Performances of Low-Income, African American Preschool and Kindergarten Children on the Peabody Picture Vocabulary Test-Revised

Julie A. Washington
Holly K. Craig
University of Michigan

This study examined test performances of 105 low-income, urban, African American preschool and kindergarten boys and girls on the Peabody Picture Vocabulary Test-Revised (PPVT-R). Most children performed more than one standard deviation below the mean, and a scoring adjustment to the test failed to change this distribution substantially. The findings indicate that the PPVT-R is not appropriate for use with this population.

KEY WORDS: low income, African American, preschooler, kindergartner, Peabody Picture Vocabulary Test-Revised

The Peabody Picture Vocabulary Test (PPVT) is a long-standing and widely used diagnostic instrument. Although the original edition of the PPVT (Dunn, 1959) fared well in terms of reliability and construct validity (Weiner, 1979), it was criticized as racially biased (Adler & Birdsong, 1983; Kresheck & Nicolosi, 1973; Wolfram, 1983). The original edition was normed on 4,012 white children in the Nashville, Tennessee, area. Investigations comparing the performance of African American and white children on the PPVT indicated that African American children score significantly lower than their White peers (Adler & Birdsong, 1983; Kresheck & Nicolosi, 1973).

The Peabody Picture Vocabulary Test-Revised (PPVT-R; Dunn & Dunn, 1981) reflected a number of positive changes from the original edition, including standardization conducted on a national basis and inclusion of subjects representing various ethnic groups and socioeconomic levels. Ten percent of the individuals in the normative sample were African American, and 7% were from low-income homes, as indexed by occupations in the household. Although the number of African Americans and low-income individuals selected to participate in the standardization sample appears small, as noted by Vaughn-Cooke (1986), these numbers reflect the percentage of African Americans in the general population as indexed by U. S. census data at the time of the test's construction (Dunn & Dunn, 1981).

By implication, the PPVT-R should be appropriate for use as part of a testing protocol for African American children from low-income families. However, rather than directly addressing the issue of the appropriateness of the PPVT-R for use with low-income and African American children who speak BE, the test developers state that "the test is designed for use with persons ... [who] understand Standard English to some degree." (See page *x*). This, then, would include African American children who speak BE. Unfortunately, the PPVT-R reports lower test reliability for minority children than for other populations (p. 58). Because it is the most widely used vocabulary test for children (Weiner, 1979; Saville-Troike, 1986), it seems important to confirm that the PPVT-R is inappropriate for use with this population.

Studies examining the performance of this population on standardized speech and language assessment instruments have found that low-income African American children who speak BE often are penalized for their dialectal differences (Taylor & Payne, 1983; Saville-Troike, 1986; Vaughn-Cooke, 1986). The test administration procedures, scoring systems, and test interpretations developed for use with most speech and language assessment instruments use a Standard English (SE) linguistic model, potentially jeopardizing their validity with children who speak different cultural dialects such as BE.

The need for valid test instruments with low-income African American children is critical. Examination of census data for the Detroit area, where the subjects in this investigation resided, revealed that from 1979 to 1989, the percentage of Black residents in the city increased from 58% to 75% and that Blacks living below the poverty line (i.e., less than $10,000 annual household income) increased from 33% to 38% (University of Michigan, Detroit Area Study, 1989). These statistics reflect a national trend toward an increase in the number of low-income

minority residents living in large urban centers (University of Michigan, Detroit Area Study, 1989). The purpose of this study was to examine the appropriateness of the PPVT-R as a test of single-word hearing vocabulary for a population of low-income, urban, African American children who were speakers of BE.

METHOD

Subjects

The subjects were 105 low-income, urban, African American preschool (*n* = 59) and kindergarten (*n* = 46) boys (*n* = 50) and girls (*n* = 55), ranging in age from 4:5 (years:months) to 6:3. All of the children were speakers of BE and attended schools in the Metropolitan Detroit area. The subject's use of BE was confirmed by the judgment of a certified speech-language pathologist who was a speaker of BE, using a 1/2-hour spontaneous language sample audiorecorded during a free play interaction. The preschool program was new in the public school system; thus the kindergartners were in their first year of school, as were the preschoolers. The preschool program stressed a basic skills curriculum, including social skill development. This curriculum was designed to provide social and academic skills judged to be deficient in the children's homes. All were functioning normally in the classroom according to teacher reports, and none of the children were receiving special education services or had been referred for special services.

The preschoolers were selected from a preschool program in which 396 4- and 5-year-old children were enrolled, 250 of whom were African American. The preschool program was designed for children considered to be "at risk" for educational achievement. At-risk status was determined based upon several criteria established by the school system, including low family income (less than $10,000), mother's age at child's birth, number of other children in the home, the child's health, and development history. At least two of the at-risk criteria had to be met for a child to be enrolled in the preschool program. None of the children participating in this study was defined as at-risk based on poor health or unusual development histories. The preschoolers were stratified by gender and then randomly selected. Each preschooler passed a hearing screening at 25 dB for 500, 1000, 2000, and 4000 Hz (American National Standards Institute, 1969), administered at the time of the data collections.

The kindergartners were selected randomly from five different elementary schools in the same public school system. They were part of a larger population of 300 low-income African American kindergarten children. They attended five different elementary schools and were selected by the school system for participation in this project. The kindergartners were determined to be low-income by the school system based upon their eligibility for participation in a free lunch program designed for low-income families. Each kindergartner had passed a bilateral hearing screening administered by the county prior to entering school.

Data Collection and Scoring

The PPVT-R (Dunn & Dunn, 1981) was administered to each child individually. It was administered and scored according to published guidelines in a room free from distractions at each school. The test was administered by a certified speech-language pathologist who was an African American female.

Each subject's responses were first scored according to the established scoring criteria for the PPVT-R, and a raw score and standard score were derived. A procedure for adjusting each child's raw scores was developed and then applied. Adjusting the scores of children who speak BE has been advocated by several investigators (e.g., Cole & Taylor, 1990, Taylor & Payne, 1983, Vaughn-Cooke, 1986) as an alternative to the development and standardization of completely new tests appropriate for children who speak BE. Cole and Taylor (1990) developed adjusted scores for articulation tests by determining the test items that might be affected by BE phonology and then crediting the child's total scores for these items. In the present study, test items that occurred prior to the final set of six errors identified by Dunn and Dunn (1981) as ceiling scores and that were missed by at least half of the children were identified. The children were given credit for these items, and the values were added to their overall raw scores.

Interjudge reliability was established by having an independent observer—a certified speech-language pathologist with past research experience—recalculate a randomly selected subset of the standard scores of 10% of the subjects. Reliability scores were 100%. An additional 10% of the subjects' raw scores were recalculated using the list of items identified for scoring adjustments. A point-to-point agreement for the resulting standard scores between the examiner and independent observer was also 100%.

RESULTS

The mean standard score equivalent obtained for the 105 children was 79.7. This score corresponds to approximately the 10th percentile. There were no statistically significant differences among children in the three age groups ($F[2, 102] = .29, p > .05$), nor between grades ($t[1] = 1.71, p > .05$) (see Table 1). Most children (91%) performed below the standard score mean of 100 established for the Dunn and Dunn (1981) sample. The mean for the present sample was more than one standard deviation below the mean obtained by Dunn and Dunn. If one standard deviation below the mean is used as a clinically significant cutoff, then 65% of the children in this sample appeared to have vocabulary problems based on their unadjusted standard scores. Figure 1 displays the standard score equivalents obtained by the children in

TABLE 1. Standard score means and standard deviations for each age group and for preschool (P) and kindergarten (K) grades.

	Ages*			Grades**		
	4	5	6	P	K	Combined
Mean	80.2	80.2	83.4	77.4	82.7	79.7
SD	19.1	13.5	9.1	17.3	13.4	15.9
n	35	55	15	59	46	105

*$F(2,102) = .29, p > .05$
**$t(1) = 1.71, p > .05$

the present study compared to the norms established by Dunn and Dunn (1981).

Table 2 lists the 16 test items missed by at least 50% of the children in the sample. The kindergartners and preschoolers both missed many items on the PPVT-R. However, the kindergartners did not get as many of the same items wrong ($n = 5$) as did the preschoolers ($n = 11$).

Examination of the children's error responses revealed random selections for most items. The children pointed to one of the three incorrect pictures at chance levels (33%) for 13 of these 16 items. Only items 37 and 59 for the preschoolers and item 56 for the kindergartners showed consistent alternative selections at levels approximately twice that predicted by chance (65%, 60%, and 72%, respectively).

Each child's raw score was adjusted by crediting the total raw score with one point for each of the items appearing in Table 2. These gains resulted in a statistically significant shift between the unadjusted and adjusted standard scores across the sample [pairwise $t(104) = 2.72$, p < .001]. However, these statistically significant gains failed to make a difference clinically for the sample as a whole. Most of the children (86%) still scored below the mean according to the PPVT-R norms, and 51% scored more than one standard deviation below the mean for the test's normative data (see Figure 1).

DISCUSSION

The changes in the standardization sample from the PPVT (Dunn, 1959) to the PPVT-R (Dunn & Dunn, 1981) did not increase the test's usefulness as a receptive vocabulary measure for low-income, urban, African American preschool and kindergarten children. Despite the inclusion of African American and low-income children in the test sample, the PPVT-R appears to be racially or economically biased. Ninety-one percent of the low-income, urban, African American children in this investigation scored below the mean, and more than half (65%) scored more than one standard deviation below the mean of the Dunn and Dunn sample. Further, the error analysis failed to identify consistent patterns of responding by the children. For most test items, the children seemed as likely to select one of the foils as another, indicating that they truly did not know the vocabulary item.

A scoring adjustment to allow credit for items missed by at least 50% of the children also produced no clinically significant differences in performance. Although the 50% level was arbitrary, it was generous. The preschoolers received additional credit for more than twice as many items as the kindergartners. However, even with twice as many items qualifying for possible credit, the preschooler's performance was not significantly different from the kindergartners, and most still performed well below the mean score. On the basis of the performance of these children, score adjustment does not appear to solve the validity problem of the PPVT-R for low-income, urban, African American children.

Perhaps African American children who belong to this segment of the U. S. population do have some deficiencies in vocabulary as a result of their impoverished environment. They may, indeed, have lower than average scores on the PPVT-R when compared to children in the nation as a whole. In this sense, the PPVT-R tells us something only about the population, and what it tells us is fairly predictable. The value of a diagnostic test, however, is to tell us something about the individual. Does the child have a language-learning disorder? The current data indicate that the scores attained by these children do not represent enough performance spread on the normal curve for this test to be revealing for these children.

The preponderance of subjects scoring at the low end of the performance curve indicates that this population does not have many of the vocabulary items tested by the PPVT-R in their repertoire. Unfortunately, it does not provide a clear picture of the vocabulary that these children do possess. Since the PPVT-R does not provide much information about the size or composition of the lexicon, it does not appear to be helpful either diagnostically or descriptively with this population.

On the basis of the information provided in the test manual (Dunn & Dunn, 1981), it appears that the members of the sample representing different minority groups and socioeconomic levels were simply included as members in the large standardization sample. Their scores were not evaluated separately to determine whether differences in vocabulary development existed. The results of this study suggest that these children probably cluster at the low-scoring end of the PPVT-R standardization sample. This raises critical questions about how representative the normal distribution of the PPVT-R is for the other minority groups, as well as for the white children included in the standardization sample. It is possible that nonminority children and children from higher socioeconomic groups are overrepresented at the high end of the performance curve.

Several alternatives have been recommended to handle the cultural inequities apparent in many current standardized testing instruments. These include scoring adjustments (Vaughn-Cooke, 1986) and re-norming (Haynes & Moran, 1989). The former alternative was employed unsuccessfully in this investigation, and the latter alternative does not appear possible with the PPVT-R. Re-norming the PPVT-R might have been a viable alternative if there had been a greater performance spread on the
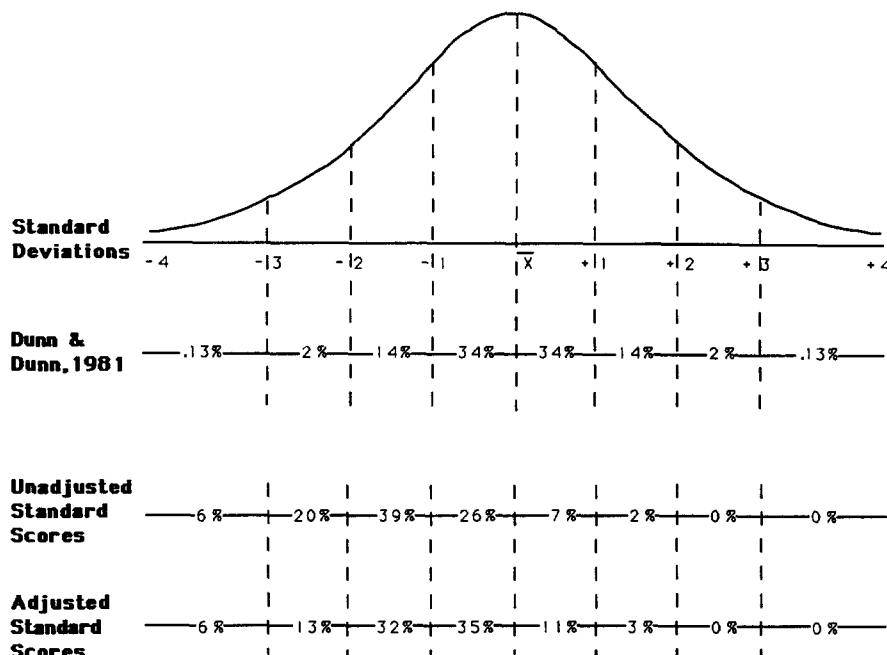
FIGURE 1. The distribution of current subjects on the PPVT-R compared to Dunn and Dunn (1981).

normal curve. However, because this is clearly not the case, attempting to re-norm this test seems counterproductive. Instead, it appears that developing a new test of vocabulary is necessary. A new test could be developed based upon the existing lexicon of these children, thus aiding both language stimulation and vocabulary learning by identifying what the children know and building on it.

In conclusion, the findings of this study indicated that the PPVT-R, like the original version of the PPVT (Dunn, 1959), is not appropriate for use with African American preschool and kindergarten children from low-income backgrounds. The testing problems for minority children, described by the test developers in terms of low-performance reliabilities, are supported by this study. The vocabulary selected by Dunn and Dunn (1981) does not appear appropriate for all children, particularly for those from different cultural backgrounds and low socioeconomic status. Consequently, clinicians utilizing the

TABLE 2. The item numbers missed by 50% or more of the subjects.

	Item #	Prompt	Subjects (%)	Alternative Selection
Preschool				
	27	somersault	64	—
	33	cobweb	52	—
	37	pail	66	Picture #4 (can)
	41	saddle	70	—
	43	eagle	59	—
	46	uniform	52	—
	53	floating	61	—
	58	printing	58	—
	59	waiter	50	Picture #1 (jumper/hurdler)
	60	root	70	—
	61	walrus	56	—
Total #	11			
Kindergarten				
	27	somersault	67	—
	56	hive	62	Picture #1 (observatory)
	60	root	56	—
	61	walrus	54	—
	73	flaming	62	—
Total #	5			

Note. — = no consistent response.

PPVT-R in their assessment protocols must be aware of the dangers of using the PPVT-R as a diagnostic instrument with low-income, African American children.

REFERENCES

ADLER, S., & BIRDSONG, S. (1983). Reliability and validity of standardized testing tools used with poor children. *Topics in Language Disorders, 3*, 76–88.

AMERICAN NATIONAL STANDARDS INSTITUTE. (1969). *Specifications for audiometers.* (ANSI S3.6-1969). New York: American National Standards Institute.

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*(3), 171–176.

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

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

HAYNES, W. O., & MORAN, M. J. (1989). A cross-sectional developmental study of final consonant production in southern Black children from preschool through third grade. *Language, Speech, and Hearing Services in Schools, 20*(4), 400–406.

KRESHECK, I., & NICOLOSI, S. (1973). A comparison of Black and White children's scores on the PPVT. *Language, Speech, and Hearing Services in Schools, 4*, 37–40.

SAVILLE-TROIKE, M. (1986). Anthropological considerations in the study of communication. In O. L. Taylor (Ed.), *Nature of communication disorders in culturally and linguistically diverse populations* (pp. 47–72). San Diego, CA: College Hill Press, Inc.

TAYLOR, O. L., & PAYNE, K. T. (1983). Culturally valid testing: A proactive approach. *Topics in Language Disorders, 3*, 8–20.

UNIVERSITY OF MICHIGAN DETROIT AREA STUDY (1989). *Separate and unequal: The racial divide.* University of Michigan Institute for Social Research.

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: College-Hill Press.

WEINER, P. (1979). *Peabody Picture Vocabulary Test (PPVT).* In F. Darley (Ed.), *Evaluation of Appraisal Techniques in Speech and Language Pathology* (pp. 49–52). Reading, MA: Addison-Wesley.

WOLFRAM, W. (1983). Test interpretation and sociolinguistic differences. *Topics in Language Disorders, 3*, 21–24.

Contact author: Julie A. Washington, PhD, Communicative Disorders Clinic, University of Michigan, 1111 E. Catherine Street, Ann Arbor, MI 48109-2054.