the same children's CDI-II and CDI-III scores and found that children's CDI expressive vocabulary scores were fair to good predictors of their linguistic performance one year later, with negative predictive power being better than positive predictive power. From this, Feldman et al. (2005) concluded that CDI-II expressive vocabulary scores at age 2 are clinically useful to screen for language impairments at age 3, but that a substantial number of children with low CDI scores at age 2 will be clinically normal at age 3 and a substantial number of children with high CDI scores at age 2 will be clinically impaired at age 3. Feldman et al.'s (2005) results for American singletons are consistent with the results of a British study with over 8000 twins (Dale, Price, Bishop, & Plomin, 2003).

Over 20 published papers report using Rescorla's (1989) Language Development Survey (LDS). In the LDS, parents of children ages 18 to 35 months complete a 310-word expressive vocabulary checklist and give three examples of their child's longest multi-word utterances. Like the CDI, several studies have shown that the LDS is a reliable test and is useful as a screening test for language impairments in toddlers (see Rescorla, 1989; Klee et al., 1998; Rescorla & Alley, 2001; Rescorla & Achenbach, 2002; Rescorla, Ratner, Jusczyk, & Jusczyk, 2005 and references therein). Whereas the CDI was designed as a way of measuring the language development of typically developing children and only secondarily as a means of identifying children with language impairments, the LDS was designed primarily as a screening test for identifying children with language delays.

The Communication and Symbolic Behavior Scales - Developmental Profile (CSBS-DP, Wetherby & Prizant, 2002) is designed for children who are 6 to 24 month old. The CSBS-DP includes a one-page CSBS-DP parent-completed checklist and a four-page CSDB-DP parent-completed questionnaire that ask parents about their child's articulation, understanding and production of gestures, words and sentences. Although the CSBS-DP parent assessments were only recently developed, several studies suggest that they may have good concurrent and predictive validity (Wetherby, Allen, Cleary, & Kublin, 2002; Wetherby, Goldstein, Cleary, Allen, & Kublin, 2003).

A fourth parent-completed questionnaire is the Children's Communication Checklist-2 (CCC-2, Bishop, 2006). The CCC-2 is primarily used to identify children with language impairments, and to determine whether they suffer from semantic-pragmatic language impairments or other types of language impairments. As such, the CCC-2 asks parents about their child's phonological, morphosyntactic, semantic (e.g., does the child confuse words with similar meanings or sounds, use words that are too general or specific, etc.) and pragmatics-discourse skills (e.g., appropriate initiation of conversations, choice of words, understanding of non-literal language, discourse coherence, etc.). Over 20 studies have used the CCC-2, and several studies suggest that the CCC-2 can be used to discriminate between children who do and do not have a variety of language impairments (Bishop, 1998, Bishop, 2006 #4599).

Notice, that the CDI, LDS and CSBS-DP are all designed for children under 37 months and the CCC-2 is designed for children who are 4 and older. Put simply, no parent-completed language assessment tool exists for children who are between three and four years old. (Bricker and Squire's (1999) Ages and Stages Questionnaire can be used with three-year olds, but because it only asks 6 questions about children's communication ability, it provides a rather crude measure of children's linguistic skills.) Notice also that the CDI, LDS, CSBS-DP, and CCC-2 are all questionnaires. A general concern with estimates of linguistic abilities that are obtained from parent-completed questionnaires is that parents may have selective memories or biased judgments with respect to their child's linguistic accomplishments. Whereas one could argue that completing a vocabulary checklist requires no special skills, estimating a child's morphological, syntactic and phonological abilities may. Except when it comes to assessing discourse-pragmatic skills (which are notoriously difficult to test experimentally), professionally-administered language tests are generally considered to provide more accurate, reliable and precise measures of preschool children's language skills than measures derived from questionnaires and observation because the subjective component is less in language tests. However, one often-voiced concern is that young children's scores on standardized language tests reflect not only their linguistic abilities but also their level of comfort with unfamiliar people, settings and activities.

A parent-administered test of language could combine the best features of language tests and language questionnaires. Unlike with professionally administered language tests, children wouldn't have to interact with a stranger or be in an unfamiliar setting. In addition, parent-administered tests could provide a more objective measure of children's linguistic abilities than parent-completed questionnaires because tests do not require that parents remember what their child did in the past or estimate what their child can do now. A quick, valid test of preschool children's language that is simple enough for parents to administer could make it feasible to screen large numbers of preschool children and identify children whose language should be evaluated further by speech-language pathologists. In addition, such a test could provide researchers with a 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 clinical screening test would be comprehensive enough that it would not only identify children who may suffer from language delays, but identify what aspects of their language is 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 appears to exist for preschool children. A comprehensive review of the literature and of existing speech-language tests and queries to the CHILDES and ASHA electronic mailing lists failed to uncover a parent-administered test for assessing preschool children's language abilities. In addition, in a recent study of existing tests of preschool children's language abilities, no parent-administered language test was uncovered (Nelson, Nyren, Walker, & Panosha, 2006; US Preventive Services Task Force, 2006). As part of an ongoing study, we developed a parent-administered test of preschool children's language, the **P**arent **A**dministered **L**anguage (PAL) test. The PAL test is a simple-to-administer, 10-minute test that assesses preschool children's articulation, lexical, and syntactic skills.[1]

In order for the PAL test to be useful as a research tool, it must provide valid measures of preschool children's current linguistic abilities. In order for it to be a useful clinical screening test, it must be a reasonably sensitive and specific test for identifying children who are language impaired. To be useful as a clinical screening test for language impairments that may only become clinically apparent in older preschool children and young elementary school children, the PAL test must have good predictive validity such that children's early PAL scores predict their language skills at a later age. The results of the following two experiments indicate that the PAL test exhibits these properties.

EXPERIMENT 1: QUESTIONNAIRE-BASED STUDY OF THE PAL TEST

ANALYSIS 1: THE CONCURRENT VALIDITY OF THE PAL

Method

Participants

We used data from the first 528 children enrolled in the Perinatal Environment and Genetic

[1] The PAL test also has a letter-naming task. Because we have no other measures of pre-reading skills, we cannot assess the validity of the letter-naming task. Thus, we do not discuss these data in this paper.

Interactions (PEGI) study who were between 36 and 72 months old when they first took the PAL tests. Of these children, 50.8% were male, and 49.2% were female. The mean gestational age at birth (GA) of the children was 35.1 weeks (SD = 3.4 weeks), and children's GA-corrected ages were used to determine which tests were administered and when children acquired developmental milestones.

The children who participated in this study were recruited because they had perinatal risk factors for developmental delays. Consistent with this, most of the children in the study were born prematurely (GA < 37 weeks), had low birth weights (birth weights of less than 2500 grams) and were twins. According to the Center for Disease Control (Martin et al., 2005), 12% of US babies born in 2003 were premature and 8% had low birth weights. In this study, 5% of the children had GAs of 28 weeks or less, 8% had GAs between 29 and 31 weeks, and 45% had GAs between 32 and 36 weeks and 41% had GAs of 37 weeks or more. Consistent with the rate of prematurity, the mean birth weight (BW) of the children was 2345.9 grams (SD = 769.3 grams), with 6% of the children having BWs of less than 1000 grams, 7% having BWs between 1000 and 1500 gram, 42% having BWs between 1500 and 2500 grams, and 45% having BWs of 2500 grams or more. According to the CDC, 3% of US babies born in 2003 were twins, whereas 87% of the children in our study were twins.

Although we encouraged parents from all racial, ethnic, educational and financial backgrounds to participate in the study, the parents who did so were more likely to be white, non-Hispanic, well educated and wealthy than the general US population. According to the US Census Bureau in 2000, 69% of the US population were non-Hispanic white, 13% were Hispanic, 12% were African American, 4% were Asian, and 1% were Native American, whereas in our study 96% were non-Hispanic white, 2.5% were Hispanic, 1% were African American, and 0.5% were Native American. According to the US Census Bureau, in 2004, 28% of US adults over the age of 25 had a bachelor's degree or more. Of the mothers who participated in this study, 3% had a high school diploma or less, 20% had some post-secondary vocational training or education, and 78% had graduated from college. Of the fathers, 9% had a high school diploma or less, 25% had some post-secondary training or education, and 65% had graduated from college. According to the US Census Bureau, in 2003, the median family income for 4-person families was $65,000, whereas in our study, 9% of the families had family incomes of less than $35,000, 14% had family incomes between $35,000 and $50,000, 27% had families incomes between $50,000 and $75,000, and 50% had family incomes of $75,000 or greater.

Design and development of the PAL test

The PAL tests assess children's abilities in 3 critical aspects of spoken language development that are frequently assessed in standardized language tests: articulation, expressive vocabulary/lexical access and syntax. In developing the PAL, we were careful to only include tasks that 1) are simple to administer; 2) are quick to administer; 3) require no special equipment to administer; and 4) elicit clear (unambiguous) responses that are easy to observe and record. The Appendix gives the items tested on each PAL.

PAL Articulation Test. The PAL test uses a word repetition task to assess children's articulation. Because it is easiest to detect mispronunciations when they occur at the beginning of a word (the onset), the PAL tests whether children pronounce onsets correctly. In this test, a parent asks his or her child to repeat 12 monosyllabic words, and checks 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 addition to reporting how the child says the onsets of 12 words, the parent lists any sounds that the child regularly mispronounces and gives examples of typical mispronounced words. In order to prevent either ceiling or floor effects, the wordlists on each version of the PAL 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*. For the PAL Articulation test and all other PAL tests, two people independently hand-coded the data. The inter-coder reliability was 98% or better for all PAL tests. Table 1 gives mean scores for all tests.

PAL Lexical Access Test. The PAL Lexical Access test uses a semantic verbal fluency test (name as many animals as you can in 30 seconds) and a "name quickly" test (in 30 seconds, name a part of a face, a fruit) to assess children's expressive vocabulary/lexical access. We used verbal fluency tasks because they do not require pictures and because typically-developing children's performance on verbal fluency tasks increases with age (Riva, Nichelli, & Devoti, 2000; Koren, Kofman, & Berger, 2005). In addition, performance on verbal fluency tasks has been shown to be a sensitive measure of the lexical access abilities of typically developing children (Riva, Nichelli, & Devoti, 2000; Koren, Kofman, & Berger, 2005), specific language impaired (SLI) children (Weckerly, Wulfeck, & Reilly, 2001), and dyslexic children (Levin, 1990; Cohen, Morgan, Vaughn, Riccio, & Hall, 1999), with SLI and dyslexic children performing more poorly than typically-developing children. Table 1 gives the mean number of correct responses when repetitions and incorrect responses were excluded.

TABLE 1: Mean Scores on PAL Tests

	PAL 3 (N= 312)	**PAL 4** (N = 148)	**PAL 5** (N = 68)	**Overall** (N = 528)
PAL Articulation	9.05 (SD = 2.64)	9.66 (SD = 2.64)	10.25 (SD = 3.12)	9.38 (SD = 2.68)
PAL Lexical Access	4.49 (SD = 2.68)	5.70 (SD = 2.57)	7.54 (SD = 2.39)	5.21 SD = (2.81)
PAL Syntax	8.19 (SD = 2.27)	9.52 (SD = 2.25)	10.03 (SD = 2.14)	8.81 (SD = 2.36)
PAL Oral	21.75 (SD = 5.55)	24.89 (SD = 4.86)	27.84 (SD = 5.86)	23.42 (SD = 5.83)

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. This sentence-picture matching comprehension task was chosen because it is widely used both in research and clinical settings, yields relatively clear and unambiguous responses that are easy to record, and is arguably the easiest syntactic test to administer to children (for a review, see Gerken & Shady, 1996). One limitation of the sentence-picture matching task is that one can only test syntactic constructions whose propositional content is easy to depict clearly and unambiguously. [2]

The PAL Syntax test includes active sentences with 2 lexical NPs (e.g., *the pig kissed the sheep*), active sentences with reflexive pronouns (e.g., *the pig kissed himself*), active sentences with non-reflexive pronouns (e.g., *the pig kissed him*), active sentences that lacked an overt object (e.g., *the pig was kissing*), *by* passive sentences (e.g., *the pig was kissed by the sheep*), and truncated passive sentences (e.g., *the pig was kissed*). We included active sentences with reflexive and non-reflexive pronouns because some studies have suggested that preschool children who are linguistically normal (e.g., Chien & Wexler, 1990) and older children with SLI (van der Lely & Stollwerck, 1997) sometimes interpret sentences with non-

[2] 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 which 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).

reflexive pronouns as if they had reflexive pronouns. We included 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; Precious & Conti-Ramsden, 1988; van der Lely, 1996; Leonard, Wong, Deevy, Stokes, & Fletcher, 2006).

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 measure children's syntactic abilities (and not other linguistic or nonlinguistic abilities), we were careful to eliminate non-syntactic cues and confounds when we designed the test. For example, all of the sentences in the PAL Syntax test 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 of the sentences have verbs that are felicitous in active sentences and in verbal passive sentences that contain an animate patient and an overt animate agent. All sentences contain definite noun phrases that refer to animals that are paired such that either animal is equally plausible as the agent of the sentence (e.g., *pig* and *sheep*), and each animal in a pair is the agent and the patient equally often over the course of the test. The visual stimuli are designed so that pairs of pictures contain no visual cues as to which picture in the pairs matches a sentence (see Figure 1). 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. Over the course of the test, the animal that is the agent of the sentence appears on the left and the right of the patient equally often and the correct picture is the left and right picture equally often. In order to minimize the number of items on the PAL Syntax test and simultaneously prevent either ceiling or floor effects at any age, the composition of the sentences differs at different ages (see Appendix).

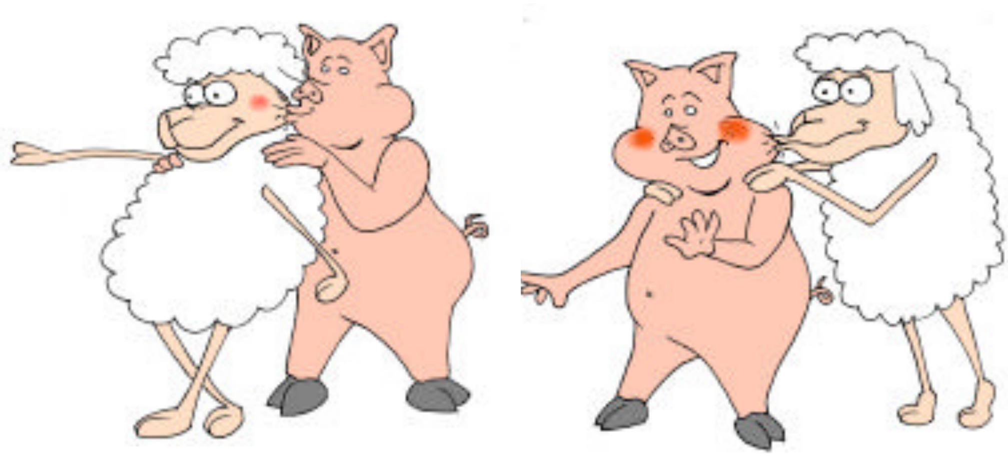

FIGURE 1: Example of Picture Pairs Used in the PAL Syntax Test
This pair of pictures was used with the sentences *The sheep was kissed by the pig* (PAL 3), *The sheep kissed the pig* (PAL 4), and *The pig was kissed by the sheep* (PAL 5).

PAL Oral score. We summed the children's scores on the three PAL tests to obtain an overall score for the PAL tests (the PAL Oral score).

Other Language Measures

We obtained additional measures of children's language development. Parents of all children completed the age-appropriate version of the communication section of the Ages and Stages Questionnaire (ASQ, Bricker & Squires, 1999). In addition, parents reported when their child reached 4 language milestones. We used a 2-stage process in choosing developmental milestones. First, we searched the literature and selected 25 linguistic and nonlinguistic milestones using the following criteria 1) good normative data existed for the milestone, 2) parents were likely to notice and remember when their child attained the milestones, 3) as a group, the milestones spanned different areas of development, and 4) as a group, milestones spanned the first 4 years of life. We had 30 parents report when their child achieved each of the 25 milestones and how confident they were in their answers. Based on parents' answers, we selected 4 language milestones: onset of babbling, onset of first words, onset of multi-word utterances, and onset of clear articulation (speaking so that a stranger can understand most of the words the child says) and 11 nonlinguistic milestones (see below). Given that most of the children who took the PAL were at high risk for language delay because they were twins, low birth weight and/or premature (see Stromswold, 2006), the GA-corrected age at which the children attained each of the 4 language milestones is consistent with the age reported in the literature (for our participants, onset of babbling mean = 8.36 months, SD = 4.74 months; onset of words = 13.61 months, SD = 5.65 months; onset of sentences = 19.49 months, SD = 6.19 months; onset of clear articulation mean = 24.44 months, SD = 8.39 months). This suggests that the parents in the study accurately reported when their children attained these linguistic milestones. Parents also rated -- on a 5-point scale – how their child's linguistic abilities compared to those of children the same age (comparative language measure). Although the comparative language measure is a crude and potentially biased measure of children's linguistic skills, the fact that the mean rating parents gave was in the middle of the scale (3.19) and parents used the full range of scores (SD = 0.99) suggests it may be informative.

Lastly, we asked parents whether their child received speech-language therapy (SLT) during each year of life. Although, like the comparative language measure, the SLT measure does not provide information about children's skills for different aspects of spoken language, the SLT measure has three desirable properties. First, speech language pathologists' assessments of children's linguistic abilities are less likely to be biased than parents' assessments. Second, the amount of SLT children receive indirectly taps speech-language pathologists' informal and formal evaluations of children's language. Third, parents' reports of how much SLT their child received may be more accurate than their reports of when their child acquired linguistic milestones or how their child's language compares to that of other children. Of the 528 children in the study, 74% received no SLT, 10% received one year of SLT, 8% received two years of SLT, and 8% received three or more years of SLT. Because the children ranged in age from 36 to 72 months, we calculated the number of years of therapy a child received divided by the child's age. The mean amount of SLT/age was 0.12 (SD = 0.24, range 0-1).

RESULTS

Correlation Results

We assessed the concurrent validity of the PAL by determined the extent to which PAL test scores were correlated with the other language measures. To compensate for the large number of correlations performed, for all analyses in Experiment 1, we set the p level for significance at $p \leq .001$. To avoid the possibility of a practice effect, in the concurrent analyses, we only included data from the first time the parents administered the PAL. We did, however, combine the data of children who took the PAL 3, 4, and 5 because doing so tripled our sample size and, thus, substantially increased the statistical power of the analyses.[3] We combined boys' and girls' data because the correlation coefficients obtained were virtually identical for the two groups and doing so increased the statistical power of the study.

If the PAL test has concurrent validity as a measure of children's language development, children who have high PAL scores should attain language milestones earlier and receive less SLT than children with low PAL scores (i.e., PAL-language milestone and PAL-SLT correlations should be negative). High PAL scores should be associated with high scores on all other language (i.e., all other language measure-PAL correlations should be positive).

PAL Articulation. Inspection of Table 2 reveals that PAL Articulation scores were significantly correlated in the predicted direction with all 10 of the other language measures. Excluding the correlation between PAL Articulation and PAL Oral scores, the absolute values of the correlation coefficients ranged from .21 (onset of words) to .42 (comparative language) with a mean correlation of .29 (median = .26). Of the language milestones, PAL Articulation scores were most highly correlated with the age of onset of clear articulation. Of the non-milestone language measures, PAL Articulation scores were most highly correlated with the more general measures of language development (comparative language and ASQ communication scores).

TABLE 2: Correlations between PAL Scores and Other Language Measures

	PAL Articulation		**PAL Lexical Access**		**PAL Syntax**		**PAL Oral†**	
	r	N	*r*	N	*r*	N	*r*	N
Babbling Milestone††	-.25**	396	-.08	396	-.10	395	-.19**	396
Word Milestone	-.21**	429	-.11	429	-.19**	428	-.23**	429
Sentence Milestone	-.23**	408	-.16*	408	-.26**	407	-.28**	408
Articulation Milestone	-.30**	413	.04	413	-.13	412	-.21**	413
Years of SLT/Age	-.25**	528	-.14*	528	-.28**	528	-.33**	528
Comparative Language	.42**	528	.21**	528	.38**	528	.45**	528
AS Communication	.38**	528	.30**	528	.45**	528528	.50**	528
PAL Articulation	-	-	.26**	528	.34**	528	.72**	-
PAL Lexical Access	.26**	528	-	-	.40**	528	.74**	528
PAL Syntax	.34**	528	.40**	528	-	-	.78**	528
PAL Oral†	.72**	528	.74**	528	.78**	528	-	-
Mean	.29		.19		.28		.31	
Median	.26		.16		.28		.28	

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

[3] When PAL score correlations were calculated separately for 3-year olds (N = 312), 4-year olds (N = 148) and 5-year olds (N = 68), with a handful of exceptions, the correlations that were significant for the children taken as a group were also significant for each age group, and vice versa. The exceptions were that the correlations between PAL Articulation and PAL Lexical Access scores ($r = .12$), PAL Articulation scores and onset of words ($r = -.07$), and PAL Lexical Access and Comparative Language scores ($r = .13$) were not significant for the 4-year olds but were significant for the children taken as a group. Three PAL-language milestones that were significant for the children taken as a group were not significant for the 5-year olds (PAL Articulation-First Words $r = -.18$, PAL Articulation-Sentences $r = -.22$, PAL Oral Score-Babbling $r = -.09$), and the PAL Syntax-Articulation correlation was significant for the 5-year olds ($r = -.55$, $p < .0001$), but not the children as a group.

† PAL subtest scores were not included in mean and median calculations for PAL Oral correlations, and vice versa

†† The number of subjects varied for the milestone correlations because some parents either didn't report when their child acquired a milestone or said they were not confident about their answer.

PAL Lexical Access. PAL Lexical Access scores were significantly correlated with scores on 7 language measures (see Table 2). The absolute values of the correlation coefficients ranged from .04 (onset of clear articulation) to .40 (PAL Syntax score), with a mean correlation of .19 (median = .17). Of the language milestones, PAL Lexical Access scores were most highly correlated with the age at which children began using sentences. Of the non-milestone language measures, PAL Lexical Access scores were most highly correlated with PAL Syntax scores. These two findings raise the intriguing possibility that semantic verbal fluency tasks tap many of the same linguistic skills as sentence comprehension tasks.

PAL Syntax. As shown in Table 2, PAL Syntax scores were significantly correlated with 8 of the other language measures. The absolute values of the correlation coefficients ranged from .10 (onset of babbling) to .45 (ASQ communication score), with a mean and median correlation of .28. Of the language milestones, PAL Syntax scores were most highly correlated with the onset of sentences. Of the other language measures, PAL Syntax scores were most highly correlated with ASQ communication scores and PAL Lexical Access scores.

PAL Oral scores. PAL Oral scores were significantly correlated with all 7 of the non-PAL language measures, with the absolute values of the correlation coefficients ranging from 0.19 (onset of babbling) to 0.50 (ASQ communication scores) with a mean correlation of .31 (median = .28, see Table 2). Given that PAL Oral scores reflect overall linguistic ability, not surprisingly, PAL Oral scores were most highly correlated with the two general measures of language development (comparative language and ASQ communication scores).

PAL Oral Scores as a Screening Test for Language Impairments.

To assess whether a screening test for developmental language disorders 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 positives will be artificially elevated if too few children are labeled language impaired. In order to test the validity of a language-impairment screening test, one needs to know which children are, in fact, 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.' The apparent validity of the language screening test will be depressed to the extent that this gold standard misclassifies language-impaired children as unimpaired or unimpaired children as language impaired. With respect to language impairments, the gold standard is a valid, reliable language test administered by a speech-language pathologist.

Unfortunately, we did not have such a measure. Therefore, we used the best measure we had available – the amount of SLT children received per year of life – to test the diagnostic validity of the PAL. While this measure is clearly less than ideal, it has two desirable properties. First, the fact that a child has received any SLT reflects that a speech-language pathologist has deemed the child to be clinically impaired. Second, the amount of SLT the child has received reflects *how* impaired the speech pathologist thinks the child is. For the purposes of these analyses, children were considered language impaired if they were in the top 15% of children with respect to the amount of SLT received. We chose this value for 3 reasons. First, although we did not specifically recruit language impaired children, as a group, the children in this study were at high risk for language impairments (see Stromswold, 2006), because they were born prematurely, had low birth weights and/or were twins. Second, a quarter of the children had received some SLT, so a lower threshold seemed imprudent. Third, a 15% impairment rate was midway between the 10th and 20th percentile values Feldman et al. (2005) explored as CDI vocabulary level-defined impairment. Fourth, the bottom 15th percentile corresponded to 1 SD below the mean amount of SLT received by children who took the PAL – the threshold Feldman et al. (2005) adopted in their analyses of the predictive validity of the CDI.

We converted children's continuous scores on our measure of language impairment (amount of SLT received) into the binary values of language impaired and unimpaired. 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) when different PAL Oral scores were used as the cutoff for failing. Low PAL Oral scores were a very specific indicator that a child was language impaired. When children whose PAL Oral scores were in the bottom 2.5th percentile were considered language impaired, the sensitivity of PAL oral scores was 11.2% and the specificity was 99.1%. When children whose PAL Oral scores were in the bottom 5th percentile were considered language impaired, the sensitivity of PAL Oral scores was 20.2% and the specificity was 97.6%. When children whose PAL Oral scores were in the bottom 10th percentile were considered language impaired, the sensitivity of PAL Oral scores was 30.4% and the specificity was 93.5%. When children whose PAL Oral scores were in the bottom 15th percentile were considered language impaired, the sensitivity of PAL oral scores was 35.4% and the specificity was 88.6%. Half of all language-impaired children having PAL Oral scores in the bottom 25th percentile (sensitivity = 49.4%, specificity = 79.3%) and almost all language-impaired children had PAL scores below the 50th percentile (sensitivity = 94.5%, specificity = 57.9%).

A Receiver Operating Characteristic (ROC) curve was generated by plotting the true positive rate (i.e. the sensitivity) against the false positive rate (i.e., 1 – specificity) for different PAL Oral score cutoffs (see Figure 2). The area under the ROC curve (AUC) provides a measure of the overall goodness of a test (i.e., the extent to which PAL scores correctly classify children as impaired and unimpaired). The AUC for PAL Oral scores was .81, indicating that PAL Oral scores do a good job discriminating between children who are and are not language impaired.

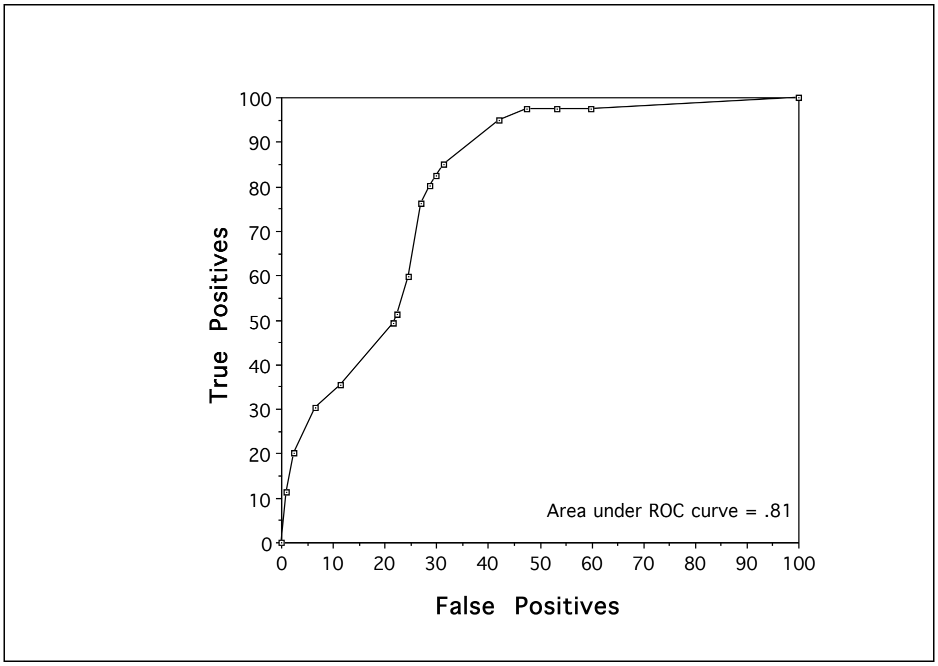

FIGURE 2: Receiver Operating Characteristic Curve of Low PAL Scores as a Screening Test for Language Impairments

ANALYSIS 2: LINGUISTIC SPECIFICITY OF THE PAL TEST

From a clinical standpoint, it may be sufficient (and even desirable) to identify children with language delays and impairments regardless of whether the delays are restricted to language (i.e., specific or selective language delays) or are a reflection of more general developmental delays. Similarly, for certain types of research, it may be sufficient to have a measure of children's language development that conflates linguistic and non-linguistic abilities. However, some studies require measures of language development that reflect children's linguistic abilities independent of their non-linguistic abilities.

TABLE 3: Correlations between PAL Scores and Non-Linguistic Outcome Measures

	PAL Articulation		**PAL Lexical Access**		**PAL Syntax**		**PAL Oral**	
	r	N	*r*	N	*r*	N	*r*	N
Gross Motor (GM)								
Sitting Milestone††	-.08	475	-.09	475	-.15	475	-.14	474
Crawling Milestone	-.07	457	-.13	457	-.12	457	-.14	456
Walking Milestone	-.13	510	-.16*	510	-.18**	510	-.21**	509
Running Milestone	-.11	377	-.13	377	-.13	377	-.16	376
Climbing Milestone	-.17	348	-.13	348	-.24**	348	-.24**	347
PT	-.06	526	-.10	528	-.20**	526	-.16*	525
Comparative GM	.03	430	-.03	430	.02	430	.001	429
AS Gross Motor	.16*	528	.06	528	.17**	528	.18**	527
Fine Motor (FM)								
Finger Feeding Milestone	-.11	422	-.04	422	-.08	422	-.11	421
Fork Feeding Milestone	-.08	385	-.04	385	-.05	385	-.08	384

Scribbling Milestone	-.04	350	-.12	350	-.06	350	-.10	349
Cutting Milestone	.00	344	-.02	344	-.073	344	-.04	344
OT	-.07	526	-.10	526	-.23	526	-.18**	525
Comparative FM	.15	429	.11	429	.20	429	.21**	428
AS Fine Motor	.20**	528	.17**	528	.20	528	.26**	527
Social								
Smiling Milestone	.04	443	-.01	443	.01	443	.02	442
Psychological Therapy	-.04	526	-.02	526	-.12	526	-.08	525
Comparative Social	.23**	430	.24**	430	.26	430	.33**	429
AS Social	.16*	527	.24**	527	.36	527	.34**	526
Cognitive								
Special Education	-.18**	526	-.10	526	-.23	526	-.23**	525
Comparative Cognition	.24**	427	.18*	427	.30	427	.32**	426
AS Cognitive	.25**	527	.25**	527	.38	527	.40**	526
Oral Motor								
Cup Drinking Milestone	-.16	381	-.03	381	-.06	381	-.12	380
Feeding Therapy	-.04	526	-.10	526	-.16	526	-.13	525

* $p < .001$
** $p < .0001$
†† The number of subjects varied for the milestone correlations because some parents either didn't report when their child acquired a milestone or said they were not confident about their answer.

We assessed the linguistic specificity of the PAL by comparing the extent to which children's PAL scores correlated with other measures of their linguistic and nonlinguistic development. The 8 gross motor measures used were ASQ gross motor scores, 5 gross motor milestones (sitting unsupported, crawling, walking, running, and climbing stairs), amount of physical therapy (PT) received per year of life, and comparative gross motor development. The 7 fine motor measures were ASQ fine motor scores, 4 fine motor milestones (finger feeding, spoon feeding, scribbling, and cutting with scissors), amount of occupational therapy (OT) received per year of life, and comparative fine motor development. The 4 social measures were ASQ social-personal scores (henceforth, ASQ social scores), onset of social smiling, amount of behavioral/psychological therapy received per year of life, and comparative social development. The 3 cognitive measures are ASQ problem-solving scores (henceforth, ASQ cognitive scores), amount of special education services received per year of life, and comparative cognitive development. Lastly, the 2 oral motor measures were onset of drinking from an open cup and amount of feeding therapy received per year of life.

As shown in Table 3, PAL Articulation scores were significantly correlated with less than 30% of the non-language measures (mean r = .12, median r = .11) and all of the language measures (mean r = .29, median r = .26). PAL Lexical Access scores were significantly correlated with only a quarter of the non-language measures (mean r = .11, median r = .10) versus 70% of the language measures (mean r = .19, median r = .16). PAL Syntax scores were significantly correlated with half of the non-linguistic measures (mean r = .17, median r = .16) and 80% of the language measures (mean and median r = .28). Lastly, PAL Oral scores were significantly correlated with half of the non-linguistic measures (mean r = .17, median r = .16) versus all of the language measures (mean r = .31, median r = .28).

ANALYSIS 3: PREDICTIVE VALIDITY OF PAL SCORES

The next question we addressed is the extent to which children's PAL scores predicted their PAL scores one year later (i.e., the predictive validity of the PAL). We collected follow-up data from some of the children who originally participated in the study. The follow-up return rate is approximately 85%, and as of March 1, 2006, we had scored, checked, and entered one-year follow up data from 184 children who took a PAL at Time 1 (T1). The mean GA-corrected age of these children at follow-up assessment was 50.4 months (SD = 9.6 months). With the exception of the age of the children at follow-up assessment, the demographic characteristics of the children who participated in the second phase of the study did not differ from those of the children who initially participated in the study (all *p*'s > .10), indicating that there wasn't selective drop out among the participants in the study.

Correlation Results

As shown in Table 4, PAL Articulation scores at T1 were significantly correlated with PAL Articulation, PAL Lexical Access, and PAL Oral scores at T2. Not surprisingly, of these correlations, the highest was between T1 and T2 PAL Articulation scores (r = .57), with the T1 PAL Articulation-T2 PAL Oral correlation being second highest (r = .49). PAL Lexical Access scores at T1 were significantly correlated with PAL Lexical Access scores at T2 (r = .37) and PAL Oral scores at T2 (r = .33). PAL Syntax scores at T1 were significantly correlated with all 6 of the T2 language measures. In contrast with the finding that for PAL Articulation and Lexical Access tests the highest T1-T2 correlations were for the same PAL test, the T1-T2 PAL Syntax correlation was not the highest. Rather, T1 PAL Syntax scores were most highly correlated with T2 PAL Oral scores (r = .37) and T2 ASQ communication scores (r = .36). Lastly, PAL Oral scores at T1 were significantly correlated with all 5 language measures collected at T2, with the T1-T2 oral score correlation being the highest (r = .52).

TABLE 4: Predictive Validity of PAL Test Scores

	T1 PAL Articulation	T1 PAL Lexical Access	T1 PAL Syntax	T1 PAL Oral
T2 ASQ Communication†	.18	.17	.36**	.31**
T2 PAL Articulation	.57**	.10	.25*	.39**
T2 PAL Lexical Access	.25*	.37**	.27*	.40**
T2 PAL Syntax	.19	.22	.25*	.29**
T2 PAL Oral	.49**	.33**	.37**	.52**

† The number of children for the PAL T1 – ASQ T2 correlations was 146 because the ASQ is not designed for 6-year olds. The number of children for PAL T1 – PAL T2 was 184.
* $p \leq .001$
** $p < .0001$

The validity of PAL scores as a screening test for future language impairments

To investigate whether children's performance on the PAL test at T1 accurately identified children with poor linguistic skills one year later, we followed the procedures Feldman et al. (2005) used to investigate the predictive validity of the CDI-II. For the purposes of these analyses, children whose PAL Oral scores were in the bottom 15th percentile were deemed to be language-impaired. Figure 3a shows the ROC curve for PAL Oral scores at T1 and T2. Consistent with the high degree of correlation between initial and follow-up PAL Oral scores, children's initial PAL Oral scores were good predictors of their PAL Oral scores 15 months later (AUC = .83), with the tradeoff between sensitivity and specificity being particularly good for low PAL Oral scores (e.g., lowest 5th percentile sensitivity = 28.6% and specificity = 99.4%; lowest 10th percentile sensitivity = 42.9% and specificity = 94.9%; lowest 15th percentile sensitivity = 53.6% and specificity = 94.2%; lowest 25th percentile sensitivity = 71.4% and specificity = 82.7%).

Let us now consider the predictive validity of individual PAL tests. These analyses revealed that children's PAL Articulation scores at T1 were good predictors of their PAL Articulation scores at T2 throughout a large range of articulation scores (PAL T1-T2 Articulation AUC = .83, see Figure 3b; lowest 5th percentile sensitivity = 21.4% and specificity = 98.8%; lowest 10th percentile sensitivity = 46.2% and specificity = 94.2%; lowest 15th percentile sensitivity = 61.5% and specificity = 90.8%; lowest 20th percentile sensitivity = 64.3% and specificity = 85.9%; lowest 50th percentile sensitivity = 85.7% and specificity = 58.2%). Overall, children's T1 Lexical Access scores were fair predictors of their T2 Lexical Access scores (AUC = .74, see Figure 3c). Although low Lexical Access scores at T1 were reasonable predictors of low poor Lexical Access scores at T1 (e.g., lowest 5% sensitivity = 21.4% and specificity = 98.1%; lowest 10% sensitivity = 28.6% and specificity = 94.2%; lowest 15% sensitivity = 39.3% and specificity = 91.7%), the tradeoff between sensitivity and specificity dropped off when 25% or more of children were considered impaired (e.g., lowest 25% sensitivity = 46.4%, specificity = 84.5%. Children's T1 Syntax scores were fairly poor predictors of their T2 Syntax scores (AUC = .64, see Figure 3d) regardless of what percentage of the children were treated as impaired based on T1 Syntax scores (e.g., lowest 5th percentile sensitivity = 10.7% and specificity = 96.8%; lowest 15th percentile sensitivity = 32.1% and specificity = 86.6%; lowest 33rd percentile sensitivity = 35.7% and specificity = 62.8%).

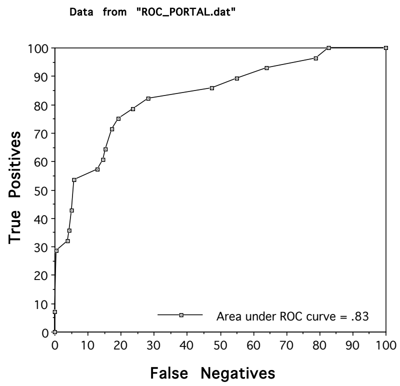

Figure 3a: Year 1 PAL Oral Score as a Predictor of Year 2 PAL Oral Score

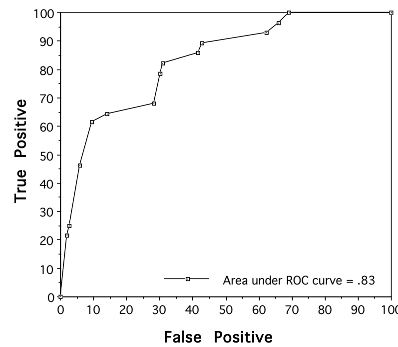

Figure 3b: Year 1 PAL Articulation Score as a Predictor of Year 2 PAL Articulation Score

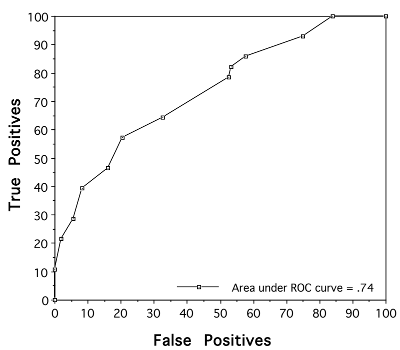

Figure 3c: Year 1 PAL Lexical Access Scores as a Predictor of Year 2 PAL Lexical Access Scores

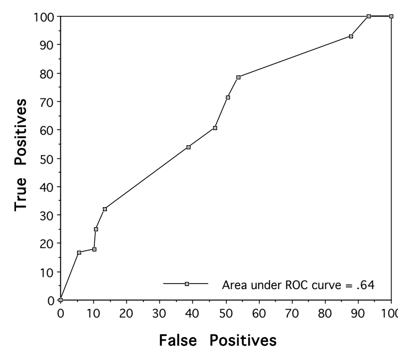

Figure 3d: Year 1 PAL Syntax as a Predictor of Year 2 PAL Syntax

FIGURE 3: Receiver Operating Characteristic Curves of the Predictive Power of PAL Scores

SUMMARY AND DISCUSSION OF EXPERIMENT 1

The results of Experiment 1 indicate that the PAL scores are a valid measure of children's current linguistic abilities and a good predictor of their future linguistic skills. Furthermore, these results suggest that the PAL test is a good screening test for language impairments, with PAL Oral scores correctly categorizing 81% of children as language-impaired or not. With the exception of PAL Syntax scores, correlation analyses revealed that children's initial scores on a PAL test were the best predictor of their scores on that same test a year later (T1-T2 PAL Articulation $r = .57$, T1-T2 PAL Lexical Access $r = .37$, T1-T2 PAL Oral $r = .52$). These correlation results are consistent with ROC results that revealed that children's initial PAL Oral, Articulation and Lexical Access scores were good at predicting which children had low and high scores 15 months later, whereas children's initial PAL Syntax scores were not.

The PAL tests were designed to provide measures of linguistic ability that are not 'contaminated' by nonlinguistic abilities. The fact that scores on each of the PAL tests were more highly correlated with linguistic measures than nonlinguistic measures indicates that the PAL tests do not simply measure children's overall development. The PAL tests also appear to do a fairly good job measuring the specific linguistic ability they were designed to measure. For example, PAL test scores were generally most highly correlated with the most similar language milestone (e.g., PAL Articulation-articulation milestone) and the composite PAL Oral measure was most highly correlated with the two general language measures (comparative language and ASQ communication).

Despite the auspicious findings outlined above, several concerns about the PAL remain. The first is that the most of the children who participated in Experiment 1 had perinatal risk factors for language delays and it is possible that the PAL test might not be valid for children without such risk factors. Second, most of the children in Experiment had parents who were wealthy, white, and well educated, and it is possible that the PAL test might not be valid for financially, racially, and educationally diverse children. Although this is worrisome, it should be noted that wealthy, white and well-educated families were also over-represented in many early studies of the CDI and LDS. The third concern is that PAL scores were compared with other parent-report measures. Parents' beliefs about their child's linguistic abilities might lead them to (consciously or unconsciously) give their child similar scores on the various language measures, thus inflating estimates of the PAL's validity. The fourth concern is that, even though PAL Oral scores were good at discriminating between children who were and were not language impaired and PAL scores were significantly correlated with most language measures, the correlation coefficients were generally modest to moderate, with only 3 being large (T1 PAL Oral-T1 ASQ Communication, T1-T2 PAL Articulation, T1-T2 PAL Oral).

The modest PAL correlations could be due to the fact that most of the children in Experiment 1 were premature, low birth weight and/or twins, and each of these factors is associated with language development that is more variable with respect to rate, developmental trajectory and final outcome than the language development of children without these perinatal risk factors (for a review, see Stromswold, 2006). Indeed, Dale et al. (1989) found that the correlation between CDI Expressive Vocabulary scores and Bayley Mental Development Scale language scores (Bayley, 1969) was almost twice as great for children who were born full term (r = .59) than children who were born preterm (r= .33). This is consistent with the large number of premature children in our study depressing PAL correlations.

Another possible explanation for the modest (but significant) PAL correlations is that PAL scores were compared with faulty language measures. Consider the hypothetical case in which PAL scores are perfect measures of children's linguistic abilities and the other language measures are poor. In such a case, the PAL-other language measure correlations will be low, even though PAL scores measure children's language perfectly. If this is the case, PAL scores should correlate more highly with better language measures. To address these concerns we conducted a second experiment in which children's PAL scores were compared with their scores on two standardized language tests

EXPERIMENT 2: A COMPARISON OF THE PAL AND STANDARDIZED LANGUAGE TESTS

METHOD

Participants

Twenty typically developing children between the ages 38 and 72 months (mean GA- corrected age = 54.2 months, SD = 9.4 months) were recruited from two daycare centers affiliated with Rutgers University. Eleven of the children were male and 9 were female. All of the children were monolingual English speakers and none had received speech or language therapy or a diagnosis of language impairment or delay. The children were not recruited because they had perinatal at-risk for developmental delays and, consistent with this, only 10% of the children were born prematurely (one at 32 weeks and one at 36 weeks), only one child was born low birthweight, and none were twins

As is to be expected from children recruited from daycare centers that serve faculty, staff, and students of a university, the children's parents were well educated, and all but one of the mothers and one of the fathers were either in college or had graduated from college. Seventeen (85%) of the children were non-Hispanic White, one child was Hispanic, one child was African American, and one child was Asian, again reflecting the demographics of Rutgers-affiliated monolingual English-speaking families. The family incomes of participants were consistent with the median national family incomes for 4-person families, with 15% of the families having family incomes less than $35,000, 30% having family incomes between $35,000 and $50,000, 30% having family incomes between $50,000 and $75,000, and 25% having family incomes that were greater than $75,000.

Tests and Testing Procedures

The children took the same parent-administered PAL tests as the children in Experiment 1. Three to 7 days after the children took the PAL, a trained experimenter administered two standardized tests to each child in his or her daycare center in a quiet room away from other children. The two standardized language tests were the Denver Articulation Screening Exam (DASE, Drumwright, 1971) and the first edition of the Clinical Evaluation of Language Fundamentals-Preschool Edition (CELF-P, Wiig, Secord, & Semel, 1992). Children always took the DASE before the CELF-P, and the person who administered the standardized tests did not know children's scores on the PAL test. All children completed all tests.

The DASE is a word repetition task that assesses 2.5- to 7-year old children's ability to correctly pronounce 30 consonants and consonant clusters that appear in word initial, medial and final positions. The CELF-P assesses preschool children's syntactic, morphological and lexical abilities. We administered 4 subtests of the CELF-P. The CELF-P's Formulating Labels subtest (renamed the Expressive Vocabulary subtest in the second edition of the CELF-P) is a test of children's expressive vocabulary. In this subtest, children say the noun or verb that corresponds to an object or action shown in a picture (e.g., what is this? answer: flag; what is this girl doing? answer: riding). The Sentence Structure subtest assesses children's receptive morphosyntactic abilities by having the 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 girl is being pushed by the boy*). The Word Structure subtest assesses children's morphological abilities by using .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 preposition *on* in *the hat is on the chair*, the pronoun *her* in *he is waving at her*, etc.). The closed class morphemes tested are prepositions (2 items), nominal inflections (2 items), verbal inflections and auxiliary verbs (9 items) and pronouns (7 items), In some of the Word Structure items, the child merely has to provide the correct closed class morpheme (e.g., *he is waving at <u>her</u>*) and in some items, the child must generate a lexical morpheme and a closed class morpheme (e.g., *this is <u>her bike</u>*).

The Linguistic Concepts subtest is a picture-pointing task in which children point to the picture(s) that corresponds to a sentence. Sentences in the Linguistic Concepts subtest require that children point to one picture (point to a dog, but not the one that is eating), two pictures (point to the elephant first and then point to the giraffe) or three pictures (point to the bear, the turtle, and the fish). The sentences become syntactically more complex over the course of the subtest and assess children's knowledge of conjunctions (point to a fish or a cat, point to the bear, the turtle and the fish), spatial prepositions (point to the elephant next to the giraffe, point to the animal in the middle), temporal terms (e.g., point to the cat and then to the bird), and quantifiers (point to one of the bears, point to some of the tigers). Thus, unlike the Formulating Labels subtest, which tests children's knowledge of nouns and verbs, the Linguistic Concepts subtest assesses children's knowledge of prepositions, quantifiers and conjunctions as well as their knowledge of nouns. Thus, the Linguistics Concepts test could be considered a test of children's lexical and morphosyntactic abilities.

Data Treatment

One person hand-scored the PAL tests and another person hand-scored the DASE and CELF-P. The person who scored the PAL test then scored the DASE and CELF-P, and the person who scored the DASE and CELF-P then scored the PAL test. When scoring the tests, neither person knew how the child had performed on the other type of test. For all tests, the inter-coder reliability was 98% or greater. In addition to determining each child's scores on individual PAL tests, we calculated an PAL Oral score for each child using the procedure described in Experiment 1. Similarly, following the procedure used in the second edition of the CELF-P (Wiig, Secord, & Semel, 2004), for each child, we obtained a composite Core Language Score (CLS) by summing the Sentence Structure, Word Structure and Formulating Labels (i.e., Expressive Vocabulary) scores.[4]

RESULTS

PAL Articulation. The mean PAL Articulation score was 10.05 (SD = 1.43, range 7-12). Consistent with the PAL Articulation test being a valid test of children's articulation, PAL Articulation and DASE scores were highly correlated (r - .66, p = .001, see Table 5).[5] In addition, PAL Articulation scores were highly correlated with CELF Formulating Labels scores (r = .53, p = .01), CELF Sentence Structure scores (r = .51, p = .02) and CELF CLS scores (p = .54, p = .01).

TABLE 5: Correlations Between PAL Test Scores and Standardized Test Scores

	PAL Articulation	PAL Lexical Access	PAL Syntax	PAL Oral
DASE	.66***	.06	.18	.43 (p =.06)
CELF Formulating Labels	.53**	.27	.11	.46*
CELF Word Structure	.33	.72***	.35	.77***
CELF Sentence Structure	.51*	.42 (p =.06)	.19	.59**
CELF Linguistic Concepts	.42 (p = .07)	.39 (p =.09)	.26	.55**
CELF CLS	.54**	.55**	.25	.70***
Mean r	.49	.40	.22	.58
Median r	.52	.41	.22	.57

* $p \leq .05$
** $p \leq .01$
*** $p \leq .001$
**** $p \leq .0001$

PAL Lexical Access. The mean PAL Lexical Access score was 5.80 (SD = 1.99, range 2-9). We had expected that PAL Lexical Access scores would be most highly correlated with CELF Formulating Labels scores, but this was not the case (r = .27). However, PAL Lexical Access scores were highly correlated with CELF Word Structure scores (r = .72, p = .0002), a test that requires that children generate lexical and grammatical morphemes. PAL Lexical Access scores were also highly correlated with CELF CLS scores (r = .55, p = .01).

PAL Syntax. The mean PAL Syntax score was 10.30 (1,42, range 8-12). Although PAL Syntax scores were not significantly correlated with any of the standardized test scores, PAL Syntax scores were most highly correlated with CELF Word Structure scores (r = .38, p = .10). At first blush, it is surprising that PAL Syntax scores correlated more highly with CELF Word Structure scores than CELF Sentence Structure scores. However, close examination of the two CELF subtests makes this seem less surprising.

[4] We also calculated a second CELF composite score by summing the scores for all 4 CELF subtests. We only present the correlations for the CELF CLS score because the correlations for the 4 subtest CELF composite score were virtually identical to those using the CELF CLS score.

[5] All children completed all subtests of the PAL and standardized language tests, and hence the N for all correlations is 20.

First reason is that 7 of the 20 Word Structure items involved pronouns as did 4 of the 12 items on the PAL Syntax test, whereas the only overlap between CELF Sentence Structure and PAL Syntax tests was passive sentences, with 2 out of the 22 items on the CELF Sentence Structure test being passive and between 4 and 6 of the 12 items on the PAL Syntax test being passive. The failure to find a significant correlation between the PAL Syntax – CELF Sentence Structure tests could also reflect that the test-retest correlation for the CELF Sentence Structure subtest is only .64 (CELF-P, Wiig, Secord, & Semel, 1992). In other words, 59% of the variance in CELF Sentence Structure scores is essentially noise ($1 – (.64)^2 = .59$), as compared to 34% of the variance for Word Structure scores ($1- (.81)^2 = .34$).

PAL Oral Scores. The mean PAL Oral score was 26.15 (SD = 3.15, range 12 – 32). PAL Oral scores were significantly correlated with all CELF subtest scores and the CELF CLS score (see Table 5), with correlations ranging from.46 (CELF Formulating Labels scores) to.77 (CELF Word Structure scores). Of particular note, the correlation between PAL Oral and CELF CLS scores was .70.

SUMMARY AND DISCUSSION OF EXPERIMENT 2

Given the small number of subjects, it is heartening that PAL Articulation and DASE scores were so highly correlated ($r = .66$). Give the size of the study and the differences in the tasks used and linguistic skills required by the tests, it is not surprising that of the 12 PAL-CELF subtest correlations, only 3 were significantly correlated (all r's > .50) and 3 were marginally correlated (all r's ~ .40). Given the fact that the PAL test is fast, inexpensive and can be administered by parents, whereas the CELF-P is long, expensive and must be administered by a trained professional, more surprising is that the correlation between PAL Oral and CELF CLS scores was as high as .70. A correlation of .70 means that children's PAL Oral scores accounted for almost half of the variance in their CELF CLS scores ($.70^2 = .49$). Comparison of the CELF-P and CELF-P2 manuals reveals that, for each CELF subtest that goes into calculating the CLS, the test-retest r is lower for the CELF-P than the CELF-P2. Thus, it is likely that the CELF-P CLS test-retest r is less than the CELF-P2's. For the CELF-P2, the CLS test-retest r is .90 (i.e., 19% of the variance is noise). Therefore, PAL Oral scores probably account for at least 60% the 'real' variance in CELF-P CLS scores (.49/.81 = .61).

GENERAL DISCUSSION

The results of Experiment 1 and 2 suggest that the PAL test has good concurrent and predictive validity. Unfortunately, we can't compare the validity of the PAL with other parent-administered language tests because the PAL is the only one. We also cannot make direct comparisons of the validity of the PAL and existing parent-completed language questionnaires because the ages of the children and aspects of language assessed are different. With these caveats in mind, how does the PAL test stack up against the language questionnaires?

Concurrent Validity. We can compare the concurrent validity of the PAL test and parent-completed questionnaires by determining the extent to which PAL test scores correlate with other language measures versus the extent to which parent-completed questionnaire scores correlate with other language measures. In order for such comparisons to be meaningful, the PAL-other language measure comparisons must be fairly similar to the questionnaire-other language measure comparisons.

Let us begin by comparing the concurrent validity of the PAL test with that of the CDI. Arguably, composite PAL Oral scores, ASQ Communication scores, our comparative language score, CELF CLS, CDI Using Language scores and McCarthy verbal scores are all measures of children's overall linguistic abilities. In Feldman et al. (2005), the correlation between CDI Using Language and McCarthy Verbal scores was .47. In Experiment 1 of this paper, the PAL Oral-ASQ Communication score correlation was .50, and the PAL Oral-comparative language correlation was .45 and, in Experiment 2, the PAL Oral Score- CELF CLS correlation was .70. The fact that the PAL Oral correlations were as high or higher than the CDI Using Language-McCarthy Verbal correlation suggests that PAL Oral Scores are as good or better than the CDI III Using Language scores as measures of children's overall linguistic abilities.

In their review of 6 studies that examined the extent to which CDI Expressive Vocabulary scores were correlated with scores on standardized tests of expressive vocabulary and expressive vocabulary measures derived from speech samples (e.g., type/token ratios; number of vocabulary items produced in a speech sample), Fenson et al. (1994) found correlations ranged from .33 to .85 (median = .61). However, in a larger study (Feldman et al., 2005), the correlation between CDI III Expressive Vocabulary scores and PPVT scores was considerably lower (r = .41). PAL Lexical Access scores are very different from CDI Expressive Vocabulary scores insomuch as CDI scores are the number of words a child has ever said and PAL scores are the number of words that match a verbal description a child retrieves from his lexicons during a timed test. Despite these differences, all of the correlations between PAL Lexical Access scores and scores on the CELF subtests (all of which have a lexical component) are on par with Feldman et al.'s (2005) correlation between CDI Expressive Vocabulary scores and Peabody Picture Vocabulary Test scores (PPVT, Dunn, Dunn, & Dunn, 1997)

We can also compare the PAL and CDI by comparing how tightly correlated PAL and CDI scores are with non-related standardized test scores. If PPVT and CELF Formulating Label scores are treated as measures of children's vocabularies (and once again CDI Using Language Scores and PAL Oral scores are treated as measures of children's overall linguistic ability), the extent to which children's overall linguistic abilities and lexical abilities are related is very similar for the CDI (CDI Using Language-PPVT r = .49 in Feldman et al., 2005) and the PAL (PAL Oral - CELF Formulating Labels r = .46). If CDI Sentence scores and PAL Syntax scores are both treated as measures of children's syntactic abilities, the correlations between children's syntactic and lexical abilities are considerably higher for the CDI (CDI Sentence – PPVT r = .48) than the PAL (PAL Syntax – CELF Formulating Labels r = .11). This difference may reflect differences between the PAL Syntax test (a picture-pointing comprehension task of syntactically minimal sentences) and CDI Sentence scores (in which parents pick which of two sentences sounds more like something their child might say). It may also reflect that PAL Syntax scores are purer measures of children's syntactic abilities than CDI Sentence scores.

How does the PAL fare relative to the LDS? Rescorla (1989) compared high SES children's scores on the LDS, the object naming and picture naming sections of the Bayley Mental Development Scale and the Reynell Expressive and Receptive Language Scales (Reynell & Huntley, 1971), and found that all of the correlations were significant and above .75. In a second study with lower SES children, the correlations between LDS and Bayley Object scores, LDS and Preschool Language Scales picture scores (Zimmerman, Steiner, & Evatt, 1969), and LDS and Bayley/LDS composite scores were all significant and above .70. Rescorla and Alley (2001) also found children's LDS scores were highly correlated with the number of Bayley objects and Stanford-Binet pictures (Thorndike, Hagen, & Sattler, 1986) that children named (r = .69 and .74, respectively). Furthermore, these correlations are considerably higher than the PAL Lexical Access-CELF Formulating Labels correlation (r = .27). However, this difference could reflect that the LDS and comparison vocabulary tests were more similar to one another than the PAL Lexical Access and CELF Formulating Labels tests. Consistent with this, the PAL Lexical Access-CELF Word Structure correlation (r = .72) was on par with the LDS correlations.

No published study has investigated the validity of the CSBS-DP checklist when compared to standardized test scores. However, Wetherby et al. (2002) compared children's composite speech scores on the CSBS-DP at 23 months with their scores 3 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-DP checklist and receptive language scores was .38 and the correlation between checklist scores and expressive language scores was .50 (both p's < .01). For the CSBS-DP 4-page questionnaire, the correlation was .47 for receptive language and .71 for expressive language (both p's < .01). The correlations between composite PAL Oral scores and CELF subtest and CELF composite scores (r's ranging from .46 to .77, with a median r of .59) were higher than the CSBS-DP checklist correlations and approximately the same as the CSBS-DP questionnaire correlations.

To the best of our knowledge, no studies have compared children's scores on Bishop's CCC-2 and their scores on standardized language tests. We also cannot compare the validity of PAL Articulation and Syntax scores as measures of children's articulation and syntactic abilities with those obtained from any of the parent questionnaires because either the questionnaires don't ask about children's articulation or syntax or no studies have investigated how highly correlated these scores are with other measures of articulation and syntax. In summary, to the extent that we can compare the PAL test with the CDI, LDS, CSBS-DP and CCC-2 questionnaires, the PAL appears to have similar concurrent validity.

Predictive Validity. We can use the longitudinal data collected in Experiment 1 to compare how well PAL scores at Time 1 correlate with PAL scores at Time 2 with T1-T2 correlations for parent-completed questionnaires scores. In Experiment 1, the correlation between PAL Lexical Access scores at Time 1 and Time 2 ($r = .37$) was virtually identical to the CDI 1 – CDI II Expressive Vocabulary correlation ($r = .39$, Feldman et al. 2000), but was lower than the CDI II – CDI III Expressive Vocabulary correlation ($r = .58$, Feldman et al. 2005). There are three plausible reasons why the CDI II-CDI III correlation is higher than the T1-T2 PAL correlation. First, the theT1-T2 interval was greater for the PAL than the T1-T2 interval for the CDI II and CDI III (15.2 months and 12.0 months, respectively). Second, the PAL Lexical Access tasks differ at T1 and T2 (name animals at age 3, name foods at age 4, name particular things at age 5), whereas there is substantial overlap in words on the CDI II and CDI III checklists. Third, most of the children in Experiment 1 had perinatal risk factors associated with variable language development (see above), whereas none of the children in Feldman et al. (2005) did.

PAL Syntax scores at T1 were significantly correlated with PAL Lexical Access scores at T2 ($r = .27$), as were CDI II Sentence Complexity – CDI III Expressive Vocabulary scores and CDI II Longest Sentence - CDI III Expressive Vocabulary scores ($r = .39$ and .47, respectively, Feldman et al 2005). The PAL Syntax scores at T1 and T2 were also significantly correlated ($r = .25$), as were CDI II Sentence Complexity -CDI III Sentence scores and CDI Longest Sentences – CDI III Sentence Scores (r's = .37 and .54, respectively). The T1 PAL Lexical Access – T2 PAL Syntax correlation was not significant ($r = .20$), whereas the CDI I Expressive Vocabulary – CDI II Sentence Complexity and CDI I Expressive Vocabulary – CDI II Sentence Length correlations were both significant ($r = .30$ and .33, respectively, Feldman et al 2000) as was the CDI II Expressive Vocabulary – CDI III Sentence score correlation ($r = .52$, Feldman et al 2005). However, that this may reflect the differences between the linguistic skills required by the PAL and the CDI, the greater T1-T2 interval for the PAL than the CDI, or the inclusion of perinatally high-risk children in Experiment. It could also be that PAL Lexical Access and Syntax scores are 'purer', more selective measures of children's lexical and syntactic abilities.

We can compare the T1-T2 PAL Oral score correlation and the T1-T2 CSBS-DP checklist composite speech score correlation (Wetherby, Allen, Cleary, & Kublin, 2002). Both T1-T2 correlations were significant ($r = .52$ for the PAL Oral composite and $r = .79$ for the CSBS-DP speech composite), with the difference in the size of the correlation plausibly reflecting (at least in part) the fact that the PAL T1-T2 interval was almost 5 times as great as the CSBS-DP T1-T2 interval (15.2 months vs. 4.1 months, respectively). As discussed below, LDS scores at age 2 were good predictors of clinical language impairment at age 3. No studies have compared the predictive validity of CCC-2 scores.

Validity as a language-screening test. Because none of the children in Experiment 2 had language impairments, only the data from Experiment 1 can be used to compare the validity of the PAL and parent questionnaires as screening tests for language impairments. Because standardized test data were not collected in Experiment 1, we restrict our discussion to studies that either compared children's questionnaire scores with clinicians' assessments of language impairment (which is more or less equivalent to our speech therapy measure) or scores on the questionnaire at Time 1 and Time 2. As demonstrated by the area under the ROC curves (AUC), PAL Oral scores were very good at distinguishing between children who do and do not have language impairments (AUC = .81), and for predicting PAL Oral scores 15 months later (AUC =. 83). T1 PAL Articulation scores are also very good predictors of T2 PAL Articulation scores (AUC = .83). The AUC for T1-T2 CDI II Expressive Vocabulary scores (.77, Feldman et al. 2005) is essentially the same as the AUC for T1-T2 Lexical Access scores (.74), indicating that scores on these measures are equally good at identifying children who have low scores on these measures at a future date (12 months and 15 months later, respectively).

PAL ORAL scores are also comparable with CCC-2 scores General Communication Composite (GCC) scores with respect to correctly classifying children with non-pragmatic language impairments. For GCC scores that are 2 standard deviations below the mean, the CCC-2 has a sensitivity of .31, a specificity of .96, and, when the rate of language impairment is 10%, a positive predictive power (PPP) of .49 and negative predictive power (NPP) of .93 when compared against speech-pathologists' diagnoses of specific language impairment (Bishop, 2006). PAL Oral scores that are 2 SDs below the mean (bottom 5th percentile) have a sensitivity of .20, a specificity of .98, a PPP of 59% and a NPV of 87%

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, when 'true' language impairment was determined by clinical evaluation, the LDS performed very well as a screening test at age 2 (sensitivity = .83; specificity = .93, PPV = 41%, NPV = 98%), and LDS scores at age 2 were good predictors of clinical language impairment at age 3 (sensitivity .67, specificity = .90, PPV = 67%, NPV = 93%). In comparison, for the bottom 10th percentile of PAL Oral scores, the sensitivity was considerably lower (.30), but the specificity (.94), PPV (45%) and NPV (88%) were comparable. In addition, PAL Oral scores were comparable in terms of their ability to predict PAL Oral scores one year later (sensitivity = .43. specificity = .95, PPV = 67%, NPV = 91%).

We cannot compare the ability of the PAL and the CSBS-DP checklist to discriminate between language impaired and non-language impaired children because the one relevant study of the CSBS checklist included children who scored poorly on the Social, Symbolic and/or Speech composite sections and compared their scores against a professional's global assessment of the children's status based on their behavior during structured play sessions (Wetherby, Goldstein, Cleary, Allen, & Kublin, 2003). In other words, this study investigated the validity of the CSBS checklist as a screening tool for a wide range of impairments and not the its validity as a language-specific screening tool. In summary, taken together, the results of Experiment 1 and 2 indicate that the concurrent and predictive validity of the PAL test is comparable with that of frequently used, parent-completed language questionnaires.

Summary. The results of Experiment 1 and 2 suggest that PAL scores may be valid measures of preschool children's current linguistic abilities and good predictors of their future abilities. Furthermore, these results suggest that the PAL may be a good screening test for current and future language impairments. These findings - coupled with the fact that the PAL test is a fast, parent-administered test – mean that the PAL holds promise as a useful tool for characterizing the linguistic abilities of preschool children and distinguishing between children who do and do not have language impairments. As such the PAL test may fill an important niche in the arsenal of language assessment tools available to researchers and clinicians.

However, before concluding, let us return to the concerns about the PAL raised in the discussion section of Experiment 1. The concern that the PAL test might not be valid for children without perinatal risk factors for language delay seems to be unfounded, because the children in Experiment 2 did not have these risk factors yet the PAL correlations were considerably higher in Experiment 2 than Experiment 1. Second, the worry that the PAL might not be valid for children from diverse economic backgrounds is allayed by the fact that the family incomes of the children in Experiment 2 reflect those of the US population at large. The third concern is that the correlations obtained in Experiment 1 were generally modest. Here, it is important to note that most large studies of the CDI have found only modest correlations between the CDI and other language measures, yet the CDI is very widely used. But, more importantly, the correlations obtained in Experiment 2 were considerably larger than those obtained in Experiment 1, with the PAL Articulation-DASE correlation (r = .66) and the PAL Oral- CELF-P CLS correlation (r = .70) comparing favorably with those obtained for the parent questionnaires. The higher correlations in Experiment 2 than Experiment 1 also allays fears that the PAL correlations in Experiment 1 were inflated by the fact that PAL scores were compared against other parent-report language measures.

Three concerns remain. First, although the number of children who participated in Experiment 2 is similar to the number who participated in most early studies comparing language questionnaire scores with standardized language scores, the number of children in Experiment 2 is small nonetheless. Thus, it is possible that the high correlations found between PAL scores and standardized test scores in Experiment 2 are statistical flukes. Second, white, non-Hispanic children were over-represented in Experiment 1 and 2 and, thus, it is possible that PAL scores may not be valid indicators of the linguistic abilities of children from ethnically and racially diverse backgrounds. Again it should be noted that this is also true of many of the early studies of the language questionnaires. The third concern is that the parents who participated Experiment 1 and 2 were generally well educated. Although the parents in Experiment 1 and 2 had no difficulty administering the PAL test, poorly educated parents might. Clearly, these concerns must be addressed in future studies. However, the results of Experiment 1 and Experiment 2 are auspicious enough and the research and clinical utility of the PAL test great enough that doing such studies is warranted.

REFERENCES

Aram, D. M., & Nation, J. E. (1980). Preschool language disorders and subsequent language and academic difficulties. *Journal of Communication Disorders, 13*, 159-170.

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

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.

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., Bates, E., Reznick, S., & Morisset, C. (1989). The validity of a parent report instrument of child language age 20 months. *Journal of Child language, 16*, 239-250.

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.

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., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*(Serial number 242).

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.

Hall, P., & Tomblin, J. B. (1978). A follow-up study of children with articulation and language disorders. *Journal of Speech and Hearing Disorders, 43*, 227-241.

King, R. R., C., J., & Lasky, E. (1982). In retrospect: A fifteen-year follow-up report of speech-language disordered children. *Language, Speech and Hearing Services in the School, 13*, 24-32.

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.

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.

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

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.

Precious, A., & Conti-Ramsden, G. (1988). Language-impaired children's comprehension of active versus passive sentences. *British Journal of Disorders of Communication, 234*(3), 229-243.

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.

Rescorla, L. A., Ratner, N. B., Jusczyk, P., & Jusczyk, A. M. (2005). Concurrent validity of the Language Development Survey: Associations with the MacArthur—Bates Communicative Development Inventories Words and Sentences. *American Journal of Speech-Language Pathology, 14*, 156-163.

Reynell, J., & Huntley, R. M. C. (1971). New scales for the assessment of language development in young children. *Journal of Learning Disabilities, 4 (10)*, 10-18.

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.

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

Stanton-Chapman, T., Chapman, D., Bainbridge, N., & Scott, K. (2002). Identification of early risk factors for language impairment. *Research in Developmental Disabilities, 23*(6), 390-405.

Stromswold, K. (1997). Specific language impairments. In T. E. Feinberg & M. J. Farah (Eds.), *Behavioral neurology and neuropsychology* (pp. 755-772). New York: McGraw Hill.

Stromswold, K. (2006). Why aren't identical twins linguistically identical? Genetic, prenatal and postnatal factors. *Cognition*.

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

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

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.

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., & Semel, E. (1992). *Clinical Evaluation of Language Fundamentals – Preschool*. San Antonio, TX: The Psychological Corporation.

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

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

APPENDIX

PAL 3 Test Items

Articulation task: *fat*, *soap*, *yuck*, *van*, *rat*, *lip*, *ship*, *cheek*, *zip*, *jeep*, *that*, *thin*
Lexical access task: Name animals quickly
Syntax task:

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 Test Items

Articulation task: *rat*, *lip*, *ship*, *cheek*, *zip*, *jeep*, *that*, *thin*, *split*, *trick*, *clock*, *frog*
Lexical access task: Name foods quickly
Syntax task:

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 Test Items

Articulation task: *split*, *trick*, *clock*, *frog*, *three*, *shrink*, *brake*, *flat*, *twin*, *street*, *scrub*, *squat*
Lexical access task: Name a part of a face, a fruit, a color, an animal, something round, a part of a car, a letter, something red, a toy, something big
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