LSHSS

Performance of African American Preschool and Kindergarten Students on the Expressive Vocabulary Test

Shurita Thomas-Tate
Florida State University, Tallahassee
Julie Washington
Wayne State University, Detroit, MI
Holly Craig
University of Michigan, Ann Arbor
Mary Packard
Wayne State University, Detroit, MI

Assessing the language of African American children is challenging because many standardized language measures are culturally biased (Stockman, 2000; Washington & Craig, 1992, 1999). It is important to examine available tests and determine which instruments are appropriate for use with this population. The primary purpose of this study was to examine the validity of the Expressive Vocabulary Test (EVT; Williams, 1997) for assessing the expressive vocabulary skills of African American students.

ABSTRACT: **Purpose:** To examine the validity of the Expressive Vocabulary Test (EVT; K. Williams, 1997) for assessing the expressive vocabulary skills of African American students.
Method/Results: One hundred sixty-five African American preschool and kindergarten students were administered the EVT. The mean EVT score for these African American students was 96.44 ($SD = 11.42$), which is not appreciably lower than the standardized mean of 100 ($SD = 15$).
Clinical Implications: Scores were normally distributed, indicating that the EVT is culturally fair and appropriate for use with some African American preschool and kindergarten children as part of an early screening battery. The importance of culturally fair vocabulary measures is discussed relative to this population.

KEY WORDS: vocabulary, African American students, assessment, validity

Historically, there has been difficulty in assessing vocabulary knowledge for culturally and linguistically diverse populations. This task is complicated by the fact that vocabulary knowledge and use are culturally bound, reflecting the specific knowledge and use of a community (Anderson & Battle, 1993; Champion, Hyter, McCabe, & Bland-Stewart, 2003; Mount-Weitz, 1996). The complexity associated with this assessment has been attributed in large part to the absence of instruments that are considered culturally fair. Even the Peabody Picture Vocabulary Test (PPVT; Dunn, 1959), one of the most popular and commonly used vocabulary tests, has a history of bias when used with African American students (Williams & Wang, 1997). In the past, African American children were shown to perform significantly lower than Caucasian children on the original PPVT (Kresheck & Nicolosi, 1973) and on the PPVT—Revised (PPVT-R; Dunn & Dunn, 1981) (Washington & Craig, 1992). One of the major criticisms of these earlier versions of the PPVT was that the ethnic minority composition of the normative sample was not representative of the U.S. population (Kresheck & Nicolosi, 1973).

Despite changes to the latest version of the PPVT, the PPVT-III (Dunn & Dunn, 1997; an increase in the percentage of persons from diverse backgrounds in the normative sample and an increase in the number of types of words sampled), there is disagreement among researchers as to whether the PPVT-III is unbiased and appropriate to use with African American students. Washington and Craig (1999) examined the PPPVT-III for use with 59 at-risk preschool-aged children from the metropolitan Detroit area. The mean performance for this group of students on the PPVT-III was 92, with a standard deviation of 11. Additionally, the students' scores were normally

0161-1461/06/3702-0143

distributed and not significantly different from the distribution of the normative sample. These researchers took these findings to suggest the PPVT–III to be unbiased and appropriate for use with African American preschoolers and kindergartners. In contrast, Champion et al. (2003) examined the PPVT–III for use with 49 typically developing preschool-aged children from impoverished families in the Tampa Bay area. The mean performance for this group of students on the PPVT–III was 86.84, with a standard deviation of 10.96. These students' scores were disproportionately lower than the normative sample and negatively skewed. Champion and colleagues found that the PPVT–III disproportionately assigned low scores to African American children from low-income backgrounds. They proposed that the differences in their results versus those of Washington and Craig may be due to income differences between the two study samples, with the students from the Washington and Craig study not being exclusively from low-income families.

Standardized tests continue to be widely used with African American and other groups of students in spite of the historical problems associated with their use with culturally and linguistically different populations. The ease and brevity of administration and interpretation makes these tests popular with clinicians who have large caseloads and limited time. Given their continued use, information regarding their validity is especially needed.

The EVT was chosen for investigation because it is conormed with the PPVT–III (Dunn & Dunn, 1997) and is widely used as well. This conorming allows for direct comparisons between expressive and receptive vocabulary skills, and these instruments are often paired in assessment to obtain a comprehensive picture of vocabulary knowledge. The EVT is an individually administered, norm-referenced test of expressive vocabulary. The EVT measures vocabulary knowledge (the number of words a person knows) as well as word retrieval (when paired with the PPVT–III). At the beginning level of the EVT, examinees are shown individual colored pictures and are asked to label them. At the advanced level, examinees are instructed to provide one-word synonyms of labeled pictures. The examiner's manual suggests that the EVT may be used for screening expressive language problems, screening preschool children, measuring word retrieval (in conjunction with the PPVT–III), understanding reading difficulties, and monitoring growth.

The EVT was standardized on a representative sample of 2,725 examinees aged 2.5 to 90 years, across four U.S. regions. The standardization sample included 49.4% females and 50.6% males. Socioeconomic status was assessed based on the education level of the examinee's parents (if the examinee was less than 24 years old) or the education of the examinee (if the examinee was older than 24 years). Seventeen percent of the population had less than 12 years of education, 31% were high school graduates, 31% had 1 to 3 years of college or technical school, and 20% had 4 or more years of college. EVT sample distribution by race or ethnicity was 18.1% African American, 12.9% Hispanic, 64.3% White, and 4.6% other. These demographic variables have distributions closely matching those of the U.S. population, and in the case of African American participants, oversampling is evident. Additionally, students receiving various special education services were represented in the norm sample in approximately the same proportion that occurs in the U.S. school population, which included 2.3% students with speech impairments and 5.5% with learning disabilities.

It is important to assess the validity of test measures with children at school entry. It is at these early grades that misclassification starts and assessment instruments fail to correctly identify at-risk students. Stockman (2000) suggested assessing bias as a component of validity, stating that "aggregate scores of the ethnic group as a whole can mask within group differences in social class, nationality and so on" (p. 350). For example, in the Washington and Craig (1999) examination of the PPVT–III, although the aggregate scores for at-risk African American preschoolers were normally distributed, additional analysis showed significant correlations between test scores and maternal education. Other researchers have noted significant differences in performance on tests within groups by social class (Champion et al., 2003; Donahue, Daane, & Grigg, 2003; Qi, Kaiser, Milan, Yzquierdo, & Hancock, 2003; Thomas-Tate & Edwards, 2001; Washington & Craig, 1999), gender (Donahue et al., 2003; Hyde & Linn, 1988; Stevenson, Chen, & Uttal, 1990), and community (Donahue et al., 2003). These variables appear to be important factors to assess. Accordingly, this article proposes to answer the following question: Is the EVT valid for use in assessing vocabulary knowledge with African American preschool and kindergarten students?

METHOD

Participants

The participants were 165 African American preschool and kindergarten students. There were 81 boys and 84 girls. Students' ages ranged from 3;2 (years;months) to 5;11, with a mean age of 4.23 (SD = .62). African American students were recruited from several schools in two communities for participation in a larger study examining the prevalence of speech and language disorders among African American preschoolers and kindergartners. Data for the prevalence study were collected over several years. Data for the current participants were obtained during one data collection year (2002–2003); a total of 166 students participated during that year. All students returning consent forms were included in the prevalence study and were administered a language battery that included the EVT (the EVT was added to the battery in 2002). One student was eliminated from the current investigation due to missing demographic data. Final analysis was performed on 165 students.

The participants were from two Michigan communities—a mid-sized central city and an urban-fringe city. These two communities represented significantly different demographics. The mid-sized central community was a college town with a large number of affluent and highly educated residents; 65.1% of adults had at least a bachelor's degree (SchoolMatters, 2005). Overall, the percentage of students scoring proficient or above average on the state assessment in reading was 84.1%. Fifteen percent of the students in this community were African American, and nearly 19% received free or reduced-price lunch as part of the federally funded lunch program. The students in the urban-fringe community were largely African American (90%), and close to half were low income, with 47.5% receiving free or reduced-price lunch. Only 25% of adults in this community had at least a bachelor's degree. The percentage of students scoring proficient or above average on the state assessment in reading was 58.1% (SchoolMatters, 2005).

Although the overall profiles of the communities differed significantly, the African American students within these communities were very similar. African American students from both

communities performed below average on the state-administered proficiency examination (SchoolMatters, 2005). African American students from both communities also were disproportionately poor and low-academic achievers. Table 1 provides a summary of participant characteristics by district.

In addition to overall community differences, there were also curriculum differences in school programs for the school districts within these two communities. Preschoolers in both districts were enrolled in public school classrooms that were part of either Head Start or the Michigan School Readiness Program (MSRP). Head Start is a federally funded program for which the children must be low income in order to participate; MSRP is designated for students who are at risk of academic failure, which includes children who are learning English as a second language. The academic curriculum is more constrained in Head Start in that there are more social/health-related requirements (e.g., tooth brushing, hand washing, meal eating). MSRP classrooms devote less time to social/health skills and thus have more time available for academic skills (e.g., color, shapes, letters, and numbers). However, no important differences were noted by the researchers in overall curriculums. Head Start teachers in the mid-sized central community's school district had bachelor's degrees and/or an early childhood endorsement, and MSRP teachers were required to have at least a master's degree and an early childhood endorsement. District policy in the urban-fringe community required that all teachers be certified; thus, all of their teachers (Head Start and MSRP) had at least a bachelor's degree (some have a master's degree), and all had an early childhood endorsement.

Kindergartners in this study were all enrolled in district-run programs. No important differences in curriculum were noted between the school districts within each community. School programs had similar foci for kindergartners, as dictated by the state of Michigan kindergarten standards. Some of the mid-sized central-city students were in full-day programs. The level of participation for students (full-day vs. half-day) was not monitored in this study.

Sixty-nine students were from homes of low socioeconomic status (SES) and 96 were from mid-SES homes. SES was determined using the Hollingshead four factor index of social status (Hollingshead, 1975). Hollingshead scores were used to assign students to social status groups. Hollingshead scores range from 1 to 5 and are obtained by a mathematic calculation that takes into account parent education level, occupation, gender, and marital status. For example, a high-school-educated single mother who works as a daycare provider receives 12 points for having a high school education and 10 points for occupation. The two point values are added together and the sum of these values determines the social strata score. In this case, a score of 22 corresponds with a social strata score of 4. Students whose parents have scores of 1 or 2 are classified as mid SES; those with scores of 3 to 5 are classified as low SES. The information used to determine Hollingshead SES scores was obtained from a brief case history questionnaire that was given to all parents at the time of consent. Parents were asked their marital status, the number of children and adults living in the home, and who contributed to the household income. Additionally, parents were asked the gender, percentage of money contributed to household, occupation, and education of the household financial contributors. The mean Hollingshead score for low-SES students was 3.81 (SD = .90); it was 2.22 (.49) for mid-SES students. The information needed to calculate Hollingshead scores was not available for 11 children. Classroom teachers provided information to determine SES for these students.

Students with atypical language skills were included in this study. As a part of the larger prevalence study, students were given a screening battery to assess language skills. The battery included the PPVT–III, the Triangles and Face Recognition subtests of the Kaufman Assessment Battery for Children (KABC; Kaufman & Kaufman, 1983), a Wh-questions comprehension task, and a picture description task (from which mean length of communication unit [MLCU] was calculated) (Washington & Craig, 2004). The EVT was given in addition to the screening battery to provide additional language information. Students with scores below age expectation on any two of the screening measures (excluding the EVT) failed the screening (Washington & Craig, 2004). Twenty-nine students (18%) failed the screening battery during the 2002–2003 school year using these criteria. (Four students were not classified as pass or fail because they did not complete the screening battery.) Mean scores for students by pass/fail status from the screening battery (including the EVT) are shown in Table 2. Students who failed to complete the screening (and who were not classified as pass/fail) were included in subsequent analyses. See Washington and Craig (2004) for a more complete description of the screening battery and subsequent outcomes.

Data Collection and Scoring

The EVT was individually administered to students at their respective schools in common areas that were relatively free of

Table 1. Summary of participant characteristics.

	Mid-sized central		*Urban-fringe*		
	Preschool	*Kindergarten*	*Preschool*	*Kindergarten*	*Total*
SES					
Low	23	5	31	10	69
Mid	18	11	43	24	96
Gender					
Male	23	7	32	19	81
Female	18	9	42	15	84
Total	41	16	74	34	165

Note. SES = socioeconomic status.

Table 2. Mean scores for students by pass/fail status from the screening battery (including the Expressive Vocabulary Test [EVT; Williams, 1997]).

	Pass[a] (n = 132)		*Fail*[a] (n = 29)			
	M	SD	M	SD	t *test* (df)	*Effect size* η^2
KABC Triangles*	9.99	2.49	8.63	1.80	$t(149) = 2.69$	.046
KABC Face Recognition	10.20	2.87	9.27	2.66	$t(70) = 1.29$	.023
MLCU*	4.47	1.32	3.57	1.20	$t(158) = 4.24$	.012
Wh-Questions*	51.37	7.36	38.72	8.13	$t(158) = 8.22$	.299
PPVT-III*	98.26	8.67	80.71	10.57	$t(158) = 9.34$	.356
EVT*	98.52	10.65	86.66	9.62	$t(159) = 5.52$	.161

Note. KABC = Kaufman Assessment Battery for Children (Kaufman & Kaufman, 1983); MLCU = mean length of communication unit; PPVT–III = Peabody Picture Vocabulary Test—Third Edition.
[a]Four students were not classified as pass or fail because they did not finish the screening battery.
*$p < .01$.

distractions. Rooms used for testing were generally non-occupied therapy rooms and offices (e.g., speech-pathology, school psychology, nurse). Whenever possible, screenings were conducted in quiet spaces. However, space was severely limited in some schools, so some screenings were conducted in less than ideal locations.

The EVT was administered according to published guidelines by female examiners who had experience working with young children and who were trained to administer the tests in the protocol. There were seven examiners, three who were African American and four who were Caucasian. A *t* test for independent samples revealed no significant variations in EVT standard scores across subjects relative to race of examiner, $t(69) = -.84$, $p = .41$.

Participants' responses were scored according to the published criteria. Raw scores were converted to standard scores and percentile ranks. Interrater agreement was established by having an independent scorer recalculate scores for a randomly selected subset of the tests (10% for each test). Scoring agreement for the EVT was 100%. Scoring agreement for the other standardized tests (KABC Triangles and Face Recognition and PPVT–III) was also 100%.

RESULTS

The mean standard score on the EVT for these 165 African American preschool and kindergarten students was 96.42, with a standard deviation of 11.45. Scores ranged from 62 to 126—a wide range spanning from approximately −2 *SD* below the mean to +2 *SD* above the mean, suggesting sufficient performance spread on this instrument.

Although the means and standard deviations for this group of students represent scores below the standardized mean of 100 and the standard deviation of 15, they were still well within the normal expectations for this instrument (within 1 *SD*). Additionally, application of the One Sample Kolmogorov-Smirnov test of normality demonstrated a normal distribution of the scores, *K-S*(165) = .674, $p = .754$, which can also been seen visually in Figure 1.

In a normal distribution, it is expected that approximately 68% of the scores will fall between −1 and +1 *SD* of the mean (34% above and 34% below). Seventy-seven percent of student scores (127 out of 165) from this population were between −1 and +1 *SD* of the mean (28.5% above and 48.5% below). Also in a normal distribution, approximately 14% of scores are expected to fall between −1 and −2 *SD* as well as between +1 and +2 *SD*. More than 13% of students (22 out of 165) in this study had scores that were between −1 and −2 *SD*, and 8.5% (14 out of 165) had scores that were between +1 and +2 *SD* (see Table 3).

The relationship between EVT scores and selected demographic variables was examined for the group of 165 students (see Table 4 means and standard deviations). A univariate between-subjects analysis of variance (ANOVA) showed no significant effects of community, $F(1, 164) = .48$, $p = .49$; grade, $F(1, 164) = 2.01$, $p = .16$; or SES, $F(1, 164) = 1.86$, $p = .18$ on the EVT. There was, however, a significant effect of gender, $F(1, 164) = 5.84$, $p = .02$, with girls performing significantly better than boys (see Table 5 for ANOVA). However, the variance accounting for effect sizes associated with this difference was negligible ($\eta_p^2 = .04$).

Figure 1. Histogram of EVT scores with normal curve overlay.

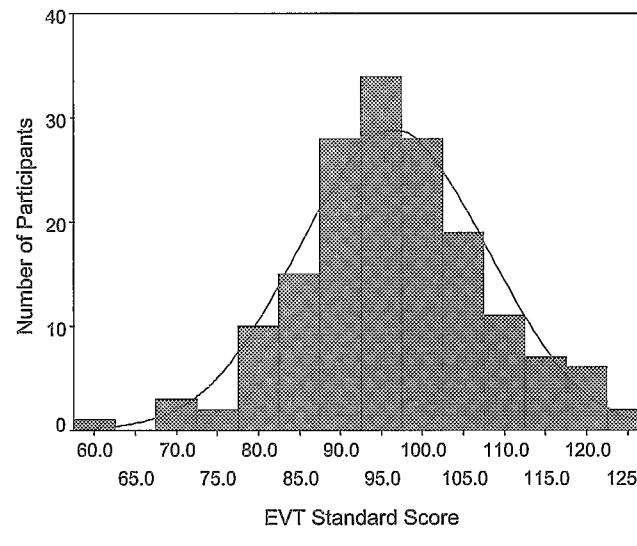

Table 3. Normal curve distributions of mean EVT standard scores.

Normal curve distribution	*Expected % distribution under the curve*	*Number of scores (%)*	*Range*
+2 *SD* and greater	2	0 (0)	131–145
+1 *SD* to +2 *SD*	14	14 (8.5)	115–130
0 to +1 *SD*	34	47 (28.5)	100–114
0 to −1 *SD*	34	80 (48.5)	85–99
−1 *SD* to −2 *SD*	14	22 (13.3)	70–84
−2 *SD* and fewer	2	2 (1.2)	69–55

Because gender was determined to be important for aggregate data, EVT scores were disaggregated and performance was analyzed by gender (see Table 6 for means and standard deviations). For girls, there was a significant effect of grade, $F(1, 83) = 4.01$, $p = .05$, $\eta_p^2 = .05$, with kindergarten girls ($n = 24$, $M = 102.54$, $SD = 13.39$) performing better than preschool girls ($n = 60$, $M = 96.68$, $SD = 11.01$). The variance accounting for effect sizes associated with this difference was negligible. There was not a significant effect for SES, $F(1, 83) = .87$, $p = .36$, or community, $F(1, 83) = .24$, $p = .63$. However, there was a significant interaction of SES by community, with a negligible effect size, $F(1, 83) = 3.82$, $p = .05$, $\eta_p^2 = .05$, with the mid-income central-city community girls having the highest mean scores ($M = 105.27$, $SD = 13.96$, $n = 15$), followed by the low-income urban fringe ($M = 99.50$, $SD = 6.72$, $n = 18$), mid-income urban fringe ($M = 97.23$, $SD = 12.31$, $n = 39$), and low-income central-city girls, respectively ($M = 91.67$, $SD = 11.20$, $n = 12$).

For boys, there were no significant effects of SES, $F(1, 80) = 1.01$, $p = .32$; grade, $F(1, 80) = .02$, $p = .89$; or district, $F(1, 80) = .24$, $p = .63$.

Twenty-nine of the 165 students (22 boys and 7 girls) participating in this investigation failed a language screening. EVT scores for the students failing the language screening ranged from 62 to 102, with a mean of 86.66 ($SD = 9.62$). An ANOVA revealed that the mean EVT scores for the 29 students who failed the screening were significantly different from the scores of the students who passed the screening: $F(1, 160) = 30.45$, $p = .001$, $\eta_p^2 = .16$. Of the 29 students who failed the language screening, only 4 of them also had failing EVT scores (more than 1.5 *SD* below the mean)—3 boys (1 kindergarten and 2 preschool) and 1 preschool girl. Relationships between EVT scores and selected demographic variables were examined for these 29 students. A univariate between-subjects ANOVA showed no significant effects of gender, $F(1, 28) = .118$, $p = .73$, $\eta_p^2 = .006$; school district, $F(1, 28) = .059$, $p = .81$, $\eta_p^2 = .003$; grade, $F(1, 28) = 1.63$, $p = .22$, $\eta_p^2 = .075$; or SES, $F(1, 28) = .001$, $p = .99$, $\eta_p^2 = .001$, on the EVT for the 29 students who failed the language screening.

The criterion validity of the EVT was examined by looking at the relationship between EVT and PPVT–III scores. There was a strong, positive, statistically significant correlation between the EVT and PPVT–III scores for the entire group of students ($r = .66$, $p = .01$), girls ($r = .69$, $p = .01$), boys ($r = .62$, $p = .01$), low-income students ($r = .56$, $p = .01$), mid-income students ($r = .72$, $p = .01$), the urban-fringe community ($r = .70$, $p = .01$), the central-city community ($r = .61$, $p = .01$), the preschool-aged students ($r = .61$, $p = .01$), and

Table 4. Means and standard deviations for the EVT and PPVT–III by selected demographics.

Demographic	*EVT* M	*EVT* SD	*PPVT–III* M	*PPVT–III* SD
Community				
Urban fringe ($n = 108$)	96.19	10.88	93.92	11.42
Central city ($n = 57$)	96.88	12.56	97.61	10.22
Grade				
Preschool ($n = 115$)	95.71	10.87	94.24	10.73
Kindergarten ($n = 50$)	98.06	12.64	97.36	11.81
SES				
Low income ($n = 69$)	94.38	9.87	94.49	10.94
Mid income ($n = 96$)	97.90	12.31	95.71	11.29
Gender				
Male ($n = 81$)	94.42	10.56*	93.76	11.50
Female ($n = 84$)	98.36	12.00*	96.57	10.65

*$p < .01$, $\eta_p^2 = .04$.

Table 5. Univariate between-subjects analysis of variance with EVT as the dependent variable.

Source	df	F	η_p^2	p
Grade	1	2.01	.01	.16
Gender	1	5.842	.04	.02*
Community	1	.478	.01	.49
Socioeconomic status (SES)	1	1.86	.01	.18
Grade * Gender	1	2.59	.02	.11
Grade * Community	1	1.39	.01	.24
Grade * SES	1	.945	.01	.33
Gender * District	1	.002	.01	.96
Gender * SES	1	.002	.01	.97
Community * SES	1	3.11	.02	.08
Grade * Gender * Community	1	.249	.01	.62
Grade * Gender * SES	1	.007	.01	.93
Grade * Community * SES	1	.111	.01	.74
Gender * Community * SES	1	1.42	.01	.24
Grade * Gender * Community * SES	1	.019	.01	.89
Error	149	(124.681)		
Total	165			

Table 6. EVT means and standard deviations by gender and selected demographics.

Demographic	*Girls* M	*Girls* SD	*Boys* M	*Boys* SD
Community				
Urban fringe	97.95	10.85	94.22	10.68
Central city	99.22	14.22	94.77	10.52
Grade				
Preschool	96.68	11.08	94.65	10.65
Kindergarten	102.54	13.19	93.92	10.55
SES				
Low income	96.37	9.45	92.85	10.03
Mid income	99.46	13.16	95.88	10.94

kindergartners ($r = .74$, $p = .01$), as well as for the students who passed the screening ($r = .60$, $p = .01$) and those who failed the screening ($r = .57$, $p = .01$).

DISCUSSION

The purpose of this study was to determine whether the EVT is a valid instrument for assessing the vocabulary skills of African American students. The mean EVT score for the 165 African American students in this study was 96.42 ($SD = 11.45$), which is not appreciably lower than the standardized mean of 100 ($SD = 15$). Additionally, the scores of these students were normally distributed and were moderately, positively correlated with receptive vocabulary scores obtained on the PPVT–III, another vocabulary measure that is already shown to be valid for use with a similar group of African American students (Washington & Craig, 1999). Additionally, statistical analyses revealed a normal distribution of EVT scores among subjects. These results suggest that the EVT is valid for assessing the vocabulary skills of this Midwestern group of African American preschoolers and kindergarteners. Even so, more detailed analyses were performed.

Statistical analysis demonstrated a significant main effect of gender on the EVT, with girls performing significantly better than boys. This finding is diminished by negligible effect sizes. A disproportionate number of male participants with language impairments may have influenced this outcome.

Twenty-nine of the 165 students participating in this investigation failed the language screening battery. Only 4 of these 29 students had EVT scores that were more than 1.5 *SD* below the mean (the criteria often used to indicate disorder). This suggests that most of the students failing the screening had EVT scores within the normal range. That is, the EVT did not distinguish students who failed a language screening battery from those who passed. Previous studies have found similar results when examining vocabulary measures for assessing language. In examining the diagnostic accuracy of four vocabulary tests including the EVT, Gray, Plante, Vance, and Henrichsen (1999) found low predictive validity for vocabulary measures, with many preschool children with language disorders scoring in the normal range on these tests, including the EVT. They concluded that the vocabulary tests examined were not singularly appropriate for identifying and screening preschool-age children for language disorders. They recommended that vocabulary tests such as the EVT not be used alone for screening or diagnosis but rather as part of a larger screening and/or assessment battery.

Criterion validity of the EVT can be assessed by examining whether a relationship is demonstrated between the scores of the test measure and some other measure of semantic knowledge. In this case, the EVT was compared with the PPVT–III. These two tests are both measures of vocabulary knowledge, although they assess vocabulary in different ways (expressive vs. receptive). Analyses of these scores showed a strong, positive, statistically significant correlation for this group of students, suggesting that the EVT accurately measures semantic knowledge (to the extent that the PPVT–III does).

Overall, these results indicate that the EVT is valid for use in assessing vocabulary skills with this group of African American children. However, these findings may not generalize to African American children in other geographic regions. The results of this study demonstrate that the EVT used alone may not identify children who may have language disorders. However, clinicians may find the EVT useful for measuring expressive vocabulary and/or as an appropriate component of a well-planned screening or assessment battery. Continued effort should be devoted to developing and assessing standardized language measures for use with culturally and linguistically diverse populations.

ACKNOWLEDGMENT

This project was supported by Grant 1 R01 DC 04273-04 from the National Institutes of Health, National Institute of Deafness and Other Communication Disorders. Special thanks to the children, administrators, and teachers of the Oak Park and Ann Arbor, Michigan public schools for their cooperation and participation in this investigation and to Connie Thompson, Stephanie Hensel, and Erin Quinn for assistance in preparation of this manuscript.

REFERENCES

Anderson, N. B., & Battle, D. E. (1993). Cultural diversity in the development of language. In D. E. Battle (Ed.), *Communication disorders in multicultural populations* (pp. 158–185). Boston: Andover

Champion, T. B., Hyter, Y. D., McCabe, A., & Bland-Stewart, L. M. (2003). "A matter of vocabulary": Performances of low-income African American Head Start children on the Peabody Picture Vocabulary Test—III. *Communication Disorders Quarterly, 24,* 121–128.

Donahue, P., Daane, M., & Grigg, W. (2003). *The nation's report card: Reading highlights 2003* (NCES 2004452). Washington, DC: U.S. Department of Education.

Dunn, L. (1959). *Peabody Picture Vocabulary Test.* Circle Pines, MN: American Guidance Service.

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

Dunn, L., & Dunn, L. (1981). *Peabody Picture Vocabulary Test—Revised.* Circle Pines, MN: American Guidance Service.

Gray, S., Plante, E., Vance, R., & Henrichsen, M. (1999). The diagnostic accuracy of four vocabulary tests administered to preschool-age children. *Language, Speech, and Hearing Services in Schools, 30,* 196–206.

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

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104,* 53–69.

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

Kresheck, J. D., & Nicolosi, L. (1973). A comparison of Black and White children's scores on the Peabody Picture Vocabulary Test. *Language, Speech, and Hearing Services in Schools, 4,* 37–40.

Mount-Weitz, J. (1996). Vocabulary development and disorders 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. 227–245). Baltimore: Brookes.

Qi, C. H., Kaiser, A. P., Milan, S. E., Yzquierdo, Z., & Hancock, T. B. (2003). The performance of low-income African American children on the Preschool Language Scale—3. *Journal of Speech, Language, and Hearing Research, 46,* 576–590.

SchoolMatters. (2005). Data retrieved March 29, 2004, from http://www.schoolmatters.com

Stevenson, H. W., Chen, C., & Uttal, D. H. (1990). Beliefs and achievement: A study of Black, White, and Hispanic children. *Child Development, 61,* 508–523.

Stockman, I. J. (2000). The new Peabody Picture Vocabulary Test—III: An illusion of unbiased assessment? *Language, Speech, and Hearing Services in Schools, 31,* 340–353.

Thomas-Tate, S., & Edwards, J. A. (2001, November). *Influence of SES and ethnicity on measures of reading readiness.* Poster session presented at the annual meeting of the American Speech-Language-Hearing Association, New Orleans, LA

Washington, J. A., & Craig, H. K. (1992). Performance 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). Performance of at-risk, African American preschoolers on the Peabody Picture Vocabulary Test—III. *Language, Speech, and Hearing Services in Schools, 30,* 75–82.

Washington, J. A., & Craig, H. K. (2004). A language screening protocol for use with young African American children in urban settings. *American Journal of Speech-Language Pathology, 13,* 329–340.

Williams, K. (1997). *Expressive Vocabulary Test.* Circle Pines, MN: American Guidance Service.

Williams, K., & Wang, Y. (1997). *Technical manual: Peabody Picture Vocabulary Test—III.* Circle Pines, MN: American Guidance Service.

Received March 11, 2004
Accepted November 28, 2005
DOI: 10.1044/0161-1461(2006/016)

Contact author: Shurita Thomas-Tate, PhD, Florida State University, Department of Communication Disorders, 307 Regional Rehabilitation Center, Tallahassee, FL 32306. E-mail: sthomast@fsu.edu