A Language Screening Protocol for Use With Young African American Children in Urban Settings

Julie A. Washington
Holly K. Craig
University of Michigan, Ann Arbor

Language screenings represent an important tool for early identification of language impairments in young children between 3 and 5 years of age. This investigation examined the utility of a well-established set of assessment measures for screening young African American children. One hundred and ninety-six children participated in the screening. Based upon the outcomes of the screening, 25 children who failed and a random sample of 56 children who passed were administered a larger language and cognitive assessment battery. Sensitivity and specificity of the screening were determined to be high. The number of different words, the Kaufman Nonverbal Scale, and nonword repetition accounted for a significant amount of the variance in performance. The screening is brief, valid, and culturally fair for use with preschool- and kindergarten-aged African American children living in urban settings.

Key Words: African American, children, screening, language

Identification of speech or language assessments that are fair and appropriate for use with children who present cultural-linguistic differences have been the focus of inquiry for many years (Vaughn-Cooke, 1986). Increased attention in recent years to the assessment of language in African American children who use African American English (AAE), in particular, has resulted in significant breakthroughs in our understanding of the impact of dialect and the appropriateness of selected standardized and nonstandardized assessments for use with these children (Cole & Taylor, 1990; Seymour & Bland, 1991; Seymour, Bland-Stewart, & Green, 1998; Stockman, 1997; Washington & Craig, 1992a, 1992b). Our own work has focused largely upon developing and evaluating assessment instruments and establishing expectations for language performance of young African American children at the time of school entry (Craig & Washington, 2002; Craig, Washington, & Thompson-Porter, 1998a; 1998b; Washington & Craig, 1992a, 1992b, 1998, 1999). These studies have resulted in the development of a battery of culturally fair expressive and receptive language measures that include both standardized instruments and non-standardized, criterion-referenced assessments.

In the field of speech and language pathology, assessment of African American children has focused primarily upon distinguishing cultural dialect use from disorder in children identified as specifically language impaired (SLI). Past attempts to make this important distinction have been complicated by the fact that several key features of AAE dialect closely approximate, at the surface structure level, the key features of impairment identified in the extant literature for children with SLI (Oetting & McDonald, 2001; Washington, 1996). A recent investigation of dialect and SLI in African American children in a rural, Southern community revealed that the surface manifestations of these morphosyntactic features differ for children who are normal language users of AAE and those identified as SLI (Oetting & McDonald, 2001), making it possible to distinguish these children from each other. This exploratory investigation has been one of a very small number that has examined AAE features for their potential to contribute to the identification of African American children with SLI. More commonly, researchers have focused on the promise of language structures unaffected by AAE for their ability to assist clinicians and researchers in their attempts to distinguish African American children with SLI from their normal language peers (Campbell, Dollaghan, Needleman, & Janosky, 1997; Craig & Washington, 1994, 2000, 2002; Seymour et al., 1998; Stockman, 1997). These investigations have provided important new information about the language skills of

African American children while avoiding the difficulties presented by dialect in the assessment process.

One important aspect of language evaluation that has received no attention in the extant literature for this population is language screening. Screenings are an important means by which to identify the presence of disease or disability as early as possible to avoid later academic, social, or emotional difficulties (Klee, Pearce, & Carson, 2000; Rescorla & Alley, 2001; Stott, Merricks, Bolton, & Goodyer, 2002). Valid and reliable screenings should (1) provide general information about speech and language skills, (2) identify children who need additional language assessment, and (3) eliminate children whose speech and language skills are age appropriate. Optimally, early language screenings should occur between the ages of 3 and 5 years (Camilleri & Law, 2001; Law, Boyle, Harris, Harkness, & Nye, 2000). Among these very young children who may be experiencing developmental delays, those whose only difficulty is language are least likely to be referred for assessment (Rescorla & Alley, 2001). This is unfortunate, since early language impairments are highly correlated with later academic difficulty (Catts, 1993; Catts, Fey, Tomblin, & Zhang, 1997; Catts, Fey, Zhang, & Tomblin, 1999; Shankweiler et al., 1995). Early identification through screening might successfully mitigate some of the problems faced by many children during later schooling. A good screening should identify both children with stark impairments and children whose performance places them at risk for disorder, minimizing the costs associated with conducting full assessments on all children.

The overrepresentation of African American children on special education caseloads in the 6- to 9-year-old age group is an important indicator of the need for early and accurate identification of disabilities. African American children are overrepresented in all 13 disability categories, making up 20.3% of students receiving services compared to only 14.5% of the national resident population (U.S. Department of Education, 2002). The most striking disparities for African American children are in the categories of mental retardation and developmental delay (which includes speech and language impairments), where they make up 34.2% and 30.5% of children served, respectively. Among 3- to 5-year-olds enrolled in early intervention preschool environments, African American children are overrepresented to a much lesser extent than these older children (15.7% of early intervention population vs. 13.7% of general preschool population; U.S. Department of Education, 2002). It would be encouraging if these data were an indication that our assessments at preschool and kindergarten age were more reliable for identifying African American preschoolers and kindergartners than those used with older children. More likely, these data suggest that like their majority peers African American children with language impairments are unlikely to be identified between the ages of 3 and 5. When identification does occur between 6 to 9 years of age, it is inaccurate and unreliable. The goal of this investigation was to examine the effectiveness of a language screening protocol for identifying young African American preschool- and kindergarten-aged children residing in an urban setting who exhibit early problems with language development.

In spite of the value of early identification through screenings, calls for universal screening of language in young children are commonly regarded to be premature. A review of the screening literature by Law et al. (2000) determined that, while beneficial, even the most accurate screenings are more useful for identifying children who do not have language impairments than for identifying those who do. Thus, the sensitivity of most instruments (i.e., the proportion of cases with language impairments that are classified correctly by a screening; also called *screen positive*) was generally lower than specificity (i.e., the proportion of true noncases who screen negative). Most screenings tended to overidentify children as language impaired.

In general, overreferrals are inevitable (Klee et al., 2000; Law et al., 2000) and are preferable to under-referral. Most screenings overidentify children who upon further assessment are determined to have no language deficits (false positives). Among children who have language problems, the more severe manifestations, not surprisingly, are easiest to detect, and the more subtle language problems are likely to be missed (Law et al., 2000; Westerlund & Sundelin, 2000). The data for African American children suggest that this trend may be true cross-culturally. In addition, for children from minority groups, misclassification based on inappropriate interpretation of cultural interaction and communication styles may contribute to difficulty with detection of true language problems (Chamberlain & Madeiros-Landurand, 1991).

Standardized instruments have been identified consistently as preferable for use in screening contexts because they are presumably reliable (Stott et al., 2002; Tomblin, Records, & Zhang, 1996). However, the validity of these instruments for use with African American children has been questioned, and using a combination of standardized and nonstandardized assessments has been recommended. One recently developed standardized instrument, the Developmental Evaluation of Language Variation (DELV; Seymour, Roeper, & deVilliers, 2003), includes a screening test that can be used with African American children. No systematic, empirical evaluation of its reliability is available yet. Despite this, such a screening is long overdue and should make a significant contribution to our ability to identify more accurately African American children in need of further evaluation and intervention services.

Most tests used to assess children, especially in schools, reflect mainstream cultural experiences and may not result in valid outcomes (Chamberlain & Madeiros-Landurand, 1991). More dynamic and criterion-referenced approaches to assessment have been endorsed and determined to be reliable for use with minority children, including African Americans (Peña & Quinn, 1997). The utility of these approaches for screening purposes is untested. Screenings frequently involve direct assessment, observation, parent report, or some combination of these. Regardless of the form that it takes, the key elements of a good screening are reliability, validity, ease of administration and interpretation, and accuracy of results (Cochrane & Holland, 1971; Westerlund & Sundelin, 2000).

The current investigation examined key assessments described in our earlier work for their usefulness in a screening context. The reliability and validity of these tasks for use with African American children have been established, a key element for any screening instrument whether standardized or nonstandardized, and they were easy to administer and interpret. Although there are currently no established guidelines for language areas to examine in screening contexts, as noted by Tomblin et al. (1996), it is customary to measure skills in both the receptive and expressive domains. The screening protocol described here includes evaluation in both domains.

The following questions were posed:

1. Does the screening protocol successfully distinguish African American children who are having language difficulty from their peers who have adequate language skills?

2. What is the sensitivity and specificity rate of the screening for accurately classifying African American preschoolers and kindergartners who exhibit early language problems?

3. Is this rate influenced by different cutoff values?

4. What is the influence of important social and demographic factors such as socioeconomic status (SES), age, and gender?

Study 1

The purpose of Study 1 was to examine and describe the performance of a sample of young African American children on a language screening instrument. The speech and language measures and data collection methods used in the screening protocol have been described in the extant literature and determined to be culturally fair and valid for use with African American children (Craig & Washington, 2000, 2002; Washington & Craig, 1992b; Washington, Craig, & Kushmaul, 1998).

Method

Participants. A community-based sample of 196 African American boys ($n = 91$) and girls ($n = 105$) who were enrolled in either preschool ($n = 140$) or kindergarten ($n = 56$) in a Metropolitan Detroit school district participated in Study 1. All participants were between 46 and 79 months of age ($M = 55.7$, $SD = 7.2$).

Consent forms were sent home by teachers, principals, or program directors with all of the African American preschoolers and kindergartners over 2 consecutive school years. One child whose parents consented to her participation but was identified by her parents as hearing impaired was excluded from the screenings. In addition, 12 children were screened twice, once in preschool and then again in kindergarten. For these children only data from their most recent screening were included in this investigation. All other children whose parents consented to their participation were included in this data set, for a total of 196. This methodology is preferable to identifying children based on teacher or clinician referral, as referred and clinical samples may be biased.

The SES of most participants was unknown at the time of the screening. SES information was obtained for students who participated in the larger language assessment phase of the study. Those demographics and data are presented in Study 2.

Data collection. Screening data were collected in the participants' school buildings. Whenever possible, screenings were conducted in quiet classrooms or libraries in each school building. However, in some schools where space was severely limited, screenings were conducted in the hallway of the lower elementary wing of the school. Although this screening context is not ideal, it is reflective of the space limitations encountered when large-scale data collections are conducted in public school contexts.

The screening protocol was designed to provide important information about each child's receptive and expressive language, and speech skills, as follows:

1. The Peabody Picture Vocabulary Test—Third Edition (PPVT–III; Dunn & Dunn, 1997) was administered using standard administration procedures and provided a measure of receptive vocabulary.

2. The Wh-Question Comprehension task (Craig & Washington, 2000, 2002; Craig et al., 1998b) was administered and scored using a laptop computer and two action pictures depicting a winter scene (shoveling) and a summer scene (barbecuing). The task, which probes responses to 12 Wh-questions, and the pictures were developed at the University of Michigan as a part of the Michigan-Project on African Language (M-PAL). The probes range from simple naming questions (e.g., What's that?) to questions that are more demanding cognitively (e.g., When this happenin'?). Probes are presented by an adult examiner in computer-generated random order. Point scores ranging from 0 to 3 were possible for each item. It was demonstrated previously that responses on the pictures were not significantly different, and the resultant raw scores could therefore be combined (Craig et al., 1998b), resulting in a total raw score of 72 points possible on this task. See Craig et al. (1998b) and Craig and Washington (2000) for a more complete description of scoring.

3. A Picture Description task was used to obtain a gross measure of mean length of C-unit screen in words (Craig & Washington, 2000; Craig et al., 1998a). A C-unit was defined as independent clauses, plus their modifiers, individual responses to adult questions, and acknowledgment by the child of a previous adult comment (Craig & Washington, 2000). Each child was shown two pictures and asked to describe them. The examiner recorded the three longest utterances produced by the child across the two pictures in order to derive the measure that is denoted as *MLCUw-s3* (mean length of C-unit in words for the three longest screening utterances).

4. The Triangles subtest of the Kaufman Assessment Battery for Children (K-ABC; Kaufman & Kaufman,

1983) was administered as a general screen of nonverbal cognitive ability to students who were at least 4 years of age at the time of the screening. This subtest examines nonverbal pattern matching ability and has been identified as a culturally fair measure of nonverbal cognition for use with African American children (Craig & Washington, 2000).

5. The Face Recognition subtest of the K-ABC was administered as a general screen of nonverbal cognitive ability to students who had not yet turned 4 years old at the time of the screening. This subtest examines the child's ability to identify the faces of individual children and adults presented in visual contexts with increasingly dense content. This nonverbal subtest has been determined to be culturally fair for use with young African American children (Lampley & Rust, 1986).

The screening took approximately 20 min to administer. Scoring was completed soon after screening to identify quickly any children who failed the screening. The gold standard for failure on the screening was a score below age expectations on any two of the screening measures, as follows:

1. PPVT–III: a standard score more than 1 *SD* (≤85) below the established mean of 100, in accordance with established scoring guidelines.

2. Wh-Question Comprehension: a score more than 1 *SD* below the established population mean (≤49), as reported in Craig and Washington (2000).

3. MLCUw-s3: a score more than 1 *SD* below the sample mean on this gross measure of expressive language. The mean on this measure was derived using criterion-referenced methodology such that an overall mean and standard deviation for the sample of children who participated in this investigation were derived (*M* = 7.01, *SD* = 2.01). The mean and standard deviation were considerably higher than previously reported, as they were based on only three utterances. Children who scored more than 1 *SD* from this mean were considered to have failed this portion of the screening.

In addition, the following two cognitive screening measures were used both to confirm the cognitive level of the participants and to identify children whose performance fell below established guidelines for SLI. Children with cognitive performance falling below established guidelines for SLI were excluded from further testing.

1. Triangles subtest of the K-ABC: a score more than 2 *SD* below the mean scaled score of 10 (*SD* = ±3) was adopted as recommended by Tager-Flusberg and Cooper (1999) to allow for identification of more children with language impairment. Children who scored below this criterion (i.e., a scaled score < 4) were not considered for further testing.

2. Face Recognition subtest of the K-ABC: a score more than 2 *SD* below the mean scaled score of 10 (*SD* = ±3) as recommended by Tager-Flusberg and Cooper (1999). Children who scored below this criterion (i.e., a scaled score < 4) were not considered for further testing.

All data were collected by female examiners who had backgrounds in speech and language and who had experience working with young children. Race of the examiners was allowed to vary. Examination of overall performance revealed that race of examiner did not influence whether students passed or failed the screening [independent *t*(194) = –1.31, *p* = .193].

Results

Examination of performance on the two subtests of the K-ABC (Triangles and Face Recognition) revealed that none of the children screened met the criterion for exclusion based on low nonverbal cognitive performance (*M* = 10.16, *SD* = 2.3, range = 5–19). Data were examined also for possible differences in performance on the screening measures by grade and gender. No gender differences were apparent on nonverbal cognitive measures [*F*(194) = .409, *p* = .523], Wh-questions [*F*(194) = 1.45, *p* = .230], PPVT–III [*F*(194) = .389; *p* = .534], or the MLCUw-s3 [*F*(194) = .084; *p* = .772]. As expected, developmental differences in performance on the language measures were evident. The group means for the kindergartners were significantly better than those of the preschoolers on the PPVT–III, on the MLCUw-s3, and on the Wh-q task (see Table 1). No grade differences in performance were evident on Triangles [*F*(195) = .043, *p* = .836].

Closer examination of individual performance indicated that 96 of 196 participants (49%) scored within the normal range on all 4 of the screening measures, and 64 children (33%) failed at least 1 measure and so did not meet the gold standard for failure of the screening. In total, 36 children (18%) met the criterion for failure of the screening by scoring below expectations on 2 or more of the measures. Of the 36 children who failed, 18 were girls (50%) and 18 were boys, indicating identical failure rates by gender. Most children who failed were preschoolers (89%). Overall, preschoolers composed 82% of the sample. The mean scores of the participants in the pass and fail groups are presented in Table 2. Most students who failed the screening (78%) scored below expectations on the PPVT–III and Wh-q (see Table 3).

Discussion

All 196 children screened in Study 1 performed in the "passing" range (scaled score ≥ 4) on the cognitive subtests of the K-ABC, indicating that no child failed the screening due to low cognitive ability. Overall, 18% of children met the criterion for failure of the screening: performance below expectations on at least two tasks. This percentage of failures was within the range (8–20%) reported for screening performance in the extant literature (Klee et al., 2000), but considerably higher than the estimates previously reported in the extant literature for African American children (Tomblin et al., 1997). Many of the children in this investigation were younger than the kindergartners screened by Tomblin and his colleagues, and this may have made a difference in the failure rate.

TABLE 1. Means, standard deviations, *F* values, effect sizes (η), and significance levels (*p*) of preschoolers' and kindergartners' performance on the Peabody Picture Vocabulary Test—Third Edition (PPVT–III), responses to Wh-questions (Wh-questions), mean length of C-units screening test (MLCUw-s3), and the Triangles subtest of the K-ABC.

Task	Group	M	SD	F	η	p
PPVT–III	Preschool	93.27	11.17	5.33	.073	.022*
	Kindergarten	96.22	11.09			
Wh-questions	Preschool	49.36	9.25	29.25	.140	.0002*
	Kindergarten	55.88	7.52			
MLCUw-s3	Preschool	6.61	2.02	8.42	.004	.004*
	Kindergarten	7.42	2.14			
Triangles	Preschool	10.27	2.29	.04	.000	.843
	Kindergarten	10.20	2.40			

*Statistically significant between-group differences.

Camilleri and Law (2001) recommend repeated and ongoing monitoring to detect problems that may emerge over time, as well as those that resolve with time and schooling. A good screening should overidentify rather than underidentify students for further assessment consideration without identifying so many false positives that caseloads become overburdened (Klee et al., 2000; Law et al., 2000). Further, although there may be a small number of students who do not perform well on the screening and are later determined to have normal language abilities during follow-up assessments, all children who have language skills below the expected range should be reliably identified in a screening (true positives). In order to examine the success with which the screening accomplished its goal of identifying true positives and true negatives, a second assessment was completed and is described below in Study 2.

Study 2

The purpose of Study 2 was to determine the accuracy with which the screening tasks differentiated children with language impairments from their peers who had age appropriate language skills. All children who failed the screening were eligible for and consented to additional language and cognitive testing. In addition, one-third of the children who passed the screening were randomly selected to receive additional testing as well. This methodology has been used successfully by others (Tomblin et al., 1997).

Method

Participants. All 36 students who failed the screening instrument were targeted for inclusion in Study 2. However, of those 36 children, 11 were preschoolers who left school before the additional testing could be completed, approximately 1 month later. Periodic checks by project staff determined that these preschoolers did not return at any point during the school year. Thus, 25 of the original 36 children who failed the screening were available for full assessments and were included in Study 2. A power analysis statistic was completed to confirm that this sample size would be sufficient to detect differences in performance between these children and those who passed the screening at the .05 confidence level [power = 0.83] and at the .01 confidence level [power = 0.80]. In addition, 56 children who passed the screening were randomly selected and participated in Study 2 as well, for a total of 81 participants (i.e., 25 children who failed the screening, and 56 who passed the screening in Study 1). SES data were obtained from children using a questionnaire that was completed by their parents during a face-to-face interview in the home or based on eligibility for the federal free and reduced lunch program as indicated by the school. From interview data, the caregiver education, marital status, gender, and occupation were collected and used to obtain a socioeconomic index score using the Hollingshead Four-Factor Index of Social Status (Hollingshead, 1975). SES data were obtained for 80 of the 81 children (99%) using these methods. The caregiver of one child who failed the

TABLE 2. Means and standard deviations for participants who passed and those who failed on the screening tasks.

Task	Pass (*n* = 160)		Fail (*n* = 36)	
	M	SD	M	SD
PPVT–III	96.9	8.9	78.1	8.2
Wh-questions	52.2	7.7	42.7	9.6
MLCUw–s3	7.04	2.2	6.04	2.1
Triangles	10.3	2.3	9.5	2.4

Note. PPVT–III = Peabody Picture Vocabulary Test—Third Edition; MLCUw–s3 = mean length of C-units screening test.

TABLE 3. Number, percentage, and combination of tasks failed by students who failed the overall screening.

Failed tasks	n	%
PD + PPVT–III + Wh-q	1	3
PD + PPVT–III	4	11
PD + Whq	3	8
Wh-q + PPVT–III	28	78

Note. PD = Picture Description; PPVT–III = Peabody Picture Vocabulary Test—Third Edition; Wh-q = Wh-Question Comprehension Task.

screening did not provide sufficient information for determining the socioeconomic index score. Overall 36 (44%) of the children participating in this phase of the study were male, and 45 (56%) were female. Boys (*n* = 11) and girls (*n* = 14) were nearly equally represented in the group of children who failed the screenings. Among the 56 who passed, 25 were male and 31 were female. Further, 48 (59%) of the children were low income, and 32 (39%) were middle income.

Data collection.

1. The Kaufman Nonverbal Scale (KNVS; Kaufman & Kaufman, 1983) was used to assess nonverbal cognition. Nonverbal skills were targeted in order to avoid penalizing children who had language impairments and/ or were dialect users. Children with language impairments and children from various cultural and linguistic communities were included in the standardization sample for this instrument. The subtests of the KNVS have been examined and determined to be culturally fair for use with African American children (Willson, Nolan, Reynolds, & Kamphaus, 1989). Scoring is based on a mean standard score of 100 with a standard deviation of 15. Again using the criterion suggested by Tager-Flusberg and Cooper (1999), students who obtained a standard score ≥ 70 were deemed to have normal cognition.

2. A 20-min spontaneous language sample was collected during a freeplay interaction in a quiet room at the participants' schools. Children were presented with 3 toy sets and asked to select the one with which they wanted to play. Language samples were audiotaped using an audiotape recorder equipped with two individual head microphones that were worn by the child and the adult female examiner during the play interaction. Although most examiners were African American, race of the examiner was allowed to vary. A statistical analysis of the numbers of utterances produced during free-play suggested that the race of examiner did not affect overall productivity for the participants [independent *t* (88) = 1.78, *p* = .070]. Examiners included certified speech-language pathologists as well as students who received extensive training in data collection methods. Three traditional expressive language measures were derived from analysis of the language samples: *mean length of communication units in words* (MLCU) was calculated on the first wholly intelligible 50 C-units; *frequency of complex syntax* (Csyn) was scored for each 50 C-unit sample using Craig and Washington's (1994) scoring criteria; and *number of different words* (NDW) in the 50 C-unit sample was calculated as a measure of lexical diversity. Craig and Washington (2002) developed performance expectations for kindergartners and preschoolers on each of these oral language measures based on cutoff values of 1 *SD* below the mean and 1.25 *SD* below the mean. These cutoffs were examined in this investigation for their influence on the accuracy with which the screening identified children who were having language difficulty. Specific scores expected for each cutoff value are provided in Table 4.

3. Comprehension of Active/Passive Sentences, a task designed to examine young children's comprehension of reversible sentence (REVS) structures, was administered as well. Ten pairs of pictured actions were presented, accompanied by 30 spoken prompts. The line drawings depicted actions that were bidirectional and ecologically valid when expressed in either passive or active form (e.g., "the boy hit the car," "the car hit the boy," "the car was hit by the boy"). The total possible point score was 20: 1 point for responding correctly to both active trials and 1 point for correct responses on the passive trial. See Craig et al. (1998b) and Craig and Washington (2000) for additional discussion. Expected scores for cutoff values for typical performance and performance at –1 *SD* and –1.25 *SD* below the mean are provided in Table 4.

4. The Nonword Repetition Task (NRT), a processing dependent measure that examines a child's ability to repeat nonwords of differing lengths, was administered to each participant. The NRT has not been included previously in our assessment battery, but it has been examined and identified as culturally fair for use with school-aged African American children (Campbell et al., 1997). Its sensitivity and specificity are reportedly high for identifying African American children who are having language difficulty (Dollaghan & Campbell, 1998). This investigation will provide data about this measure when used with preschool and kindergarten children. Sixteen nonwords as described in Dollaghan and Campbell were presented under headphones through an audiotape recorder in fixed but random order. The nonwords were recorded onto an audiotape by an African American female, because previous research has shown that using a voice that the listener perceives to be ethnically different may influence the accuracy of responding for adults (Gerken, 1979), raising the possibility that child responding also might be affected. The participants were instructed to repeat the nonwords, which were recorded into a second audio recorder using a head microphone. Nonsense words varied in syllable length from one to four syllables. Percent phonemes correct (PPC) were calculated by dividing the number of phonemes produced correctly by the total number of phonemes in the prompt. The mean

TABLE 4. Means and standard deviations on each assessment measure, at 1 *SD* below the sample mean, and at 1.25 *SD* below the mean.

	M	*SD*	–1 *SD*	–1.25 *SD*
KNVS	99.5	12.6	86.9	83.7
REVS	11.5	4.1	7.1	6.1
MLCU	3.22	.66	2.56	2.40
Csyn	4.0	2.8	1.2	0.5
NDW	74.6	13.4	61.2	57.9
NRT	82.4	9.5	72.9	70.5

Note. KNVS = Kaufman Nonverbal Scale; REVS = reversible sentences; MLCU = mean length of C-unit; Csyn = complex syntax; NDW = number of different words; NRT = nonword repetition task.

NRT score for children who passed the screening was 84.1% phonemes correct (*SD* = 8.9); for children who failed it the mean score was 78.3% phonemes correct (*SD* = 9.9). Accordingly, the gold standard passing score adopted by Dollaghan and Campbell, 81%, was adopted for this investigation as well.

5. The Arizona Articulation Proficiency Scale—Third Edition (AAPS–3) was used to confirm typical articulation development of participants. Performance on the AAPS-3 was examined to ensure that these young participants did not evidence difficulty producing phonemes targeted on the NRT. An earlier investigation indicated that this instrument successfully and fairly characterizes the articulation skills of African American children who use AAE (Washington & Craig, 1992a). Published scoring guidelines were used. Thus, children with a standard score ≥ 85 were judged to have normally developing articulation skills.

Reliability. Approximately 20% of transcripts (*n* = 18) were randomly selected and were both recoded and retranscribed by an independent observer. Transcripts for 7 children who failed the screening and 11 who passed were included. Transcription reliability was obtained both at the level of the morpheme and for C-unit segmentation. An independent observer randomly selected 3 transcripts from among the 18 described above and retranscribed the first 50 C-units. Agreement between the original transcriptions and the retranscriptions was 95.2% at the level of the morpheme, and for C-unit segmentation it was 96.2%. Recoding of the data included point-to-point comparisons for each scoring system as well by dividing the number of disagreements by the number of disagreements + agreements. Percentage agreement for Csyn types was 97%, and for Csyn tokens agreement was 100%. Agreement for NDW was high at 99% agreement, and for the NRT agreement was 94%.

Results

Gender, SES, and age. The scores for each assessment measure were converted to *z* scores in order to standardize the scale for comparison purposes. Performance was examined on each measure for the potential influence of gender and SES. There were no significant main effects or interaction effects of gender for the group on KNVS, Csyn, MLCU, REVS, NDW, or NRT (see Table 5). Further, there were no significant main effects or interaction effects for SES on any of these measures (see Table 6). Examination of grade differences on these measures revealed no significant interaction effects, and one significant main effect for the NRT [$F(1, 80) = 6.48$, $p = .013$]. The total percent phonemes correct on the NRT was significantly higher for kindergartners (*M* = 87.8, *SD* = 8.2) than for preschoolers (*M* = 81.1, *SD* = 9.4). However, the effect size ($\eta^2 = .080$) was quite small. Based on these outcomes, all measures were collapsed across SES and gender in subsequent analyses. Grade was considered separately on analyses involving the NRT. Examination of articulation skills using the AAPS revealed no speech deficits for the children who failed the screening (*M* = 94.6, *SD* = 4.4) or

TABLE 5. Statistical results of performance on each assessment measure when examined for the influence of gender and SES using a multivariate ANOVA.

		df	F	η	p
KNVS	Gender	1	.016	.000	.899
	SES	1	.125	.002	.724
Csyn	Gender	1	.002	.000	.965
	SES	1	.026	.000	.871
MLCU	Gender	1	.553	.007	.459
	SES	1	1.07	.014	.304
REVS	Gender	1	1.86	.023	.176
	SES	1	2.49	.031	.119
NDW	Gender	1	1.31	.016	.256
	SES	1	.063	.001	.803
NRT	Gender	1	.053	.001	.818
	SES	1	1.41	.018	.238

Note. KNVS = Kaufman Nonverbal Scale; Csyn = complex syntax; MLCU = mean length of C-unit; REVS = reversible sentences; NDW = number of different words; NRT = nonword repetition task.

those who passed (*M* = 94.4, *SD* = 5.4). As this measure was simply confirmatory, the AAPS was not included in any further analyses.

Group performances. Table 6 displays the means and standard deviations for the children in the pass and fail groups as identified based on the screening in Study 1. Each measure was examined for significant differences between these two groups. Using a more conservative, adjusted alpha level of .01 (.05/5 measures) significant differences in performance were evident by group on three measures: KNVS, NDW, and NRT total score. These data confirmed that as a group the students who failed the screening did indeed have language skills that were different from the larger group of children who passed the screening, but not all assessment measures were equally useful for making this differentiation. Effect sizes were small for all variables ranging from $\eta^2 = .055$ to $\eta^2 = .145$. It was notable also that although all children scored in the normal range cognitively, the difference in the mean score on the KNVS was statistically significant, with a mean difference of nearly 10 standard score points lower attained by the group that failed the screening in Study 1 compared to the group that passed.

Discriminant function analyses. Scores on the cognitive assessment and all language measures were converted to *z* scores to allow comparison using a common, standardized scale. The *z* scores were entered into a series of discriminant function analyses to permit examination of student performance on the assessment compared to the screening. Discriminant analyses allow prediction of group membership based on linear combinations of the predictor variable, from a sample of cases for which group membership is known (pass/fail on screenings). A discriminant function is generated that can be applied to new cases for the predictor variables but unknown group membership (SPSS, 1999). The first discriminant analysis predicted group membership based on differences in the mean scores, using actual group membership (pass versus fail status on the screening) as the independent variable and the language and cognitive scores as the dependent variables.

TABLE 6. Means, standard deviations, minimum and maximum scores, *F* values, effect sizes (η), and significance levels (*p*) on the language and cognitive assessment measures for children in the pass and fail groups, and for preschoolers (P) and kindergartners (K) on the NRT.

		M	*SD*	Minimum	Maximum	*F*	η	*p*
KNVS	fail	93.5	10.1	75	119	9.92	.114	.002*
	pass	102.1	12.7	75	132			
REVS	fail	9.9	4.51	3	17	5.09	.085	.026
	pass	12.1	3.8	5	20			
MLCU	fail	2.9	.597	1.82	3.88	5.72	.069	.020
	pass	3.31	.667	2.08	4.94			
NDW	fail	68.4	11.2	44	86	10.2	.118	.002*
	pass	77.3	13.4	48	111			
Csyn	fail	3.1	2.3	0	10	4.48	.055	.037
	pass	4.3	2.8	0	13			
NRT	fail							
	P*	78.01	10.2	55.1	93.8			
	K*	79.52	10.3	64.2	86.7	6.48	.145	.013*
	pass							
	P	82.5	8.8	58.1	93.8			
	K	90.8	5.1	82.6	96.9			

Note. KNVS = Kaufman Nonverbal Scale; REVS = reversible sentences; MLCU = mean length of C-unit; NDW = number of different words; Csyn = complex syntax; NRT = nonword repetition task.
*Significant at *p* ≤ .01.

Sensitivity and specificity. Results of this analysis indicated that approximately 82% of participants were correctly classified overall based upon their performance on the screenings. Specifically, approximately 93% of children (51/55) who passed the screening in Study 1 also passed this assessment, while 60% of children who failed the screening also failed the assessment. Like many screenings described in the extant literature, this screening was more successful at identifying students who did not have language impairments, and was less accurate when identifying children in the failed group. Thus, the instrument appeared to have high specificity and moderate sensitivity (see Table 7).

Influential variables. The second discriminant analysis used a stepwise discriminant function to examine the relative contribution of each dependent variable to discrimination of the groups. Using Statistical Package for the Social Sciences (SPSS), each variable was entered sequentially and then systematically added and removed from the analysis until only those variables that contributed significantly to discrimination of the two groups remained. The outcomes confirmed again that three variables contributed

most significantly to determination of group membership, as follows: KNVS [$F(1, 78) = 9.15$, $p = .003$; Wilks's $λ = .890$], NDW [$F(1, 78) = 8.44$, $p = .0005$; Wilks's $λ = .820$], and NRT [$F(1, 78) = 7.11$, $p = .0003$; Wilks's $λ = .781$]. Combined, these three instruments accounted for approximately 51% of the total variance in scores.

Clinical cutoff values. The discriminant function analyses represented application of a traditional statistical standard to the data for interpretive purposes. Participant performance was examined further based on clinical assessment standards. Similar to the methods used by Tomblin et al, cutoff values of –1 *SD* and –1.25 *SD* were applied to the data and examined for their potential influence on the classification of participants (see Table 4). Using the same gold standard adopted for the screenings in Study 1, students who scored –1 *SD* from the mean on 2 or more of the assessment measures were placed in the "fail" group, and in the second analysis students who scored more than –1.25 *SD* from the mean on 2 or more measures were considered to have failed the assessment. All of the children who failed both the screening and the assessment were identified as true positives using these clinical cutoff scores. Thus the sensitivity rate of the screening for identifying true positives was 100% when these cutoffs were used. These represent slightly better results than were obtained using the discriminant function analysis to identify these children.

False negatives. Four children were identified as false negatives (i.e., language impairments not detected by the screening) by the discriminant analysis. However, application of clinical cutoffs revealed that three of these four children exhibited borderline normal performance, placing them in the passing group. One child did indeed fail two of the assessment tasks even when clinical standards were employed, suggesting the presence of a language impairment not detected by the screenings. Most important,

TABLE 7. Classification of participants using discriminant function analysis.

Screening results		Predicted group membership		Total
		Fail	Pass	
Count	Fail	15	10	25
	Pass	4	51	55
%	Fail	60.0	40.0	100.0
	Pass	7.3	92.7	100.0

Note. 82.3% of original grouped cases classified correctly.

application of a standard clinical cutoff score improved the specificity rate from 91% to 98%. Finally, two children who failed the screening in Study 1 and then were identified as false negatives using discriminant function analysis failed the assessment based on the –1 SD clinical standard. The use of these clinical standards improved the overall classification accuracy for the group from 82% to 88%.

False positives. A closer examination of the eight children who were misclassified as language impaired revealed significant differences by gender [$F(1, 23) = 5.77$, $p = .026$] and age [$F(1, 23) = 5.51$, $p = .029$] compared to the true positives. Effect sizes for gender ($\eta^2 = .22$) and for age ($\eta^2 = .21$) were small, suggesting some overlap in these two groups despite the statistically significant differences. The majority of the false positives ($n = 6$) were preschool boys, who were an average of 6 months younger ($M = 50$ months, $SD = 4.2$) than students identified as true positives ($M = 56$ months, $SD = 6.6$). Four of the children who were false positives were low income, and four were middle income, indicating no impact of SES on misclassification. Table 8 presents the screening and assessment data for true and false positives. Whereas the false positives performed quite poorly on the screenings, their assessment performance clearly demonstrates language skills within the normal range.

Discussion and General Conclusions

This investigation presents a screening protocol appropriate for use with preschool and kindergarten aged African American children. The screening protocol comprised traditional speech and language assessments that have been examined extensively and validated for use with this population in our earlier work (Craig & Washington, 2000, 2002; Washington & Craig, 1999; Washington et al., 1998). The utility of these instruments for screening purposes is an extension of this line of inquiry that is critical for early identification and subsequent intervention and/or prevention of language impairments in this population.

The screening successfully identified all children with true language deficits. The scores obtained by these children on the screening were compared to performance on a larger assessment battery for confirmation of their language impaired status. Research has demonstrated that screenings often are most successful for identifying children with more obvious language impairments but miss children with more subtle deficits. The children in this investigation who exhibited true language deficits performed at a level significantly below their normal language peers in most areas assessed and might be considered in this group of obviously impaired children. Less severe disabilities are often overlooked in young children and more likely to be identified later (Mercer, Algozzine, & Trifiletti, 1988). Accordingly, it is recommended that students who exhibit borderline performance like the false negatives in this investigation be targeted for follow-up in subsequent school years. It is possible that many of these students will demonstrate difficulty with language and with academic subjects that depend on strong language skills.

TABLE 8. Means, standard deviations, and minimum and maximum scores on the assessment tasks for true and false positives.

Group	Task	n	Minimum	Maximum	M	SD
True positives						
Screening	PPVT–III	17	62	103	78.38	9.79
	Wh-quest	17	11	56	41.47	10.23
	MLCUw-s3	17	3.6	10.3	6.476	2.012
	Triangles	17	6	12	8.65	1.90
Assessment	KNVS	17	78	104	92.71	8.05
	REVS	17	3	16	9.44	4.68
	MLCU	17	1.82	3.64	2.8400	.5731
	Csyn	17	0	6	2.59	1.87
	NDW	17	49	82	65.59	9.21
	NRT	16	55.1	86.7	76.544	9.821
False positives						
Screening	PPVT–III	8	77	89	82.37	3.89
	Wh-quest	8	19	52	40.13	10.99
	MLCUw-s3	8	2.6	9.3	5.725	2.301
	Triangles	8	5	19	10.00	4.11
Assessment	KNVS	8	75	119	95.38	14.22
	REVS	8	5	17	11.13	4.19
	MLCU	8	2.26	3.88	3.2800	.5683
	Csyn	8	1	10	4.38	2.92
	NDW	8	44	86	74.50	13.26
	NRT	8	63.2	93.8	81.700	10.046

Note. PPVT–III = Peabody Picture Vocabulary Test—Third Edition; KNVS = Kaufman Nonverbal Scale; REVS = reversible sentences; MLCU = mean length of C-unit; Csyn = complex syntax; NDW = number of different words; NRT = nonword repetition task.

Approximately one third (32%) of the children who failed the screening were later determined to have normal language skills. These false positive cases are illustrative of one of the major challenges facing professionals involved with screening young children: it is desirable for a screening to overidentify rather than underidentify children with disabilities so that all children who need intervention services will get them, but significant overidentification is undesirable because it results in overrepresentation of nondisabled students on special education caseloads. Overrepresentation is costly, both in fiscal and human terms. Children erroneously identified for services strain the financial and human resources of school systems. In urban school systems this can be particularly costly, as the numbers of high-risk and African American children are often greater than in smaller, more suburban districts and these children are more likely to be misidentified (U.S. Department of Education, 2002). The human cost to the children and their families of being identified for special education services unnecessarily is a major concern as well.

Screening can lead to misidentification of children for many reasons. Assessment with African American children has suffered from significant concern about the cultural bias of our instruments and subsequent validity of the results obtained. There are no such concerns about the protocol used in this investigation. It was developed specifically for use with African American children who were the age of the subjects in this investigation, and its reliability and validity for use with this population are well-established. This original investigation was focused on using a subset of the tasks composing our larger protocol for screening purposes. The outcomes are promising, yet certainly preliminary given the small sample of African American children who participated. In addition, only children whose nonverbal cognitive performance was not significantly below normal are included in this investigation. It is important to note, however, that none of the students screened exhibited cognition outside of the range identified by Tager-Flusberg and Cooper (1999) and thus none were excluded based on their performance on the cognitive measure.

The brevity of most screenings has been implicated as a significant variable in misclassifying children as well (Lenkarski, Singer, & Peters, 2001). The screening described in this article took approximately 20–25 min to administer. Screenings are designed to be brief, efficient snapshots of a student's skills. For many children, particularly at risk populations and those who are culturally different, this snapshot can lead to false impressions of the child's abilities, mistaking test performance for communicative competence.

It was notable that the eight students who were misidentified in this investigation were among the youngest children and that most of them (six) were boys. Overall, boys composed 44% (11/25) of the total sample of children who failed the screening. More than half of these boys (54%) were confirmed false positive cases by the larger assessment. Thus, despite the careful attention to development of a battery that is culturally fair for use with these children, this screening misclassified the same demographic group that is consistently overrepresented in special education: young, African American boys (U.S. Department of Education, 2002). It appears that for these young boys the validity of the instrument may be only one piece of the equation, albeit a critical piece, that must be considered in order to decrease their numbers in special education.

Although it is not possible to discuss causative relationships with such a small participant sample, the results of this preliminary look at screening outcomes suggest that there may be some critical interaction between the boys and the assessment context that we do not understand and that is beyond the explanatory scope of this investigation. These results will need to be examined with a larger group of children, but the outcomes of this investigation raise some important issues that warrant further consideration when evaluating this population. Notably, perhaps the brevity of screenings affects more acutely the performance of these boys in evaluation contexts than it does other children. The screening context involves performance on demand with an unfamiliar adult in a time frame that is quite brief by design. This may be a context that is especially difficult for these children. When given the opportunity to perform in a larger assessment context that involved approximately 1–1.5 hr of interaction with the examiner, these boys demonstrated language abilities well within normal expectations. These outcomes highlight the value of more extensive testing for these children before considering placement in special education. These children who present as false positive cases may also be a group for whom follow-up observation or assessment is warranted, as they may also represent a high-risk group linguistically.

The discriminant function analyses employed in this investigation were very useful as a first-level analysis and for evaluating the overall accuracy of the screening protocol and the performance of the assessment protocol. Number of different words, the Kaufman Nonverbal Scale, and the nonword repetition task contributed most significantly to the variance for assessment performance. This combination of expressive language, nonverbal cognitive, and processing skills proved to be most informative for identifying true negative performance in this population, and is therefore recommended for use in screening protocols used with this population. Further, it seems important to note that for both the screening and the assessment a reliable combination of standardized and nonstandardized instruments were very effective for accurate discrimination of African American children with language impairments and those whose language was developing normally. The improved sensitivity and specificity that resulted from application of a clinical standard to the data in addition to the statistical standard were informative for confirming the effectiveness of the screening for detection of true language impaired cases.

This protocol was designed to assess developmental speech and language skills that are *noncontrastive* with SAE (Seymour et al., 1998). That is, language structures were chosen because they are seemingly unaffected by the child's dialect status. This approach is widely used (Campbell et al., 1997; Craig & Washington, 1994, 2000; Stockman, 1997) and has allowed researchers to expand

significantly our understanding of language use with this population. It has also enabled us to develop culturally fair assessment approaches in the absence of in-depth knowledge of AAE. Whereas this approach has been important for moving us forward, in order to fully understand the African American child as a language user, future work will need to consider the *contrastive* features of the language system as well.

This screening protocol is brief, effective, and accurate for identifying African American children with language impairments. Like all screenings it overidentifies children and should be followed by additional, more extensive testing for children who fail. This study did not include a procedure for obtaining information from the teacher or parent regarding any concerns about the child's language use or level of functioning. In the case of African American boys, including this kind of information might be informative for assisting the speech-language pathologist with interpretation of disparate findings between the screening and assessment outcomes. Despite a range of reporting accuracy, it is generally agreed that parent reporting in combination with direct screening results are invaluable for identifying children who may have language difficulty (Westerlund & Sundelin, 2000). Attention to parent reporting instruments as important additions to the screening process would be an informative line of inquiry for future research in this area. Screenings represent an important method for identification of African American children with language impairments.

Acknowledgments

References

Camilleri, B., & Law, J. (2001). Screening for speech and language impairment: A follow-up of true negatives and false positives. *International Journal of Language and Communication Disorders*, *36*(Suppl.), 493–498.

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

Catts, H. W. (1993). The relationship between speech-language impairments and reading disabilities. *Journal of Speech and Hearing Research, 36*, 948–958.

Catts, H., Fey, M., Tomblin, B., & Zhang, X. (1997, November). *Semantic-syntactic deficits in poor readers*. Poster presented at the annual convention of the American Speech-Language-Hearing Association, Boston, MA.

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. *Scientific Studies of Reading*, *3*, 331–361.

Chamberlain, C., & Madeiros-Landurand, P. (1991). Practical considerations for assessment of LEP students with special needs. In E. V. Hamayan, & J. S. Damico (Eds.), *Limiting bias in the assessment of bilingual students* (pp. 111–156). Austin, TX: Pro-Ed.

Cochrane, A. L., & Holland, W. W. (1971). Validation of screening procedures. *British Medical Bulletin*, *27*, 3–8.

Cole, P. A., & Taylor, O. L. (1990). Performance of working-class African-American children on three tests of articulation. *Language, Speech, and Hearing Services in Schools*, *21*, 171–176.

Craig, H. K., & Washington, J. A. (1994). The complex syntax skills of poor, urban, African-American preschoolers at school entry. *Language, Speech, and Hearing Services in Schools*, *25*, 181–190.

Craig, H. K., & Washington, J. A. (2000). An assessment battery for identifying language impairments in African American children. *Journal of Speech, Language, and Hearing Research*, *43*, 366–379.

Craig, H. K., & Washington, J. A. (2002). Oral language expectations for African American preschoolers and kindergartners. *American Journal of Speech-Language Pathology*, *11*, 59–70.

Craig, H. K., Washington, J. A., & Thompson-Porter, C. (1998a). Average C-unit lengths in the discourse of African American children from low-income, urban homes. *Journal of Speech, Language, and Hearing Research*, *41*, 433–444.

Craig, H. K., Washington, J. A., & Thompson-Porter, C. (1998b). Performances of young African American children on two comprehension tasks. *Journal of Speech, Language, and Hearing Research*, *41*, 445–457.

Dollaghan, C., & Campbell, T. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, *41*, 1136–1146.

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

Gerken, K. (1979). The ability of listeners to report oral responses of Black and White children. *Language, Speech, and Hearing Services in Schools*, *10*, 35–46.

Hollingshead, A. B. (1975). *Four factor index of social status*. New Haven, CT: Yale University, Department of Sociology.

Kaufman, A., & Kaufman, N. (1983). *Kaufman Assessment Battery for Children: Interpretive manual*. Circle Pines, MN: American Guidance Service.

Klee, T., Pearce, K., & Carson, D. K. (2000). Improving the positive predictive value of screening for developmental language disorder. *Journal of Speech, Language, and Hearing Research*, *43*, 821–833.

Lampley, D. A., & Rust, J. O. (1986). Validation of the Kaufman Assessment Battery for children with a sample of preschool children. *Psychology in the Schools*, *23*, 131–137.

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (2000). The feasibility of universal screening for primary speech and language delay: Findings from a systematic review of the literature. *Developmental Medicine and Child Neurology*, *42*, 190–200.

Lenkarski, S., Singer, M., & Peters, M. (2001). Utility of the early screening profiles in identifying preschoolers at risk for cognitive delays. *Psychology in the Schools*, *38*, 17–24.

Mercer, C. D., Algozzine, B., & Trifiletti, J. (1988). Early identification: An analysis of the research. *Leaning Disability Quarterly*, *11*, 176–188.

Oetting, J. B., & McDonald, J. L. (2001). Nonmainstream dialect use and specific language impairment. *Journal of Speech, Language, and Hearing Research*, *44*, 207–223.

Peña, E. D., & Quinn, R. (1997). Task familiarity: Effects on the test performance of Puerto Rican and African American

children. *Language, Speech, and Hearing Services in Schools*, *28*, 323–332.

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

Seymour, H., & Bland, L. (1991). A minority perspective in the diagnosis of child language disorders. *Clinics in Communication Disorders*, *1*(1), 39–50.

Seymour, H. N., Bland-Stewart, L., & Green, L. J. (1998). Difference versus deficit in child African American English. *Language, Speech, and Hearing Services in Schools*, *29*, 96–108.

Seymour, H. N., Roeper, T. W., & deVilliers, J. (2003). *Diagnostic Evaluation of Language Variation*. San Antonio, TX: Psychological Corporation.

Shankweiler, D., Crain, S., Katz, L., Fowler, A. E., Liberman, A. M., & Brady, S. A. (1995). Cognitive profiles of reading disabled children: Comparison of language skills in phonology, morphology, and syntax. *Psychological Science*, *6*, 149–156.

SPSS. (1999). *SPSS Base 10.0 for Windows user's guide*. Chicago: SPSS.

Stockman, I. J. (1997). The promises and pitfalls of language sample analysis as an assessment tool for linguistic minority children. *Language, Speech, and Hearing Services in Schools*, *27*, 355–366.

Stott, C. M., Merricks, M. J., Bolton, P. F., & Goodyer, I. M. (2002). Screening for speech and language disorders: The reliability, validity and accuracy of the General Language Screen. *International Journal of Language and Communication Disorders*, *37*, 133–151.

Tager-Flusberg, H., & Cooper, J. (1999). Defining the phenotype of SLI. *Journal of Speech, Language, and Hearing Research*, *42*, 1275–1278.

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

Tomblin, J. B., Records, N. L., & Zhang, X. (1996). A system for the diagnosis of specific language impairment in kindergarten children. *Journal of Speech and Hearing Research*, *39*, 1284–1294.

U.S. Department of Education. (2002). *Twenty-third annual report to Congress on the implementation of the Individuals with Disabilities Education Act*. Washington, DC.

Vaughn-Cooke, F. B. (1986). The challenge of assessing the language of nonmainstream speakers. In O. L. Taylor (Ed.), *Treatment of communication disorders in culturally and linguistically diverse populations* (pp. 23–48). Boston, MA: College-Hill Press.

Washington, J. (1996). Issues in assessing language skills in African American children. In A. G. Kamhi, K. E. Pollock, & J. L. Harris (Eds.), *Communication development and disorders in African American children: Research, assessment, and intervention* (pp. 35–54). Baltimore: Brookes.

Washington, J., & Craig, H. (1998). Socioeconomic status and gender influences on children's dialectal variations. *Journal of Speech and Hearing Research*, *41*, 618–626.

Washington, J. A., & Craig, H. K. (1992a). Articulation test performances of low-income, African-American preschoolers with communication impairments. *Language, Speech and Hearing Services in Schools*, *23*, 203–207.

Washington, J. A., & Craig, H. K. (1992b). Performances of low-income, African-American preschool and kindergarten children on the Peabody Picture Vocabulary Test-Revised. *Language, Speech, and Hearing Services in Schools*, *23*, 329–333.

Washington, J. A., & Craig, H. K. (1999). Performances of at-risk, African American preschoolers on the Peabody Picture Vocabulary Test-III. *Language, Speech, and Hearing Services in the Schools*, *30*, 75–82.

Washington, J. A., Craig, H. K., & Kushmaul, A. J. (1998). Variable use of African American English across two language sampling contexts. *Journal of Speech, Language, and Hearing Research*, *41*, 1115–1124.

Westerlund, M., & Sundelin, C. (2000). Screening for developmental language disability in 3-year-old children: Experiences from a field study in a Swedish municipality. *Child: Care, Health, and Development*, *26*, 91–110.

Willson, V. L., Nolan, R. F., Reynolds, C. R., & Kamphaus, R. W. (1989). Race and gender effects on item functioning on the Kaufman Assessment Battery for Children. *Journal of School Psychology*, *27*, 289–296.

Contact author: Julie A. Washington, PhD, University Center for Development of Language and Literacy, University of Michigan, 1111 E. Catherine St., Ann Arbor, MI 48109-2054. E-mail: washja@umich.edu