



Oral Language Expectations for African American Children in Grades 1 Through 5

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

Reference profiles for characterizing the language abilities of elementary-grade African American students are important for assessment and instructional planning. H. K. Craig and J. A. Washington (2002) reported performance for 100 typically developing preschoolers and kindergartners on 5 traditional language measures: mean length of communication units, amount of complex syntax production, number of different spoken words, responses to *wh*-questions, and understanding of active/passive sentence construction. The present study reports performances on the same measures for 295 typically developing African American children in the 1st through 5th grades. Findings revealed

increasing performance scores with increasing grades on 4 of the tasks. A ceiling effect was evident on the task that assessed comprehension of active and passive voice. Gender, socioeconomic status, and community influenced the values in systematic ways, and responses to requests for information varied relative to vocabulary skill. These measures are recommended for inclusion in culturally fair assessment protocols designed to characterize the language abilities of elementary-grade African American students.

Key Words: African American, language, minorities, assessment

Expectations for performance variations by typically developing children are fundamental to accurate identification of students with language impairments and to appropriate classroom planning. Performance standards for mainstream students are well established (Bloom & Lahey, 1978; Brown, 1973; Miller, 1991) and have been informative to classroom instruction, curriculum development, and construction of valid testing instruments. Unfortunately, there is limited availability of reference profiles for characterizing the skills of nonmainstream populations such as African American children, limiting our ability to establish comparable performance standards. Although reference profiles are available for preschool- and kindergarten-age African American children (Craig & Washington, 2002), comparable profiles have not been developed for elementary-grade African American students. Reference points that define the range for typically developing African American students are necessary to enable clinicians to better identify children who fall outside these ranges and thus may demonstrate language deficits in particular skill areas. The purpose of this investigation, therefore, is to extend the prior work of Craig and Washington (2002), which described the oral language performances of typically developing African American preschool and kindergarten students on five fairly

traditional assessment measures: mean length of communication units (MLCU), production of complex syntax (Csyn), number of different words (NDW), responses to *wh*-questions (Wh-q), and comprehension of active/passive voice construction using reversible sentences (RevS). The current study examined the same measures for first- through fifth-grade African American students.

Findings from Craig and Washington (2002) indicated that performance on the three oral language production measures (MLCU, Csyn, NDW) did not change significantly between preschool and kindergarten. The students' performances on the comprehension measures (Wh-q, RevS) were mixed. Wh-q improved by grade, with the kindergartners outperforming the preschoolers. However, no grade level differences were evident between preschoolers and kindergartners in their ability to comprehend active and passive voice. Performances on these five measures at later grades may remain stable or may increase. For mainstream students, oral production measures increase with grade, although not necessarily showing significant differences between consecutive grades (Loban, 1976; Nippold, 1998; Scott, 1988). It is not clear, therefore, whether the values reported by Craig and Washington (2002) are appropriate reference levels for African American students as they

progress through the elementary grades. It is likely that higher scores are more appropriate.

Extending the descriptions of Craig and Washington (2002) to characterize the receptive and expressive language skills of African American students in the elementary grades will be particularly valuable for understanding their language at a time when language skills must support the development of reading. Measures of these types will complement scores obtained from the growing number of culturally fair standardized tests of oral language currently available for African American students. Tests shown to be culturally fair (Thomas-Tate, Washington, Craig, & Packard, 2005; Washington & Craig, 1992, 1999) include the Peabody Picture Vocabulary Test—Third Edition (PPVT—III; Dunn & Dunn, 1997), the Expressive Vocabulary Test (Williams, 1997), and the Arizona Articulation Proficiency Scale—Second Edition (Fudala & Reynolds, 1986). The Diagnostic Evaluation of Language Variation (Seymour, Roeper, & de Villiers, 2003), a criterion-referenced instrument, also provides a comprehensive set of tasks designed to characterize the oral language status of African American students who speak African American English (AAE). Moreover, this full set of culturally fair language assessment measures can be used in conjunction with standardized tests of reading, also recently shown to be culturally fair (Craig, Thompson, Washington, & Potter, 2004).

The five measures used in Craig and Washington (2002) are long-standing and have been sensitive to performance changes associated with development. Loban (1976) found that increases in communication unit (C-unit) length across grades were related to increases in linguistic complexity. Complex syntax production has been important to characterizing language growth (Craig & Washington, 1994; Scott, 1988) and predicts later reading achievement (Craig, Connor, & Washington, 2003), and failure to acquire complex syntax is a core symptom of language disorder (Scott, 2004; Scott & Windsor, 2000). NDW, a measure of vocabulary size or breadth, is a sensitive measure for distinguishing typically developing from language impaired preschool-age children (Watkins, Kelly, Harbers, & Hollis, 1995). In addition, children's responses to requests for information and comprehension of active/passive voice construction change with age. These changes are associated with semantic and syntactic growth for both mainstream and African American preschool-age children (Craig & Washington, 2002; Tyack & Ingram, 1977) and have strong ecological face validity to the types of instructions presented to students across elementary-grade classrooms. The Wh-q task is notable among the measures in this regard. As students read more in the academic content areas and increasingly engage in instructional activities in literature, science, and math, more complex and cognitively demanding requests are made of them to synthesize information and make inferences and deductions (Van Den Broek & Kremer, 2000). The Wh-q task probes their ability to respond to demands of these types.

Many African American students speak AAE (Craig & Washington, 2002, 2004; Washington & Craig, 1994). Child AAE can involve the production of as many as 30 features that contrast with comparable renderings of the same meanings in Standard American English (SAE);

Craig, Thompson, Washington, & Potter, 2003; Craig & Washington, 2004). For example, a major feature characterizing the spoken discourse of students is the variable subject-verb agreement feature, such that the morpheme *-s* is variably included (e.g., “because he looks like a baby”) and excluded (“she look_ cute for the wedding”), and both forms are appropriate within the dialect. Only the inclusion of the *-s* morpheme is appropriate within SAE, and thus these features that represent contrasts to SAE are referred to as *contrastive features* of the dialect. Many aspects of spoken discourse involve larger discourse units than morphemes and phonemes (e.g., clauses, C-units, and lexical stems). The MLCU, Csyn, NDW, Wh-q, and RevS tasks, selected for analysis by Craig and Washington (2002) and used again in the current study, involve units that are larger than the morpheme or phoneme and are noncontrastive. They can be scored without reference to whether the morphemic or phonemic structure of the response includes a contrastive AAE feature. Together, therefore, these five measures represent critical language acquisition skills, offer a broad look at the oral language abilities of African American children, and are culturally fair.

In addition to grade, any description of major nondialectal characteristics of the discourse of African American students needs to probe for systematic differences in performance relative to gender, socioeconomic status (SES), and community. Gender (Donahue, Daane, & Grigg, 2003; Stevenson, Chen, & Uttal, 1990), SES (Donahue et al., 2003), and community (Donahue et al., 2003; Thomas-Tate et al., 2005) can influence oral language performances and reading achievement for African American students in systematic ways. Craig and Washington (2002) found systematic differences for gender in NDW, so this variable may continue to exert an influence on lexical diversity at later grades and should be examined. Further, it is not yet known whether demographic variables such as these are systematically related to oral language performances of African American students when the measures are non-dialectal in nature. Characterization of the oral language skills of African American students will be critically incomplete until the influence of potentially important demographic variables is available. Gender, SES, and community type are particularly relevant for the study of oral language skills in this population because these three demographic variables exert systematic influences on oral language production skills that are dialectal in nature (Craig & Washington, 2004; Washington & Craig, 1998). Are gender, SES, and community influencing AAE feature production rates only, or are these influences observable more broadly across other major aspects of oral language?

Therefore, in order to better understand the oral language production and comprehension skills of elementary-grade African American students, the following questions were posed:

1. What are the means and standard deviations for first-through fifth-grade African American students' performances on MLCU, Csyn, NDW, Wh-q, and RevS?
2. Are there systematic relationships among performances on MLCU, Csyn, NDW, Wh-q, and RevS, and gender, SES, or community?

3. Are there systematic performance trends for MLCU, Csyn, NDW, Wh-q, and RevS related to grade?

Method

Participants

The participants were the same typically developing elementary-grade ($N = 295$) African American boys and girls who were the focus of the recent study of AAE production by preschool and elementary-grade students (Craig & Washington, 2004). For the purposes of the present investigation, separate analyses of the five oral language measures were conducted on the same first through fifth graders. The students were from low ($n = 98$) or middle ($n = 197$) SES homes. All of the children spoke AAE as demonstrated by production of at least two different types of AAE features during picture description, as in prior research by Craig, Thompson, et al. (2003, 2004). The elementary-grade students resided in two communities, either an urban-fringe community in the metropolitan Detroit area ($n = 163$) or a midsize central city ($n = 132$). The grade distributions for SES and community are presented in Table 1.

Students were judged to be typically developing based on the following factors: (a) parent and teacher report, (b) no history of referral to or enrollment in special education services, (c) performance within 2 *SDs* of the mean scaled score of 10 (≥ 4) on the Triangles subtest of the Kaufman Assessment Battery for Children (Kaufman & Kaufman, 1983), and (d) a score of 90% or greater on the Percentage of Consonants Correct—Revised (Shriberg, Austin, Lewis, McSweeney, & Wilson, 1997). Performance values were obtained for all 295 students on the MLCU, Csyn, and NDW. In addition, PPVT–III scores were available for 245 students (83% of the participant sample). PPVT–III standard scores ranged from 73 to 137. Thirty-seven of these children (15%) scored at least 1 *SD* below the mean of 100 on the PPVT–III, falling between 1 and 2 *SDs* below the mean. These students were not excluded from the study for two reasons: (a) approximately 14% of scores of a normally distributed sample are expected to fall between 1 and 2 *SDs* below the mean, consistent with the normal curve on the PPVT–III (Dunn & Dunn, 1997), and (b) many typically developing African American students fall below published normative means on standardized tests that are based primarily on mainstream students (Craig & Washington, 2004; Donahue et al., 2003; Washington & Craig, 1999). However, prevailing recommendations for cutoffs indicative

of language disorder would determine that students falling more than 1 *SD* below the mean demonstrate clinically significant performances on the language measure (Aram, Morris, & Hall, 1992; Tomblin, Records, & Zhang, 1996). For the purposes of the present study, therefore, students were retained in the study but assigned to one of two groups: below the mean ($n = 37$, standard score [SS] = 73–84) or within or above the mean ($n = 209$, SS = 85–137), based on their PPVT–III scores. Performances on MLCU, Csyn, NDW, Wh-q, and RevS were then examined for any systematic relationships to vocabulary group membership.

In the present study, gender and SES were allowed to vary. SES was determined by students' eligibility or ineligibility to participate in the federally funded free or reduced-price lunch program or by the Hollingshead Four Factor Index of Social Status (Hollingshead, 1975). Free or reduced-price lunch eligibility is established annually by the federal government, taking into account annual increases in the cost of living (U.S. Department of Agriculture, Food and Nutrition Service, 2004). The Hollingshead Index assigns points to caregivers based on a combination of factors: marital status, the gender of the primary caregiver, the highest level of education completed, and occupational status. Total points ranged from a minimum of 8 points to a maximum of 66 points based on responses to the queries. Point scores corresponded with one of five levels, which were used to characterize the caregivers' socioeconomic levels. For example, a single female head of household with a college degree who was employed as a teacher was assigned 53 points (Level 2), which placed her within the middle SES range. In contrast, a single mother with less than a high school education who was employed as a skilled worker obtained 25 points (Level 4), which placed her within the low SES range.

Data Collection and Analysis

Data collection included administration of a picture description task and two comprehension tasks: Wh-q and RevS. Tasks were administered individually in random order to reduce the risk of systematic order effects.

Collection and analysis of picture description samples. Oral language samples were collected during picture descriptions in which the participants described three action pictures (Numbers 5, 7, and 24) from the Bracken Concept Development Program (Bracken, 1986). Picture descriptions provided the language elicitation context in the present investigation. It was the context selected by Craig and

TABLE 1. Distribution of participants by grade, community, and socioeconomic status (SES), and their mean standard score performance on the Triangles subtest of the Kaufman Assessment Battery for Children.

Grade	<i>n</i>	Community		SES		Triangles	
		Midsize central city	Urban-fringe community	Low	Middle	<i>M</i>	<i>SD</i>
1st	39	9	30	11	28	10.6	2.5
2nd	60	21	39	19	41	10.7	1.8
3rd	69	32	37	28	41	10.7	2.4
4th	63	36	27	21	42	10.4	2.2
5th	64	34	30	19	45	10.6	2.2
Total	295	132	163	98	197	10.6	

Washington (2002) in their earlier examination of these five measures when the participants were preschoolers and kindergartners. Pictures were randomly ordered for each participant. During administration of this untimed task, the participants were given the following prompt: "Tell me as much as you can about this picture." If the child simply named the pictured actions and objects, he or she was provided an additional prompt: "Tell me what's happening/going on in this picture." Both the child and examiner were audio-recorded during this task.

After language sample collection, the samples were transcribed orthographically using the Coding for Human Analysis of Transcripts conventions of the Children's Data Exchange System (CHILDES; MacWhinney, 1994). Samples were segmented into C-units following the scoring guidelines that were developed by Loban (1976). C-units included independent clauses and their modifiers.

The language samples were scored for MLCU, frequencies of Csyn, and NDW. The *FREQ* command of the Computerized Language Analysis Program from CHILDES generated frequencies of Csyn types and tokens, and word lists for NDW. The *MLT* command generated mean length of turn in words. There were differences in how some of the measures were calculated by Craig and Washington (2002) compared to what was done in the present study, due to the older ages of the students in this investigation. A discussion of each of the measures used in the present study follows.

MLCU. The picture descriptions were untimed, the completion of the language sample was determined by the child, and the resulting number of C-units available for analysis varied across students. The length of the language samples did not vary systematically by grade, $F(4, 290) = 2.2, p = .07$. The mean number of total C-units that were produced per language sample was 26.2 ($SD = 9.8$). These C-unit corpora constituted the base for calculations of Csyn and NDW.

Calculation of the MLCU was based on the varying but total number of C-units collected during each picture description language sample. The number of words produced was divided by the number of C-units in each student's sample.

Csyn. The participants' language samples were coded for 11 Csyn types (Craig & Washington, 1994). Csyn types included simple sentence constructions such as simple infinitives with the same subject (e.g., "he don't need to stand up") and more advanced forms such as relative clauses (e.g., "that's the noise that I like"). Each Csyn type had the potential to be produced multiple times within each language sample. Multiple instances of a Csyn type were considered as a single type with multiple tokens. In the present study, in order to control for variable production of C-units across the participants, Csyn frequencies (tokens) were divided by the total number of C-units produced in the language sample of each participant and are reported as proportional frequencies.

NDW. NDW is a measure of diversity of expressive vocabulary. The *FREQ* command automatically generated NDW word lists. Adjustments were made to the word lists in order to replicate traditional methods of calculating NDW

(Miller, 1982; Templin, 1957). In calculating NDW, root words were considered as single words regardless of the morphological endings of the nouns and verbs that were produced. Therefore, words such as *stand*, *standing*, and *stands* were considered as three examples of the single word type *stand*. Irregular noun and verb forms such as *calf*, *calves*, and *see*, *saw* were calculated as separate word types. In the present investigation, NDW was divided by the total number of C-units in each sample and reported as rates.

Wh-q. The Wh-q task assessed students' ability to respond to requests for information. The Wh-q task in the current investigation was a longer version than that used by Craig and Washington (2002). In addition to the 12 original probes used by Craig and Washington, the task presented 7 new probes per picture. The new probes were more difficult than the original queries because (a) they required processing of more advanced cognitive relationships that included comparatives, predictions, and explanations; (b) they were longer; and (c) they were more syntactically complex. Probes included queries such as "What do you think will happen when the man/woman (action + object)?" "How is this like this?" and "How would you describe this person/people?" The Appendix presents scoring guidelines and some probes for the snow-shoveling scene with scoring possibilities. The participants were presented with two randomly ordered, colored action pictures; one depicted a snow-shoveling scene, and the second depicted a beach scene. In order not to rely on a single trial for determining the responses to specific probes that were easier or more difficult, one probe for each request type was administered per picture.

For the Wh-q, scores were assigned based on a system developed by Craig and Washington (2000). A total of 114 points were possible. Wh-q is based on a 3-point scale. Participants were assigned full credit (3 points) if they produced the target response; 2 points were given if the participant responded to the probe with a nonspecific or vague referent or misnamed the referent; 1 point was given if the student responded, however, to a different *wh-* probe, and zero points were assigned if the child failed to respond, produced an unacceptable response to the probe, produced an irrelevant utterance, simply labeled or described the picture, or indicated "I don't know" in response to the query.

RevS. RevS examined the students' understanding of word order strategies. A total of 30 spoken probes were presented to each participant in random order in groups of 3. No modifications in the probes administered or scoring procedures were made to the RevS task for the present study compared to the one used by Craig and Washington (2000). Over the course of data collection, it became apparent to the examiners in the field that students, even first graders, were achieving 100% accuracy on this task. Therefore, in the 3rd year of the project, the RevS task was not administered to any additional students. As a result, this task was administered to 61% of the participants, so that scores were available for 180 of the 295 students in the sample.

A total of 20 points could be obtained on this task. One point was assigned for correct responses to both of the active voice stimuli, and 1 point was given for a correct response to the passive voice probe. Credit was given for a correct response to a passive voice prompt only if both of the active

voice prompts were responded to correctly. A computerized scoring program generated total raw scores.¹

Reliability

Several reliability checks were applied to the data analyses. Transcription reliability at the level of the morpheme and C-unit segmentation reliability were conducted on one third (one picture) of the C-units of every language sample. Scoring reliability for Csyn was conducted on 60 language samples that were randomly selected from the first through fifth graders, representing 20% of the total participant samples that were rescored for Csyn. Reliabilities were calculated by dividing total agreement by agreements plus disagreements. Transcription and C-unit segmentation reliabilities were high across subjects, with 97% and 99% agreement, respectively. Point-to-point agreement for Csyn types was 95%, and agreement for Csyn (tokens) was 88%.

Results

This study describes the performance of first- through fifth-grade African American students on five traditional language measures: MLCU, Csyn, NDW, Wh-q, and RevS. These five measures all examine oral language performances and have the potential to be highly related. In order to avoid Type I errors among potentially related variables, in which statistical differences are interpreted as real differences when they are not, an alpha level of .05 was set for the study as a whole and then adjusted to .01 for the interpretation of main effects. The .01 level was determined by dividing the preset level by the number of measures (.05/5 = .01). In the analyses that follow, all main effects are interpreted as statistically significant only when $p \leq .01$. The findings provide means and standard deviations for each measure and report systematic variations based on gender, SES, community, and grade, and are discussed below.

Whereas a number of students (15%, $n = 37$) scored at least 1 *SD* below the mean of 100 on the PPVT-III, the first analysis explored whether vocabulary skill influenced performance on the five oral language measures. The performance distributions of students assigned to the below-the-mean group ($n = 37$, $SS < 85$) and the within-or-above-the-mean group ($n = 208$, $SS \geq 85$) were examined first. Application of the one-sample Kolmogorov-Smirnov test of normality confirmed that the distributions of both groups were not different from a normal distribution, $Z(37) = .865$, $p = .442$; $Z(208) = .824$, $p = .506$. The performances of the two groups then were compared on the five measures of oral language. Results from independent *t* tests revealed no significant differences between the two groups on four of the five oral language measures. However, the within-or-above-the-mean group outperformed the below-the-mean group on the Wh-q task, $t(243) = 4.7$, $p = .000$.

Therefore, in subsequent analyses, the data were collapsed for the four measures that evidenced no difference between vocabulary skill groups, but the Wh-q performances were examined separately by group.

Each of the language measures also was examined for systematic sources of variation related to gender, SES, community, and grade using multivariate analysis of variance (MANOVA). For the MANOVA, proportional frequencies of Csyn and the NDW per C-unit were treated with an arcsine transformation in order to manage some of the variability inherent in proportional data. RevS was examined in a separate univariate analysis of variance because only a subset of the participants (61%, $n = 180$) was administered this task.

MLCU

MLCU evidenced a significant main effect for gender, $F(1, 254) = 7.2$, $p = .008$. A small variance-accounted-for correlation effect size ($\eta^2 = .03$) was associated with this main effect. As discussed by Meline and Paradiso (2003), effect sizes represent the amount of individual variability that can be explained by the group membership of the participants and can be interpreted in terms of the practical importance of the finding. In the current study, the small effect size for gender indicated that it had a negligible impact on average C-unit lengths despite its statistical significance.

On average, the girls produced longer C-units ($M = 7.1$, $SD = 1.6$) than the boys ($M = 6.8$, $SD = 1.7$). No main effects were observed for SES, $F(1, 254) = 1.2$, $p = .28$, or community, $F(1, 343) = 2.3$, $p = .13$.

For MLCU, there was a significant main effect for grade, $F(4, 254) = 7.4$, $p = .000$. The variance-accounted-for correlation effect size ($\eta^2 = .11$) indicated that grade predicted only a small amount of the variance of MLCU. The small effect size appeared related to the pattern of differences across grades. Table 2 presents means and standard deviations for MLCU by grade. Tukey honestly significant difference post hoc comparisons revealed significant differences between grades, but only for alternating grades. For example, the mean performance of the first graders was significantly different from the third through fifth graders but not from the second graders. As can be seen in Table 2, a noncontiguous relationship was observed between MLCU and grade for the second through fifth graders.

A significant interaction effect was observed between grade and SES, $F(4, 254) = 2.9$, $p = .02$, but the independent variables accounted for only a small amount of variance ($\eta^2 = .04$) for MLCU. The fifth graders from middle SES backgrounds ($M = 8.2$, $SD = 2.0$) produced longer C-units than the fifth graders from low SES homes ($M = 7.0$, $SD = 1.2$).

Csyn

For Csyn, a significant main effect was observed for gender, $F(1, 254) = 8.3$, $p = .004$. A small variance-accounted-for correlation effect size ($\eta^2 = .03$) indicated that gender predicted a small but negligible amount of the variance of Csyn. On average, the girls produced more Csyn per C-units ($M = 0.66$, $SD = 0.29$) than the boys ($M = 0.60$, $SD = 0.32$). No significant main effects were evident for SES, $F(1, 254) = 0.82$, $p = .37$, or community, $F(1, 254) = 1.6$, $p = .21$.

¹The computerized task is available from the authors.

TABLE 2. Means and standard deviations for mean length of communication units (MLCU), proportional frequencies of complex syntax (Csyn), and rates for number of different words (NDW) for all students (N = 295).

Grade	n	MLCU		Csyn		NDW	
		M	SD	M	SD	M	SD
1st	39	6.0 ^{1,2,3}	1.2	.45 ^{7,8,9}	.24	3.0 ^{13,14}	0.61
2nd	60	6.4 ^{4,5}	1.4	.52 ^{10,11}	.25	3.2 ^{15,16}	0.72
3rd	69	7.0 ^{1,6}	1.5	.64 ^{7,12}	.28	3.4 ¹⁷	0.74
4th	63	7.3 ^{2,4}	1.4	.69 ^{8,10}	.28	3.5 ^{13,15}	0.77
5th	64	7.9 ^{3,5,6}	1.9	.82 ^{9,11,12}	.32	3.8 ^{14,16,17}	0.84

Note. Matching superscripts (¹⁻¹⁵) indicate significant relationships between grades. For example, 6.0^{1,2,3} indicates that the first graders differed significantly from the third, fourth, and fifth graders on MLCU.

¹⁻¹⁵ $p \leq .05$.

A significant main effect was observed for grade, $F(4, 254) = 10.1, p = .000$. Grade explained a small amount of the variance of Csyn ($\eta^2 = .14$). Similar to MLCU, Csyn showed noncontiguous relationships across grades, with significant increases every other grade. For example, the mean performance of the first graders was significantly different from the third through fifth graders but not from the second graders. Similarly, the performances of the second graders varied systematically from the fourth and fifth graders but were not significantly different from the third graders. Table 2 shows the relationships among the five grades.

The grade-related differences were associated with different profiles of Csyn types across grades. Some types were produced rarely regardless of grade: let(s)/lemme and infinitive, tag questions, and *wh*- infinitive clause. Other types were produced at comparable levels regardless of grade: gerunds and participles, simple infinitive with same subject, and unmarked infinitive. Five types increased with increasing grades: simple noninfinitive *wh*- clause, infinitive with a different subject, relative clause, noun phrase complement, and clauses joined by conjunctions. These five types included embedded clauses, subject changes, and the addition of new sentence-level meanings. The statistically significant higher percentage frequencies of Csyn with increasing grade, therefore, appeared to reflect greater use of more advanced syntactic relationships at third grade and then again at fifth grade.

Csyn evidenced a significant interaction effect between gender, SES, and community, $F(1, 254) = 4.5, p = .03$, with a very small variance-accounted-for correlation effect size ($\eta^2 = .02$). The boys and girls from low and middle SES homes who lived in the midsize central city ($M = 0.69, p = .31$) produced significantly more Csyn per C-unit than the boys from low SES backgrounds ($M = 0.46, p = .23$) who resided in the urban-fringe community.

NDW

Analysis of NDW indicated no main effects for gender, $F(1, 254) = 0.66, p = .42$, SES, $F(1, 254) = 0.15, p = .70$, or community, $F(1, 254) = 0.32, p = .58$.

A significant main effect was observed for grade, $F(4, 254) = 5.2, p = .001$, and again accounted for a small

amount of the variance for NDW ($\eta^2 = .08$). As observed for the prior measures, the small effect sizes appeared related to the lack of statistically significant differences between consecutive grades on this measure. Post hoc comparisons revealed systematic differences for every two to three grade increases. For example, the first graders produced significantly fewer different words than the fourth and fifth graders but were not significantly different from the second and third graders. As can be seen from Table 2, noncontiguous relationships were evident through fifth grade. There were no interaction effects for NDW.

Wh-q

Whereas vocabulary group impacted performances on the Wh-q task but not the other four measures, the average and the below-the-mean groups were analyzed separately. For responses to Wh-q by the average group, no significant main effects were observed for gender, $F(1, 208) = 0.36, p = .55$, SES, $F(1, 208) = 0.03, p = .86$, or community, $F(1, 208) = 1.4, p = .24$.

A significant main effect was evident for grade for the within-or-above-the-mean vocabulary group, $F(4, 208) = 3.8, p = .005$. The variance-accounted-for correlation effect size was small ($\eta^2 = .08$), which indicated that grade predicted a small amount of the variance of Wh-q. Like the prior measures, the small effect sizes appeared related to the pattern of grade-related differences, but for Wh-q the pattern was different than the other measures. For Wh-q, the first graders were significantly different from the second through fifth graders. There were no significant performance differences between the second through fifth graders. No significant interaction effects were observed for Wh-q.

Correct responding on the Wh-q was probed further for the within-or-above-the-mean vocabulary group. Participants' correct responses to both trials of each question form were examined to determine whether grade-related differences simply reflected more accurate overall responding regardless of question type or were associated with increased comprehension of more advanced *wh*-question forms. First, the number of request types eliciting correct responses (3 points) were examined for distributional differences. For the first graders, the total number of correct responses per

request type ranged from 8 to 17, and for the second through fifth graders, total correct responses to both trials ranged from 9 to 18. Interestingly, nearly 70% of the second through fifth graders responded correctly to both trials of 14 or more questions, whereas less than one third of the first graders responded correctly to a comparable number of questions. These findings indicate that, at least in part, grade-related differences reflected more accurate responses overall. In a second analysis, however, the two most difficult question forms “how like” and “how describe” were examined. In response to these two most difficult question types, the third and the fifth graders outperformed the first graders. When the Wh-q task was expanded to include more advanced probes, significantly higher percentage frequencies of correct responses at second through fifth grades appeared to reflect advances in the ability to process more cognitively and syntactically complex linguistic structures.

The small sample size of the below-the-mean group ($n = 37$) precluded statistical treatment by gender, SES, community, and grade. However, the primary data trend for the within-or-above-the-mean vocabulary group was a difference between first and second through fifth grade as discussed previously. In order to determine whether comparable patterns might characterize students in the below-the-mean group, first graders were compared to second through fifth graders. An independent t test revealed a statistically significant difference in the scores obtained by first compared to second through fifth graders in the below-the-mean group: $t(35) = 4.6, p = .000$, with a very large effect size ($d = 1.7$). These findings are suggestive that performances on Wh-q present grade level differences between first and second through fifth grades, regardless of vocabulary skill level. However, the large effect size for the below-the-mean vocabulary skill group indicates that the developmental differences represented by their significant grade effects were stronger than for the within-or-above-the-mean vocabulary group, which evidenced only small effect sizes for grade.

RevS

For RevS, there were no significant main effects for gender, $F(1, 145) = 1.7, p = .19$, SES, $F(1, 145) = 0.77, p = .38$, or community, $F(1, 145) = 3.5, p = .07$. Unlike the findings for the other language measures, for RevS there was not a significant main effect for grade, $F(1, 145) = 0.33, p = .86$. The combined mean for the first through fifth graders on this task was 18.7 ($SD = 1.7$). For RevS, a significant interaction effect was evident between gender, SES, and grade, $F(4, 145) = 2.9, p = .02$. The variance-accounted-for correlation effect size was very small ($\eta^2 = .07$). The second-grade girls from middle SES homes ($M = 19, SD = 1.3$) outperformed the boys in second grade from low SES homes ($M = 16.3, p = 3.5$) on RevS.

Discussion

The present study characterizes, in quantitative terms, the performances of typically developing first- through fifth-grade African American students on five traditional

descriptive oral language measures and as such is an extension of the earlier work with preschoolers and kindergartners by Craig and Washington (2002) on the same measures. The findings from the present study contribute new reference information about the oral sentence structure and comprehension skills of typically developing elementary-grade African American students. These findings should be helpful to practitioners and teachers as they develop culturally fair assessment batteries and address the educational needs of school-age African American students. The major findings are as follows:

1. The vocabulary skills of this cohort of elementary-grade African American students with reportedly typical language acquisition skills distributed into two groups. The students with low ($SS < 85$ on the PPVT-III) and high vocabulary scores ($SS \geq 85$) performed differently on the Wh-q task but not the other measures.
2. Of the potential external sources of variation examined in this study (gender, SES, and community), gender influenced the outcomes on two of the measures and SES interacted with the other variables on three measures. These findings were detected even though the correlation effect sizes were small.
3. Grade was a systematic source of variation for African American students' performances on these traditional language measures. Statistically significant grade differences were detected for all measures except RevS, again despite small effect sizes, and revealed that most performance increases were across nonconsecutive grades.

Influence of Potential Sources of Variation on the Oral Language Measures

Approximately 15% of the participants scored between 1 and 2 SDs below the mean on the PPVT-III. This level of low vocabulary skill is consistent with the structure of the PPVT-III and predictable from the normal curve. Students grouped by low vocabulary skill level performed like their peers with within-or-above-the-mean vocabulary skills on MLCU, Csyn, NDW, and RevS, indicating that low vocabulary may be a fairly circumscribed deficit for these particular students. The one measure affected by vocabulary group membership was the Wh-q task, with the students in the within-or-above-the-mean vocabulary skill group significantly outperforming those in the low vocabulary skill group when responding to requests for information. This association between the two tasks may be due, at least in part, to the spoken responses required for the Wh-q task. A response that indicated recognition of the question type but was formulated using vague or incorrect lexical choices received a lesser score of 2 (out of 3). Therefore, students with limited vocabularies would score lower on Wh-q than students with stronger vocabularies.

Gender differences were observed for MLCU and Csyn in the present study. Girls produced longer sentences and more complex syntax than boys. The observation of small circumscribed gender effects on language development is consistent with the mainstream literature (Bauer, Goldfield, & Reznick, 2002; Ely, Gleason, & McCabe, 1996). Girls

tend to be more verbal than boys (Hyde & Linn, 1988). These differences generally do not reflect cognitive differences between boys and girls but likely reflect different ways that girls and boys are socialized and the differential treatment they receive from adult role models (Cherry & Lewis, 1976). For example, adults use longer utterances, more repetitions, and more questions when addressing girls and more directives and imperatives when communicating with boys (Cherry, 1975; Cherry & Lewis, 1976). These discourse practices have the potential to influence the language use of children over time. Measures of language use that reflect verbal productivity (e.g., longer sentences and more embedded sentences) may be particularly sensitive to gender-based differences in socialization practices. Therefore, girls may outperform boys as observed in this study, when measures reflect verbal productivity. It is important, therefore, not to place clinical significance on the lower performance of boys on measures such as MLCU and Csyn. Unless otherwise indicated with additional testing, lower performances by boys on measures reflecting verbal productivity do not represent a language disorder or delay, simply a gender-based difference in oral language production. It may be useful, if teachers expect longer, more elaborated responses from their male students than they are receiving, to provide opportunities during instruction and discourse to facilitate increased verbal production. Boys might be prompted simply to “say something more.” Increasing length of C-units should yield more frequent formulation of complex syntax as well. Whereas there were no differences on the other measures, with additional prompts for length and elaboration, the quality of the boys’ responses should be equivalent to the girls.

There were no main effects of SES or community on the dependent measures. Three interaction effects were observed in the present investigation. For MLCU, there was an interaction between SES and grade. For Csyn, there was an interaction between gender, SES, and community, and for RevS, there was an interaction between gender, SES, and grade. Although the participant sample as a whole was quite large ($N = 295$), it was not large enough to prevent some small cell sizes (see Table 1), limiting interpretations of these findings. Potential interactions among important child external variables such as these need to be replicated and probed in greater depth. However, the current statistical interactions underscore the need for practitioners to recognize the complexities involved in determining the potential causes for any specific child’s language performances. Practitioners should exercise caution when interpreting differences observed among students in classrooms enrolling students from diverse backgrounds. The present data indicate that boys and girls from the same or different SES backgrounds, residing in the same or different communities, may vary in language performances, and all of these differences may fall within the range of typical variations.

Systematic Grade Differences on the Oral Language Measures

Grade was a systematic source of variation on the language performances of elementary-grade African

American students. Performances on four of the five language measures increased as grade level increased. Considered together, with the prior research by Craig and Washington (2002), systematic grade increases in the present study suggest that MLCU, Csyn, NDW, and Wh-q will be useful for profiling growth in the linguistic skills of African American students from preschool through the elementary grades. Practitioners can refer to means and standard deviations on each measure to determine approximate grade level expectations on the language measures that were examined in the present investigation, exercising caution in these interpretations while larger sample studies remain forthcoming.

In this study of elementary-grade students, when grade level differences occurred they were primarily nonconsecutive, occurring every other year or more. Differences in MLCU and Csyn were apparent at every other grade level (i.e., Grades 1, 3, and 5). The grade differences for NDW occurred approximately every two to three grades. These findings are consistent with prior studies that have shown (a) systematic increases by grade on a variety of measures for elementary-grade and preadolescent mainstream students and (b) that longer intervals, spanning more than one grade, were required to detect significant grade level differences (Loban, 1976; Nippold, 1998; Scott, 1988). Overall, the findings indicate that these three measures can be used to characterize major aspects of oral language for African American elementary-grade students as well as for their mainstream peers.

The data for Wh-q revealed a somewhat different relationship to grade. In the present study, Wh-q scores differed significantly between first grade and second through fifth grades. Considered together with prior research, it appears that at the time of school entry, preschoolers and kindergartners make rapid growth in their ability to respond to requests for information (Craig & Washington, 2002). Comprehension of request types continues to evidence rapid growth through first grade as reflected by the significant differences between consecutive grades on the Wh-q task for preschoolers compared with kindergartners, kindergartners compared with first graders (Craig & Washington, 2002; Craig, Washington, & Thompson-Porter, 1998), and first graders compared with second graders (the data from this study). This developmental change was even more pronounced for students with lower vocabulary skills. In this study, students performing below the mean on the PPVT-III evidenced large effect sizes for the grade-related increase in Wh-q comprehension between first and second grade. After first grade, comprehension of these request types no longer evidences statistically significant growth. It is not surprising that this period of rapid growth corresponds to times when the curriculum focuses on learning classroom routines, transitioning to more of the school day spent on academic subjects, and moving from half- to full-day schedules. The data from the present investigation indicate that one achievement during this period of preschool to second grade is the acquisition of comprehension skills for understanding major types of requests for information.

The second through fifth graders scored high on the Wh-q task. Of a potential 114 points, the average across all

participants was 106.0. The data from the present study indicate that clinicians and teachers can expect African American students to have essentially mastered all of the types of requests probed by this task at second grade. High scores on the Wh-q task (>100 , i.e., $106.0 - 5.3 = 100.7$) appear appropriate as a criterion-referenced measure of comprehension. Students can be expected to perform at this level on this task after first grade. Further, when clinicians identify students who are having difficulty responding to requests for information, the specific trials on the Wh-q task should be informative for helping to identify which types of requests must be taught explicitly.

Increasing performance scores appeared to represent increases in linguistic sophistication. Two of the measures, Csyn and Wh-q, permitted analysis of responding to easier versus more difficult forms. In both analyses, the students at the higher grades with the higher scores were using the most difficult forms. Five types of complex syntax increased from first to third and third to fifth grades, and these included some of the most complex syntactic types. For example, relative clauses, in which one clause is used to modify a main referent in another clause, were produced by 75% of the fifth graders but only 41% of the first graders. Further, the version of the Wh-q task presented to the students in this study included seven cognitively and syntactically difficult probes. The second through fifth graders outscored the first graders on the two most difficult types of Wh-q. These more qualitative analyses confirm that the increasing scores on these measures across the elementary grades reflect linguistic growth by the students.

Unlike the other measures, RevS showed no significant differences between grades for the first- through fifth-grade students in this study. This finding contrasts with performances at younger grades on this measure. Craig et al. (1998) found significant differences between preschoolers and first graders on this task. In the present study, a ceiling effect was observed at first grade on RevS. Considered together, the current data indicate that African American students can be expected to understand the passive voice distinction at first grade. Therefore, RevS should be useful as a criterion-referenced measure because comprehending differences in active and passive voice can be expected at first grade.

Considered as a whole, the quantitative and qualitative data reported in this study indicate that these five measures are appropriate and valid ways to characterize growth and language production and comprehension skills for African American students in the elementary grades. Further, two measures (Wh-q and RevS) offer criterion-referenced measures for the acquisition of important comprehension skills.

Implications

The goal of the current study was to gather oral language reference information for first- through fifth-grade African American students. This information should prove valuable to practitioners who work with elementary-grade African American students. Using the means and standard deviations for the language measures presented in the current study, practitioners can profile areas of strength and weakness for

African American students having difficulty learning and reading in a classroom context. The examples below illustrate how the values from the present study may be used to inform the assessment and instructional processes.

Sherel. Sherel was an AAE-speaking student from a low SES home, enrolled in a first-grade classroom in an urban-fringe community. Her teacher reported that Sherel had difficulty answering questions, following oral and written directions, and contributing information during class lessons. The school speech-language pathologist evaluated Sherel's language at the teacher's request and confirmed the basis for the teacher's concerns. Sherel's SS on the PPVT-III was low at 82. In addition, she performed below the expected mean on two out of the five language measures described in this study, and they were the two measures of comprehension. Her Wh-q score was 88. For first graders with low vocabulary like hers, based on the findings in this study the expected Wh-q mean is 94.1. Whereas the standard deviation associated with that mean is 3.1, her score should be no lower than 91.0 ($94.1 - 3.1 = 91.0$). At 88, her performance was low.

Sherel's RevS was 14. Whereas there were no grade level differences with RevS in the present study, her score should approximate the mean on this task at 18.0. As with Wh-q, the clinician determined the acceptable performance range by subtracting the standard deviation from the mean ($18.0 - 2.1 = 15.9$). At 14, Sherel's score fell below that lower limit of 15.9. In contrast to her comprehension scores, she performed at expected levels for MLCU (6.2), Csyn (.44), and NDW (3.5).

Overall, Sherel's performance profile revealed receptive language problems. Her receptive vocabulary, comprehension of requests, and understanding of a major syntactic sentence form all were below expectations based on the reference profiles available from the current study for African American students with backgrounds like Sherel. Her difficulties in understanding the teacher's directions, and in comprehending enough of the classroom instructions so that she could respond appropriately, are predictable from and consistent with her profiles of receptive language deficits.

Sherel was recommended for speech-language intervention directed at improving her comprehension skills. In addition, the clinician provided the teacher with a list of the requests for information, which Sherel understood, and those that she did not, based on her Wh-q responses. Armed with this information, the teacher was able to adapt many classroom question and answer routines to include those understood by Sherel. The teacher's reframing of instructions allowed this student to participate better in instructional activities while ongoing speech-language intervention focused on improving her understanding of other types of requests. As Sherel learned new request types, the teacher added these to her instructional set whenever possible.

Kwame. Kwame was an AAE-speaking student from a low SES home, enrolled in a fourth-grade classroom in an urban-fringe community. His teacher observed that he was experiencing significant difficulty keeping up with his classmates during reading activities and requested a speech-language evaluation as part of the planning for this student.

As part of a comprehensive speech-language evaluation, the PPVT-III and the five measures described in this study were administered. Kwame's SS on the PPVT-III was low (84). He also performed below the expected mean on MLCU (4.5), Csyn (.35), Wh-q (91), and RevS (14.8). An analysis of his response errors on the Wh-q task revealed a preponderance of 2-point responses. He responded to the specific question word but with a nonspecific referent, or he misnamed the target referents. His NDW of 3.0 was low but within the expected range, above 2.7 for fourth graders ($3.5 - 0.77 = 2.7$).

Kwame's speech and language profile revealed a broad set of expressive and receptive language deficits. Vocabulary (PPVT-III and Wh-q 2-point answers) and syntax (Csyn and RevS) both were problem areas for him and consistent with his difficulties in reading. Deriving meaning from text requires a number of convergent skills, including recognizing a word meaning when the lexical form is pronounced, and processing sentences constructed with even more complex syntactic relationships than the levels that characterize spoken discourse. Kwame's low MLCUs were judged to be a byproduct of his limited vocabulary and syntactic skills.

The speech-language pathologist designed a program of intervention for Kwame that targeted growth in his vocabulary and syntax skills. The teacher prioritized word concepts and sentence structures in upcoming lessons and provided them to the clinician as targets for intervention. In this way, the teacher and clinician were successful in coordinating Kwame's intervention goals with classroom demands.

Summary

This descriptive study provides reference information for performance on five oral language tasks. This information contributes to the increasing literature on valid language measures for African American students and thus improves the ability of practitioners to conduct culturally fair and informative assessments. This information is particularly important in the current educational climate where the consequences for failure to meet academic standards are large for both educators and students throughout the elementary and secondary school years.

Future research will be important to help clinicians fully utilize the current information. In particular, despite an overall participant sample size of 295 students, the cell sizes for the MANOVA that examined these effects were very low. This was no doubt due to the number of variables entering into the analyses, reducing the cell sizes for some analyses to very low numbers. Additional studies that include larger sample sizes by grade will permit better examination of the influence of important factors such as gender, SES, and community and their interrelationships on linguistic skills.

Overall, this study demonstrated that fairly traditional measures of oral language production and comprehension are informative and appropriate for use with elementary-grade African American students. Only five measures were targeted in this investigation. It will also be important for

future research to continue to provide clinicians with a larger repertoire of descriptive measures appropriate for assessment purposes with African American students that, like these five, are valid, administered efficiently, and representative of a child's language abilities.

Acknowledgments

This work was supported by the U.S. Department of Education, Office of Educational Research and Improvement, Grant R305T990368. The authors extend a special acknowledgment to the students, families, and school personnel who participated in this study, and to Stephanie Hensel, Erin Quinn, and Mary Packard for their valuable comments.

References

- Aram, D. M., Morris, R., & Hall, N. E. (1992). The validity of discrepancy criteria for identifying children with developmental language disorders. *Journal of Learning Disabilities, 25*, 549-554.
- Bauer, D. J., Goldfield, B. A., & Reznick, J. S. (2002). Alternative approaches to analyzing individual differences in the rate of early vocabulary development. *Applied Psycholinguistics, 23*, 313-335.
- Bloom, L., & Lahey, M. (1978). *Language development and language disorders*. New York: Wiley.
- Bracken, B. A. (1986). *Bracken concept development program*. San Antonio, TX: The Psychological Corporation.
- Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.
- Cherry, L. (1975). The preschool teacher-child dyad: Sex differences in verbal interaction. *Child Development, 46*, 532-535.
- Cherry, L., & Lewis, M. (1976). Mothers and two-year-olds: A study of sex-differentiated aspects of verbal interaction. *Developmental Psychology, 12*, 278-282.
- Craig, H. K., Connor, C. M., & Washington, J. A. (2003). Early positive predictors of later reading comprehension for African American students: A preliminary investigation. *Language, Speech, and Hearing Services in Schools, 34*, 31-43.
- Craig, H. K., Thompson, C. A., Washington, J. A., & Potter, S. L. (2003). Phonological features of child African American English. *Journal of Speech, Language, and Hearing Research, 46*, 623-635.
- Craig, H. K., Thompson, C. A., Washington, J. A., & Potter, S. L. (2004). Performance of elementary-grade African American students on the Gray Oral Reading Tests. *Language, Speech, and Hearing Services in Schools, 35*, 141-154.
- 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. (2004). Grade-related changes in the production of African American English. *Journal of Speech, Language, and Hearing Research, 47*, 450-463.

- Craig, H. K., Washington, J. A., & Thompson-Porter, C.** (1998). Performances of young African American children on two comprehension tasks. *Journal of Speech, Language, and Hearing Research, 41*, 445–457.
- Donahue, P., Daane, M., & Grigg, W.** (2003). *The nation's report card: Reading highlights 2003* (NCES No. 2004452). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Educational Statistics.
- Dunn, L., & Dunn, L.** (1997). *Peabody Picture Vocabulary Test—Third Edition*. Circle Pines, MN: AGS.
- Ely, R., Gleason, J. B., & McCabe, A.** (1996). “Why didn’t you talk to your mommy, honey?”: Parents’ and children’s talk about talk. *Research on Language and Social Interaction, 29*, 7–25.
- Fudala, J. B., & Reynolds, W. M.** (1986). *Arizona Articulation Proficiency Scale—Second Edition*. Los Angeles: Western Psychological Services.
- Hollingshead, A. B.** (1975). *Four Factor Index of Social Status*. New Haven, CT: Yale University, Department of Sociology.
- 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: AGS.
- Loban, W.** (1976). *Language development: Kindergarten through grade twelve*. Urbana, IL: National Council of Teachers.
- MacWhinney, B.** (1994). *The CHILDES project: Tools for analyzing talk* (2nd ed.). Pittsburgh, PA: Carnegie Mellon University.
- Meline, T., & Paradiso, T.** (2003). Evidence-based practice in schools: Evaluating research and reducing barriers. *Language, Speech, and Hearing Services in Schools, 34*, 273–283.
- Miller, J.** (1982). *Assessing language production in children*. Baltimore: University Park Press.
- Miller, J.** (1991). Quantifying productive language disorders. In J. F. Miller (Ed.), *Research on child language disorders: A decade of progress* (pp. 211–220). Austin, TX: Pro-Ed.
- Nippold, M. A.** (1998). *Later language development: The school-age and adolescent years*. Austin, TX: Pro-Ed.
- Scott, C. M.** (1988). Spoken and written syntax. In M. Nippold (Ed.), *Later language development: Ages nine through nineteen* (pp. 49–96). Boston: College Hill Press.
- Scott, C. M.** (2004). Syntactic contributions to literacy learning. In C. A. Stone, E. R. Silliman, B. J. Ehren & K. Apel (Eds.), *Handbook of language and literacy: Development and disorders* (pp. 340–362). New York: Guilford Press.
- Scott, C. M., & Windsor, J.** (2000). General language performance measures in spoken and written narrative and expository discourse in school-age children with language learning disabilities. *Journal of Speech, Language, and Hearing Research, 43*, 324–339.
- Seymour, H. N., Roesper, T. W., & de Villiers, J.** (2003). *The Diagnostic Evaluation of Language Variation*. San Antonio, TX: The Psychological Corporation.
- Shriberg, L. D., Austin, D., Lewis, B. A., McSweeney, J. L., & Wilson, D. L.** (1997). The Percentage of Consonants Correct (PCC) metric: Extensions and reliability data. *Journal of Speech, Language, and Hearing Research, 40*, 708–722.
- 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.
- Templin, M. C.** (1957). *Certain language skills in children: Their development and interrelationships*. Westport, CT: Greenwood.
- Thomas-Tate, S. R., Washington, J. A., Craig, H. K., & Packard, M. E. W.** (2005). *Performances of African American preschool and kindergarten students on the Expressive Vocabulary Test*. Manuscript submitted for publication.
- 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.
- Tyack, D., & Ingram, D.** (1977). Children’s production and comprehension of questions. *Journal of Child Language, 4*, 211–224.
- U.S. Department of Agriculture, Food and Nutrition Service.** (2004). *School meals programs: Income eligibility guidelines*. Retrieved June 11, 2004, from <http://www.fns.usda.gov/cnd/Governance/notices/iegs/IEGs.htm>
- Van Den Broek, P., & Kremer, K. E.** (2000). The mind in action: What it means to comprehend during reading. In B. M. Taylor, M. F. Graves & P. Van Den Broek (Eds.), *Reading for meaning: Fostering comprehension in the middle grades* (pp. 1–31). New York: Teachers College Press.
- Washington, J. A., & Craig, H. K.** (1992). Performances of low-income, African American preschool and kindergarten children on the Peabody Picture Vocabulary Tests-Revised. *Language, Speech, and Hearing Services in Schools, 23*, 329–333.
- Washington, J. A., & Craig, H. K.** (1994). Dialectal forms during discourse of urban, African American preschoolers living in poverty. *Journal of Speech and Hearing Research, 37*, 816–823.
- Washington, J. A., & Craig, H. K.** (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.** (1999). Performances of at-risk, African American preschoolers on the Peabody Picture Vocabulary Test—III. *Language, Speech, and Hearing Services in Schools, 30*, 75–82.
- Watkins, R. V., Kelly, D. J., Harbers, H. M., & Hollis, W.** (1995). Measuring children’s lexical diversity: Differentiating typical and impaired language learners. *Journal of Speech and Hearing Research, 38*, 1349–1355.
- Williams, K.** (1997). *Expressive Vocabulary Test*. Circle Pines, MN: AGS.

Received October 21, 2003

Revision received June 11, 2004

Accepted February 25, 2005

DOI: 10.1044/1058-0360(2005/013)

Contact author: Holly K. Craig, University Center for the Development of Language and Literacy, University of Michigan, 1111 E. Catherine Street, Ann Arbor, MI 48109-2054. E-mail: hkc@umich.edu

Julie A. Washington is now at Wayne State University, Detroit, MI. Connie A. Thompson is now with the Ypsilanti Public Schools, Ypsilanti, MI.

Appendix**Responses to *Wh*-Questions Scoring Instructions and Probes**

Scoring guidelines

3 points: Target response is provided.

2 points: Child responds to the specific question word with a nonspecific referent or misnames the referent.

1 point: Participant responds but provides answer to a potentially different question.

0 points: Participant provides no response, provides a response that was not an acceptable target to any of the question probes, or produces an unrelated utterance.

Prompt	Response	Score
What this? (points to snowman)	a snowman	3
	snow	2
	he holdin' a shovel	1
Who car this? (points to car in garage)	the mama, daddy, theirs (pointing to family in house)	3
	the man who shoveling the snow in the garage	2
		1
How long will it take to dig out the car? (points to man in picture)	2 hours, a long time	3
	a little, five, seven	2
	he shovelin' the snow	1
Which kids are going sledding?	the ones on the hill	3
	the ones with the hats	2
	three	1
Why do you think the mom wants the kid to stand still?	so she can fix his hat	3
	so she can wipe her face	2
	go inside and play	1
When this boy throws his snowball, what do you think this boy will do?	throw one back	3
	make a snowman	2
	snowball	1
