Performances of young African American children on two comprehension tasks
Holly K Craig; Julie A Washington; Connie Thompson-Porter
Journal of Speech, Language, and Hearing Research; Apr 1998; 41, 2; Research Library
pg. 445

Performances of Young African American Children on Two Comprehension Tasks

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

This investigation examines the performances of 63 urban 4- to 7-year-old African American children from middle-income homes on two tasks designed to assess the development of comprehension skills. Performances on a task designed to elicit responses to wh-questions, and another to distinctions between active and passive sentence constructions, revealed grade effects and a positive relationship to age. The findings are discussed in terms of the appropriateness of using tasks of these types with young African American boys and girls who are dialect users.

KEY WORDS: language comprehension, African American, children, language development

Very little is known about the language development of the young African American child. This is problematic because the African American segment of the population of the United States is significant, currently the largest minority group, and is increasing at a steady rate, especially in large urban centers (Statistical Abstract of the United States, 1990). In addition, the numbers of African American children on the nation's special education caseloads are disproportionately large (Statistical Profile of Special Education in the United States, 1994), and the reasons for this are unclear. At least part of the explanation may be that the course of normal language acquisition has not been determined for African American children.

At the University of Michigan our research program focuses on characterizing the language development of young African American children, an important prerequisite to establishing appropriate assessment procedures for this population. When strong normative statements and profiles of key language behaviors across development are available, clinicians will be better situated to distinguish dialectal variations from language disorder.

Most of what is known about the language development of African American children focuses on aspects of language production. Young African American children residing in metropolitan Detroit primarily are speakers of African American English (AAE) (Washington & Craig, 1994; Washington, Craig, & Kushmaul, 1998). Their discourse is characterized by a set of 16 different morphosyntactic features (Washington & Craig, 1994) with little variation in articulation forms from those of Standard American English (SAE) (Washington & Craig, 1992, 1994, in press). Amounts of dialect in the spontaneous discourse of young African American children vary considerably across children. Gender,

socioeconomic status, and the narrative character of the discourse task influence frequencies of dialect marking by children in systematic ways (Washington & Craig, in press; Washington et al., 1998).

In addition to dialect use, our research program is attempting to characterize other aspects of the young African American child's developing linguistic system. Craig, Washington, & Thompson-Porter (1998) have reported average sentence lengths, calculated as mean communication unit (C-units) lengths in words, for African American preschoolers and kindergartners from urban, low-income homes. Average C-unit lengths in words were found to increase across the 4 to 6 1/2 year age span, increased significantly by grade, and related positively to increasing syntactic complexity and to amounts of dialect in the children's discourse. These data indicate that production performance gains are sensitive to developmental advances, indexed by age and/or grade.

Any profile of typically developing language skills would be critically incomplete if just oral production skills were considered, without complementary characterization of comprehension development. It is important to understand comprehension skills because they relate in fundamental ways to the child's ability to internalize and interpret external events, processes fundamental to building mental models (Golinkoff & Hirsh-Pasek, 1995). The study of children's comprehension skills can contribute to our understanding of the acquisition of spoken constructs as well, particularly the steps leading up to the spoken use of key forms (e.g., passive constructions). In addition, comprehension probes can avoid the sampling error potentially affecting outcomes of spontaneous production tasks when targeted behaviors may not occur readily during interaction. Accordingly, comprehension warrants study as an important set of developmental skills, and because of its potential to inform explanations of oral language acquisition. Unfortunately, studies focusing on the comprehension skills of African American children have been narrowly defined, addressing the impact of AAE compared to SAE renderings of the same sentences on children's comprehension (Isaacs, 1996; Skupas & Tweney, 1979; Torrey, 1983). Investigations of this type reveal that African American children may respond significantly better to AAE probes than SAE alternatives. Other aspects of comprehension have been neglected.

The purpose of this paper is to examine the comprehension abilities of young African American children on two tasks designed to assess syntactic-semantic comprehension abilities. The first task examined the children's ability to respond to questions. Appropriate responses to questions are acquired developmentally (Chapman, 1988; Parnell, Patterson, & Harding, 1984;

Tyack & Ingram, 1977) and responding as though a different question has been asked offers potential insight into the child's comprehension status. Connected speech sampling is a difficult context for examination of a child's ability to comprehend questions as not all types of questions may arise in natural discourse nor lend themselves to specific props. Question types vary in cognitive demand, from fairly simple requests to identify and label ("What that?"), to those probing causality ("Why___?"), manner ("How___?"), and temporal relationships ("When___?"). Close examination of response errors can help determine which aspects of the semantic-pragmatic demands are problematic for the child. In particular, the child may have difficulty with the request function itself, the specific concepts under development, or the preciseness of the vocabulary required to respond appropriately. For the purposes of this investigation, therefore, a comprehension task was developed to probe responses to requests for a range of information, but during picture description.

A second comprehension task explored children's word order strategies for comprehension of active and passive sentence constructions. Children acquiring SAE comprehend active sentence constructions in the form of agent-action-patient, using word order cues, at around 4 years of age (Beilin & Sack, 1975; Bever, 1970; Chapman, 1988; Fraser, Bellugi, & Brown, 1963; Maratsos, Kuczaj, Fox, & Chalkley, 1979; Slobin, 1973), but they may overgeneralize this strategy so that passive sentence constructions are misunderstood. Comprehension of the passive can provide evidence of syntactic understanding, in addition to the semantic processing of the sentence.

Both of these tasks have the potential to show differences by chronological age, an important attribute when the characterization of a child's developmental status is the ultimate goal. It will be important, however, also to understand the development of comprehension skills in the larger context of the child's overall process of language acquisition. Increasing comprehension skills may correspond to increasing oral language skills, or may reflect more autonomous progress. In addition, the child with greater cognitive skills may have an advantage in understanding language probes, or in contrast, the sentence level processing demanded in the current tasks may tap more independent skills. An understanding of performances on these comprehension tasks within the context of a larger set of developing skills, and of the extent to which they contribute new information or overlap information obtained from other domains, should help clinicians and educators determine normal expectations for young African American children.

The lack of language developmental information has become increasingly troublesome with accumulating

evidence that instruments developed for mainstream children are not revealing for African American children. The need for new formal and informal assessment instruments appropriate to the African American child is pressing. Prevailing views are that major language tests are culturally biased and should not be used for distinguishing children with atypical language from those with typical language skills when the children are dialect users (Adler & Birdsong, 1983; Stockman, 1986; Terrell & Terrell, 1993; Vaughn-Cooke, 1986; Washington, 1996). Understanding normal language acquisition by African American children is fundamental to making progress on these issues. The study of comprehension development will make an important contribution to our information on normal development and ultimately to these clinical concerns.

In order to begin to explore the comprehension abilities of young African American children, the following questions were posed.

1. Are there systematic relationships between performances by young African American children on a Response to Wh-Questions task, an active/passive sentence task, and chronological age or grade?

2. How do performances on these tasks relate to other aspects of the child's development, more specifically to major aspects of oral production and cognitive abilities?

Method
Subjects

The subjects were 63 African American boys (n = 31) and girls (n = 32) living in the metropolitan Detroit area (see Table 1). African Americans constituted more than 75% of the children enrolled in the participating school districts. Children who met the subject selection criteria for this investigation were selected on a continuous enrollment basis from our larger research program centered in the Detroit area. The data collection for the larger project is ongoing and will include African American children from both low- and middle-income homes, a broader age span, and both typically developing children and those with language impairments.

Table 1. Numbers of male and female subjects by grade.

Grade	Males	Females	Totals
Preschool	6	6	12
Kindergarten	17	15	32
First	8	11	19
Total	31	32	63

The subject selection criteria for this particular investigation included typically developing language, middle socioeconomic status, and chronological ages from 4 through 7 years. In the absence of culturally fair assessment instruments for determining normal-language status in this population, teacher judgment and school history were used to rule out children with language disorders. Accordingly, all of the children in this investigation were apparently normally developing, based upon teacher and parent reports, and had no history of referral for, nor enrollment in, special education services. In addition, each child passed a bilateral hearing screening at 25 dB for 500, 1000, and 4,000 Hz (ANSI, 1978) prior to data collection.

Middle socioeconomic status was determined from the demographics of the communities and the subjects' eligibility or ineligibility to participate in the federally funded free or reduced-price lunch program for low-income families. The federal guideline for participation is an annual income between 135% and 185% of the federally established poverty level for a family of four during the year in which data were collected. The subjects lived in school districts where 80% of the families were *not* eligible for the free or reduced-price lunch programs and none of these subjects was participating in these programs.

Sixty-four subjects were identified in this way for potential participation. One child, a preschooler, was unable to attend to the tasks, however, and subsequently was dropped from the study. The remaining 63 children comprised the subject sample for this investigation.

The children were enrolled in preschool (n = 12), kindergarten (n = 32), or first grade (n = 19). They ranged in age from 48 to 86 months.

Data Collection and Analysis

Two comprehension tasks constituted the core analyses for this investigation: a *Responses to Wh-Questions task*, and a *Reversible Sentences task* probing active/passive sentence forms. These two tasks were selected because prior research with mainstream children has revealed that they are sensitive to developmental changes (Chapman, 1988; Parnell et al., 1984; Pinker, Lebaux, & Frost, 1987), and therefore, have potential to be informative for the purposes of this project. In addition, as recommended elsewhere (Craig, 1996), at a time when so little is known about dialect use by children, research designs would be most informative if response formats avoided any potential co-influences of being a dialect user. For comprehension, the concepts reflected by wh-question words, and the content targeted for responses, should be unrelated to the morphosyntactic structures of AAE. Similarly, the syntax reflected in the active/passive sentence distinctions should not

be dependent upon those morphemes varying in AAE features, and the response format is a nonverbal one. The potential developmental sensitivity of the tasks, and their apparent independence from dialectal forms used by children, together recommended their selection.

One cognitive task and two oral language measures were available for each subject from the data collection protocol of the larger research project. The cognitive task, the Triangles subtest of the Kaufman Assessment Battery for Children (K-ABC), was selected as a nonverbal measure of cognition. This subtest required the subjects to assemble a set of 3–7 identical triangles to match a pictured design presented by the examiner. Triangles were chosen for use because as a simple matching task they should estimate general cognitive abilities. The larger protocol included a home sampling context for some subjects and another oral language sample, this time during a picture description task. Not all children participated in all aspects of the full data collection protocol so all components of the protocol were randomly ordered for each subject to minimize systematic effects of one activity on another during data analysis.

Comprehension Tasks

The comprehension tasks were developed specifically for this study. One, designated the *Responses to Wh-Questions task*, probed comprehension of requests for information. It seemed important not to rely on a single trial when determining which responses to specific probes were easier or harder for the children. Therefore, two activity pictures were presented, depicting barbequing and snow shoveling, picture numbers 33 and 35, respectively, selected from the Bracken Concept Development Program (Bracken, 1986), with one probe for each request type per picture. The performances of the children were not different on the two pictures, pairwise $t(62) = 1.00$, $p > .05$, based upon a comparison of their total points for each, so responses were combined for data analysis purposes.

Each picture was accompanied by 12 questions presented by the examiner using AAE. The referents for each question prompt were selected in part because they appeared related to only one possible item in the picture. Each question prompt therefore was associated with a targeted response, established a priori[1]. For SAE, these questions are acquired across the time span represented by the chronological ages of the current subjects, and presumably reflect processing of increasingly demanding concepts. Appendix A presents the 12 prompts for each picture, the targeted response, and examples of partial responses. This scoring system was developed to assign full credit, partial, or no credit as follows.

3 points (full credit) if the child produced the target response.

2 points if the child responded to the specific question word with a nonspecific label, or misnamed the referent. For example, one child responded "with a vacuum" to the question, "how he movin' the snow?" when the target was "with a snowblower."

1 point if the child responded, but to a potentially different wh-question prompt. For example, one child responded "at the park" to the "When this happening?" probe. This required that the examiner could formulate a question for which this would have been an acceptable answer.

0 points if the child produced an unrelated utterance that was not an acceptable target to any of the question probes (e.g., Adult: "when this happenin?" Child: "and the cat"), said "I don't know," or if there was no response.

This scoring system was developed to provide more than a simple correct/incorrect analysis. If a child earned fewer than the fully credited 3 points, the alternative scores attempted to rank order the closeness-of-fit of their responses to that required. Responses that reflected understanding of the semantic-pragmatic intent of the specific wh-question probe but resulted in factual inaccuracies, were distinguished from those that failed to comprehend the concept underlying the specific wh-question probe from off-task behaviors. The total possible score for this task was 72 points. The order of the pictures and the order of questions were randomly determined for each subject. The second task, designated the *Reversible Sentences task*, explored the children's word order strategies for comprehension of active and passive sentence constructions using a forced choice picture pointing task. Picture pointing tasks have been used to examine comprehension skills since the earliest stages of systematic inquiry into child language (see for example, Fraser, Bellugi, & Brown, 1963). The stimulus set of picturable agents and actions was pretested with each subject for familiarity. All subjects were readily able to point to the agents and actions when named during pretesting.

The experimental task consisted of 30 spoken prompts, three for each of 10 pairs of pictures, and they are presented in Appendix B. The order of presentation of each prompt was determined randomly for each subject.[2] The pictures were black line drawings presented

[1]Non-target responses were established from pilot data collected with 20 African American children who were the same age as the experimental subjects.

[2]Pilot testing performances with 20 African American children who were the same age and from the same schools as the experimental children were compared to performances of 20 experimental subjects. No statistically significant differences between randomized and nonrandomized trials on the Wh-question task were apparent [$t(38) = 0.35$, $p > .05$], but there was a statistically significant advantage for the Reversible Sentences task if the active trials consistently preceded the passive ones [$t(38) = 3.43$, $p = .001$].

on 5 × 7 inch cards. The prompts were constructed to form reversible sentences after the works of others, for example Roberts (1983). Ten common nouns and verbs were selected that avoid potential asymmetries in response probabilities relative to the likelihood of the reversibility of the actions occurring in real life experiences. One pair of pictures, for example, depicted a mom hugging a baby, and a baby hugging a mom.

The three spoken prompts were:

"the baby hug the mom"
(Target active voice trial),

"the mom hug the baby"
(Foil active voice trial), and

"the mom was hugged by the baby"
(Passive trial).

Simple present tense active verbs potentially use one of the phonologically weakest types of AAE, and passives are unaffected by the morphosyntactic forms of AAE. These attributes recommended the task and stimulus materials for presentation to young African American children because they minimize the potential (and largely unknown) effects of AAE on the task performances obtained, as recommended by Craig (1996).

One point was credited for each match between the spoken prompt and the child's picture selection, so that a total of 20 points was possible for the task. Pilot administrations of the task during development indicated that the presentation of more than one passive trial to the younger subjects occasionally resulted in children discontinuing the task. The passive trials were more likely to be in error for the preschoolers. In order to maintain interest in the task and to evaluate the child's use of word order cues for active sentences when prompted to respond to the passive, the two active sentence trials but only one passive trial were presented for each picture pair to each child. In order for a response to receive credit for comprehension of the passive, the passive trial and both active constructions had to be scored as correct. For analysis purposes, the two active voice trials were segmented into the targeted counterpart to the passive probe and an active voice foil. Scores on the 10 passive trials were added to those on only the 10 active voice trials that were the counterparts to the passive probes. This maximized responding but maintained a balance in the contribution of the active and passive trials to the total scores, and avoided weighting the total scores toward the correct active voice trials.

Cognition task. In addition to the comprehension tasks, a task designed to assess nonverbal cognition was administered to all subjects, the Triangles subtest of the Kaufman Assessment Battery for Children, K–ABC (Kaufman & Kaufman, 1983). Most cognitive tasks, although revealing for Anglo-American children, are not appropriate for application to African Americans (Cole, Gay, Glick, & Sharp, 1971; Kaufman, 1973). Cultural differences in cognition seem not to be due to the differential presence or absence of a process, but to good and poor correspondences between assessment situations and the function-specific practices of the culture (Cole et al., 1971; Roof, 1982).

The K–ABC fares better than other widely used tests of cognition in terms of cultural fairness. The test as a whole does not predict achievement of African Americans and Hispanics, and African Americans score lower on the complete test than Anglo-Americans (Lampley & Rust, 1986; Palmer, Olivarez, Willson, & Fordyce, 1989). Only specific subtests appear to be the problem, however, particularly the Gestalt Closure subtest (Willson, Nolan, Reynolds, & Kamphaus, 1989). The Triangles subtest contains only two biased items, one affecting African American and the other Anglo-American performances. For this reason, and the following additional ones, the Triangles subtest was selected for administration to our African American subjects: (a) as an abstract shape, triangles are visually simple, and appropriate, therefore, for our young subject ages; (b) response adequacy depends less upon a priori concept knowledge in that the figures are present; (c) responses are nonverbal, avoiding the effects of dialect; and (d) as noted by Scribner and Cole (1981), grouping objects by class membership is a more generalized cognitive skill, and, therefore, appears relatively free of the effects of culture-specific literacy.

Triangles examines a child's ability to assemble identical triangles, to match abstract designs, and evaluates mental processing and nonverbal concept formation using the visual-motor modality. It is appropriate for children ages 4 to 12 1/2 years, and, using a correct/incorrect scoring system, yields scaled scores.

Oral Language Production task. A key component of the larger research project is the collection and analysis of a spontaneous oral language sample from each child during freeplay with an unacquainted examiner. The examiners were one of the larger project's team of six African American females experienced in testing children, who spoke AAE to the children during the samplings. Each child wore an individual microphone, and the samples were audiorecorded using a microphone mixer in a quiet room in the child's school. Freeplay samples were approximately 15 to 20 minutes in duration.

The play contexts were created by allowing each child to select one toy set from three action toy sets composed of Barbie and Ken dolls with a Burger King play set, action figures and props, and the Fisher-Price School. This presentation procedure attempted to standardize interest levels in the play materials across the subjects, and the toys were successful in maintaining the interest of all children for the duration of the language sampling.

The language samples were transcribed orthographically using the CHAT conventions of the Children's Data Exchange System, CHILDES, (MacWhinney, 1994). The transcripts were segmented into Communication Units (C-Units) using the scoring criteria of Loban (1976). The first 50 wholly intelligible C-units were segmented into secondary CHAT files for analysis using the CLAN programs of CHILDES. A more detailed discussion of the C-unit segmentation scoring is provided in Craig, Washington, and Thompson-Porter (1998).

The 50 C-unit files were scored for the occurrence of one or more types of AAE, using Washington and Craig's (1994) definitions developed for LSES children that more recently have been applied to the discourse of middle-class African American children (Washington & Craig, in press). The frequency commands (FREQ) of CLAN were used to compute total numbers of occurrences of these dialectal forms in the children's 50 C-unit files. The *M* Length of Turn command (MLT) of the CLAN programs yielded average C-unit lengths in words for each subject's 50 C-unit file as well.

Reliability

A number of reliability checks were applied to the data analyses. Ten administrations of the Wh-questions task were rescored by an independent observer. The percentage of agreement between coders was high, 98% for the points assigned by each, when the number of points in agreement were divided by the number of agreements plus disagreements.

Eight of the Reversible Sentences task and eight of the Triangles subtest of the K–ABC were scored independently by a second observer at the time of their administration as an online check for scoring adequacy. Scoring agreements were high at 99% agreement for Reversible Sentences and 98% agreement for assignment of raw scores on the Triangles task.

Average sentence lengths for freeplay and the coding of AAE are two core analyses performed for every child participating in the larger research program. As a result, a transcription reliability at the level of the morpheme, a C-unit segmentation reliability for the first 50 C-units of every file, and a coding reliability for AAE for every 50 C-unit file have been developed for all subjects. Reliabilities are calculated as the number of agreements divided by the number of agreements plus disagreements. Transcription, segmentation, and coding reliabilities are very good across all subject files. For the subjects included in this investigation, transcription and segmentation reliabilities were high, 91% and 99%, respectively. Coding agreements for types of AAE were 97% and for their frequencies of occurrence 82%.

Results

Table 2 presents the means and standard deviations for responses obtained on the two comprehension tasks. Two-way Analyses of Variance were computed for each measure, by grade (3 levels) and gender (2 levels).

Wh-Question Analyses

For the children's responses to Wh-questions, there was no statistically significant interaction between grade and gender, $F(2, 57) = 1.24$, $p > .05$, and no statistically significant main effect for gender, $F(1, 57) = 0.03$, $p > .05$. However, there was a statistically significant main effect for grade, $F(2, 57) = 12.79$, $p = .000$. A Tukey Honestly Significant Difference (HSD) post hoc analysis revealed that the number of points received by the First graders was significantly greater than the Kindergartners ($p < .05$) and Preschoolers ($p < .05$), but the Kindergartners and Preschoolers were not significantly different from each other ($p > .05$). See Table 2.

An analysis of the children's errors revealed that it was rare, less than once per child on average, for any of the children, at any grade, to fail to respond to the question probes or to ask an unrelated question. A pairwise comparison revealed that the low frequency of this error type was significantly less than factual errors, $t(62) = 5.52$, $p < .0001$, which were the next most frequent error type ($M = 2.57$). See Table 3. These findings indicate that the Wh-question task was successful in eliciting response behaviors from children at these grades.

A pairwise comparison, $t(62) = 8.55$, $p < .0001$, indicated that errors in the form of a response to a different Wh-question probe occurred significantly more often ($M = 6.05$) than factual errors ($M = 2.57$). Factual errors decreased significantly across grade levels so that the

Table 2. Means (*M*) and standard deviations (*SD*) for the responses to Wh-questions (Wh-q) and Reversible Sentences (Rev) tasks

	n	Wh-q (maximum = 72 points)		Rev (maximum = 20 points)	
		M	*SD*	*M*	*SD*
Grade					
Preschool	12	53.7*	5.4	13.3*	2.6
Kindergarten	32	57.9**	5.8	13.6	3.1
First	19	63.1*,**	4.0	15.7*	2.5
Gender					
Male	31	58.5	6.3	14.2	3.5
Female	32	58.8	6.0	14.1	2.5

*First graders > preschoolers, $p \le .05$.
**First graders > kindergartners, $p \le .05$.

Table 3. Mean (*M*) frequencies and standard deviations (*SD*) of a No Response (A), a Factual Error (B), or a Response to a Different Question (C) for the three grade levels.

Grade	A		B		C	
	M	SD	M	SD	M	SD
Preschool	.92	1.00	4.33*,**	2.77	7.00**	2.92
Kindergarten	.84	1.78	2.63*	2.01	6.38	2.42
First	.32	.95	1.34**	1.34	4.89**	2.02
Combined	.70***	1.44	2.57***	2.18	6.05***	2.50

*Kindergartners < preschoolers, $p < .05$.
**First graders < preschoolers, $p < .05$.
***$p < .001$

kindergartners and first graders evidenced fewer of these error types (*M* = 2.63 and 1.34, respectively) than the preschoolers (*M* = 4.33). The frequencies of mismatching responses to specific Wh-question probes decreased significantly (*p* < .001) between the preschoolers (*M* = 7.00) and first graders (*M* = 4.89) as well. Other relationships were nonsignificant. See Table 3.

Some probes were more difficult for the children than others. Figure 1 displays the number of children in each grade who made an error on both trials of each specific question form, organized from probes eliciting the fewest errors to those with more errors. As reported quantitatively above, the preschoolers and kindergartners overall looked more like each other than did the first graders. Probes that required labeling and naming responses elicited the fewest errors and reflected a decrease in errors across increasing grade level. Requests

for information that required manner, time, or cognitively more advanced relationships like distance and frequency consistently evidenced more errors than the other probes.
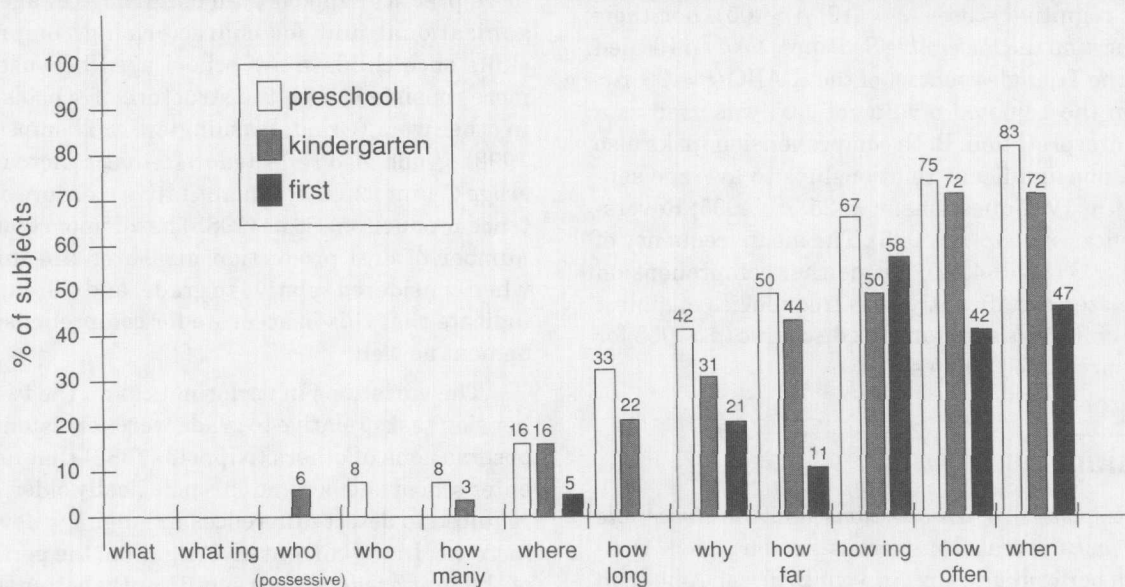
Reversible Sentence Analyses

Like the Wh-question task, the children's scores on the *Reversible Sentences task* evidenced no statistically significant interaction between grade and gender, *F*(2, 57) = 0.15, *p* > .05, and no main effects for gender, *F*(1, 57) = 0.12, *p* > .05. Statistically significant main effects for grade were again apparent, *F*(2, 57) = 4.10, *p* = .02. A Tukey HSD post hoc analysis revealed a statistically significant difference between the mean scores for the first graders and the preschoolers (*p* < .05), but other contrasts again were nonsignificant (*p* > .05). See Table 2.

An examination of the children's error responses on the *Reversible Sentences task* revealed the same grade relationship for just the passive trials as that obtained for the task's total scores. A one-way analysis of variance for frequencies of correct passive trials by grade (3 levels) revealed a statistically significant relationship, *F*(2, 57) = 5.34, *p* < .01. A Tukey-HSD post hoc analysis revealed that the number of correct passive trials was significantly greater for the first graders (*M* = 6.89) than the preschoolers (*M* = 4.41), but the Kindergartners were not different from the other two grade levels (*M* = 5.15). Further, 2 of the 3 children who responded correctly to all 10 passive trials were first graders and were 85 and 82 months of age.

Considered together, both comprehension tasks generated grade-based differences in performance, with

Figure 1. Percentage of subjects having both trials wrong.

comparable response outcomes for boys and girls. Grade effects were somewhat more pronounced for the *Responses to Wh-Questions* than the *Reversible Sentences task*.

Correlational Analyses

The relationships among scores on the two comprehension measures, and selected other developmental aspects were then examined. Table 4 reports a series of Pearson product moment correlations and their associated probability levels. Whereas nine interrelationships were examined, a conservative significance level was established for interpretation by dividing the experiment-wise p-value set at .05 by the nine statistical tests, resulting in $p \leq .005$. Using this conservative level, performances on the two comprehension tasks showed a low, positive, statistically significant correlation to each other ($r = .39, p < .002$), indicating that although there might be overlap in the processes being measured by the two tasks, the *Responses to Wh-Questions* and *Reversible Sentences* were not tapping identical skills.

The strongest relationship was between scores on the Wh-question task and chronological age ($r = .54, p < .000$). This moderately strong, positive, statistically significant correspondence indicated that the numbers of target and partial credit responses received on the *Responses to Wh-questions* task increased across the subject sample for children who were chronologically older. Reversible sentences also showed a statistically significant, positive, but considerably weaker correspondence between increases in chronological age and higher scores ($r = .35, p = .005$).

Performances on the Wh-questions task did not correlate at a statistically significant level with the children's cognitive scores ($r = .12, p > .05$), nor their performances on the *Reversible Sentences task* and scaled scores on the Triangles subtest of the K-ABC ($r = .28, p > .005$) when the adjusted p value of .005 was used as a basis for interpretation. Both comprehension tasks also evidenced nonsignificant relationships to average sentence lengths (Wh-questions: $r = .25, p > .005$; Reversible Sentences: $r = .32, p > .005$). The mean frequency of AAE was 7.03 ($SD = 4.52$), but neither comprehension task correlated significantly with frequencies of dialect use in the children's spontaneous discourse ($p > .005$ for both measures). See Table 4.

Discussion

The *Responses to Wh-questions* and the *Reversible Sentences tasks* revealed numerous similarities in comprehension performances by our young African American subjects, and some interesting differences. The ways in

Table 4. Intercorrelations among responses to Wh-questions (Wh-q), Reversible Sentences (Rev), Chronological Age, Standard Scores (SS) on the Triangles (Tri) subtest of the K-ABC, average C-unit lengths (C-MLCU), and frequencies of African American English (AAE) and their corresponding *p*-values.

	Wh-q (points)	Rev (points)	Age (months)	Tri (SS)	MLCU (words)	AAE (tokens)
Wh-q (points)	—	.39	.54	.12	.25	−.05
p		.002*	.000*	.333	.044	.685
Rev (points)		—	.35	.28	.32	−.08
p			.005*	.028	.010	.539

*$^*p \leq .005$ (.05/9).

which the tasks yielded similar performances as well as the importance of their differences are discussed below.

Performances on both the *Responses to Wh-questions* and *Reversible Sentences tasks* related positively to the chronological age of the child and to the child's grade, indicating that the skills involved in successful responding are developmental. The need for language developmental tasks appropriate to assessment of the African American child is an important one (Washington, 1996). The findings of this investigation suggest that tasks of the types described here have a place in the forthcoming assessment protocols of educators, clinicians, and researchers when examining the developmental language skills of African American children.

Moderate and low-moderate positive correlations were found between chronological ages and performances on the Wh-questions task ($r = .54$) and the *Reversible Sentences task* ($r = .35$), and systematic differences were also found by grade. Grade, however, has more practical appeal than chronological age as an organizational unit for characterizing comprehension skills once children are school aged because grade is more consistent with the structure of schools. Further, in other work (Craig, Washington, & Thompson-Porter, 1998), grade also reflected systematic increases in average C-unit lengths, a quantitative measure of oral sentence production. Scott (1988) has demonstrated that a number of oral production measures are informative when considered relative to grade, and the current data indicate that this is accurate for comprehension development as well.

The variations in performances on the two comprehension tasks relative to grade were consistent with the observations of others (Nippold, 1988) that as children enter school and become chronologically older, the spans required to detect differences in language development increase. In the current investigation, the performances of the first graders were significantly better than those of the kindergartners or preschoolers on the total scores

of the Wh-question task. In addition, the first graders performed significantly better than the preschoolers on the *Reversible Sentences task*. Significant grade differences between the preschoolers and kindergartners were not apparent. Examination of the errors on each task supported the general view that first graders demonstrated performances superior to those of the younger children, and that the younger children performed in remarkably similar ways.

Performances on both tasks showed nonsignificant relationships to the amount of dialect used by the children. This finding is consistent with a recent comparison by Isaacs (1996) of the comprehension skills of nonstandard dialect users, some of whom were African American. Although the discourse of children from middle-class homes evidences reduced amounts of dialect compared to that of their low-income peers (Washington & Craig, in press) and the subjects in this study were all middle-class, their discourse still reflected an average of seven occurrences of dialect use per 50 C-unit sample. It is unlikely, therefore, that the lack of correspondence between performance on either comprehension task and dialect frequencies was an artifact of low dialect use overall. It seems more likely that, as hoped, these tasks are free from the systematic influences of AAE. If this is the case, they particularly recommend themselves at this early stage of inquiry, when so much remains unknown about dialect development and systematic influences on its use by young African American children (see further discussion of this issue by Craig, 1996).

It would be interesting to know if the lack of a relationship between performances on these tasks and the amounts of dialect occurring in the children's samples reflects a more generalized independence of dialect and sentence comprehension. In the present investigation, these tasks were selected as a first look at sentence-level comprehension development by African American children in part because dialect influences were assumed to be minimized in the required response formats. However, the examiners did use AAE during all interactions with the children, including the comprehension probes, consistent with prevailing recommendations that a shared cultural context between examiner and child maximizes comfort level and responding (Skupas & Tweney, 1979; Terrell & Terrell, 1993; Torrey, 1983; Wheldall & Joseph 1985–6). Perhaps this was unnecessary. It will be important for future research to evaluate more specifically which assessment tasks are influenced by shared culture, and which are less dependent on shared race, dialect, and life experiences.

Task performances also showed nonsignificant relationships to average sentence lengths when interpreted relative to a conservative alpha level. However, even if a more liberal significance level had been selected, the tendency for higher comprehension scores to be obtained by children with longer C-units would reflect a very low association (Wh-questions: $r = .25$; Reversible Sentences: $r = .32$). These findings indicate that tasks involving responses to specific wh-question forms and to the active/passive voice distinction would add information to the developmental profile of an African American child and should be useful as a complement to oral production tasks. It would be informative in future research to determine how children with language disorders perform on these tasks, especially whether the relative lack of association between these comprehension tasks and gross production measures allow clinicians to distinguish African American children with circumscribed expressive language problems from those with more pervasive receptive-expressive impairments.

Lastly, no statistically significant differences were found in the performances on either task relative to gender. This finding is consistent with most recent research on gender differences in the language of children that finds few differences and then primarily for discourse behaviors (Craig & Evans, 1991; Ely, Gleason, & McCabe, 1996; Sheldon & Rohleder, 1996). For African Americans, very few studies have addressed gender differences. The extent to which discourse is marked by dialectal forms appears greater for African American males than females, whether adults or children (Washington & Craig, in press; Wolfram, 1969), but language structural behaviors fail to show systematic variations by gender when measured as average sentence lengths (Craig, Washington, & Thompson-Porter, 1998) and amounts of complex syntax (Craig & Washington, 1994). Considered together with the findings of the current investigation and consistent with the prior literature for adults and children, gender differences appear to be fairly circumscribed, and for African American children, do not appear to affect major aspects of language structural skills beyond dialect usage.

The performances of the children reflected differences on the two tasks as well as the similarities discussed previously. The tasks evidenced a low, positive, statistically significant relationship to each other, but they were not overlapping. The findings indicate that the two tasks may differ in processing demands. However, the response requirements between tasks differed in their production demands as well. The *Responses to Wh-questions* task required expression of specific targets whereas the *Reversible Sentences* task was nonverbal and could be accomplished successfully by pointing to the appropriate picture. Accordingly, a closer correspondence in the children's performances on the two tasks may have been prevented by unknown production factors. In addition, the children's performances on the two comprehension tasks did not relate significantly to

their scores on the Triangles subtest of the K–ABC, using a conservative alpha level. Even if a more liberal alpha level were employed (e.g., $p \leq .05$), the *Responses to Wh-questions* would remain unrelated to performances on the Triangles task ($p > .05$). Performances on the *Reversible Sentences* task would be considered statistically significant with a p value of .05 but the association would remain very low. These findings indicate that these sentence-level processing demands probably reflect autonomous progress from those skills measured by generalized cognitive abilities, such as grouping objects by class membership.

In summary, the findings of this investigation indicate that tasks probing responses to requests for information in the form of Wh-questions, and recognition of active and passive sentence constructions, have merit for assessing the development of comprehension skills of young African American children. Reference profiles of normal-language development are sorely needed for this population. These two comprehension tasks appear to be good candidates for the comprehension portion of protocols designed to contribute to the formulation of normal developmental expectations for language of the young African American child. Tasks like these are well established in the comprehension literature. The findings of this investigation indicate that they are appropriate for this population and that for African American children the development of unique tasks to evaluate comprehension development may not be necessary. Alternatively, comprehension tasks like these might be renormed and serve a useful role in improving our understanding of language development.

Acknowledgments

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. (1978). *Methods for manual pure-tone audiometry* (ANSI S. 321-1978). New York: ANSI.

Beilin, H., & Sack, H. (1975). The passive: Linguistic and psychological theory. In H. Beilin (Ed.), *Studies in the cognitive basis of language development* (pp. 6–83). New York: Academic Press.

Bever, T. (1970). The cognitive base for linguistic structures. In J. Hayes (Ed.), *Cognition and the development of language* (pp. 279–362). New York: Wiley.

Bracken, B. A. (1986). *Bracken concept development program*. San Antonio, TX: Psychological Corporation.

Bureau of the Census. (1990). *Statistical Abstract of the United States 1990. 110th ed.* Washington, DC: U.S. Department of Commerce.

Chapman, R. (1988). Language acquisition in the child. In N. Lass, L. McReynolds, J. Northern, & D. Yoder, *Handbook of speech-language pathology and audiology* (pp. 120–124). Toronto: B. C. Decker.

Cole, M., Gay, J., Glick, J., & Sharp, D. W. (1971). *The cultural context of learning and thinking*. New York: Basic Books.

Council for Exceptional Children. (1994). *Statistical profile of special education in the United States*. Reston, VA: Author.

Craig, H. K. (1996). The challenges of conducting language research with African American children. In A. Kamhi, K. Pollock, & J. Harris (Eds.), *Communication development and disorders in African American children* (pp. 1–17). Baltimore, MD: Paul H. Brookes Publishing Co.

Craig, H. K., & Evans, J. L. (1991). Turn exchange behaviors of children with normally developing language: The influence of gender. *Journal of Speech and Hearing Research, 34*, 866–878.

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

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

Ely, R., Gleason, J., & McCabe, A. (1996). "Why didn't you talk to your mommy honey?": Gender differences in talk about past talk. *Research on Language in Social Interaction, 29*, 7–25.

Fraser, C., Bellugi, U., & Brown, R. (1963). Control of grammar in imitation, comprehension and production. *Journal of Verbal Learning and Verbal Behavior, 2*, 121–135.

Golinkoff, R. M., & Hirsh-Pasek, K. (1995). Reinterpreting children's sentence comprehension: Toward a new framework. In P. Fletcher & B. MacWhinney (Eds.), *The handbook of child language* (pp. 420–461). Oxford: Blackwell Publishers.

Isaacs, G. J. (1996). Persistence of non-standard dialect in school-age children. *Journal of Speech and Hearing Research, 39*, 434–441.

Kaufman, A. (1973). Comparison of the performance of matched groups of black children and white children on the Wechsler Preschool and Primary Scale of Intelligence. *Journal of Consulting & Clinical Psychology, 41*(2), 186–191.

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

Lampley, D. A., & Rust, J. O. (1986). Validation of the Kaufman Assessment Battery for Children with a sample of preschool children. *Psychology in the Schools, 23*(2), 131–137.

Loban, W. (1976). *Language development: Kindergarten through grade twelve.* Urbana, IL: National Council of Teachers of English.

MacWhinney, B. (1994). *The CHILDES project: Tools for analyzing talk* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Maratsos, M., Kuczaj, S., Fox, D., & Chalkley, M. (1979). Some empirical studies in the acquisition of transformational relations: Passives, negatives, and the past tense. In A. Collins (Ed.), *Children's language and communication* (pp. 1–46). Hillsdale, NJ: Erlbaum.

Nippold, M. A. (1988). The literate lexicon. In M. A. Nippold (Ed.), *Later language development: Ages nine through nineteen* (pp. 29–47). Boston: College-Hill.

Palmer, D. J., Olivarez, A., Willson, V. L., & Fordyce, T. (1989). Ethnicity and language dominance: Influence on the prediction of achievement based on intelligence test scores in nonreferred and referred samples. *Learning Disabilities Quarterly, 12*(4), 261–274.

Parnell, M., Patterson, S., & Harding, M. (1984). Answers to Wh- questions: A developmental study. *Journal of Speech and Hearing Research, 27,* 297–305.

Pinker, S., Lebaux, D., & Frost, L. (1987). Productivity and constraints in the acquisition of the passive. *Cognition, 26,* 195–267.

Roberts, K. (1983). Comprehension and production of word order in Stage I. *Child Development, 54,* 443–449.

Roof, B. (1982). Integrating context and cognitive development. In M. E. Lamb & A. L. Brown (Eds.), *Advances in developmental psychology* (Vol. 2, pp. 125–170). Hillsdale, NJ: Erlbaum.

Scott, C. M. (1988). Spoken and written syntax. In M. A. Nippold (Ed.), *Later language development: Ages nine through nineteen* (pp. 49–95). Boston: College-Hill.

Scribner, S., & Cole, M. (1981). *The psychology of literacy.* Cambridge, MA: Harvard University Press.

Sheldon, A., & Rohleder, L. (1996). Sharing the same word, telling different stories: Gender differences in co-constructed pretend narratives. In D. I. Slobin, J. Gerhardt, A. Kuratzis, & J. Guo (Eds.), *Social interaction, social context, and language* (pp. 613–632). Mahwah, NJ: Lawrence Erlbaum Associates.

Skupas, R., & Tweney, R. (1979). Race, social class, and response bias effects in grammatical processing. *Journal of Psycholinguistic Research, 8,* 73–82.

Slobin, D. (1973). Cognitive prerequisites for the development of grammar. In L. Bloom & M. Lahey (Eds.), *Studies of child language development* (pp. 175–208). New York: Holt, Rinehart & Winston.

Stockman, I. J. (1986). Language acquisition in culturally diverse populations: The black child as a case study. In O. L. Taylor (Ed.), *Nature of communication disorders in culturally and linguistically diverse populations* (pp. 117–155). San Diego, CA: College-Hill Press, Inc.

Terrell, S. L., & Terrell, F. (1993). African-American cultures. In D. E. Battle (Ed.), *Communication disorders in multicultural populations* (pp. 3–37). Stoneham, MA: Butterworth-Heinemann.

Torrey, J. (1983). Black children's knowledge of Standard English. *American Educational Research Journal, 20,* 627–643.

Tyack, D., & Ingram, D. (1977). Children's production and comprehension of questions. *Journal of Child Language, 4,* 211–224.

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.

Washington, J. (1996). Issues in assessing the language abilities of African American children. In A. Kamhi, K. Pollock, & J. Harris (Eds.), *Communication development and disorders in African American children* (pp. 35–54). Baltimore, MD: Paul H. Brookes Publishing Co.

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

Washington, J., & Craig, H. (1994). Dialectal forms during discourse of poor, urban, African-American preschoolers. *Journal of Speech and Hearing Research, 37,* 816–823.

Washington, J., & Craig, H. (in press). *Socioeconomic status and gender influences on children's dialectal variations.*

Washington, J., Craig, H., & Kushmaul, A. (1998). *Picture description as a context for eliciting African American English.* (Manuscript submitted for publication.)

Wheldall, K., & Joseph, R. (1985-86). Young Black children's sentence comprehension skills: A comparison of performance in Standard English and Jamaican Creole. *First Language, 6,* 149–154.

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

Wolfram, W. A. (1969). *A sociolinguistic description of Detroit Negro speech.* Washington, DC: Center for Applied Linguistics.

Appendix A. Responses to Wh-Questions scoring instructions.

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: Subject responds, but provides answer to a potentially different question.

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

A. Picture: Shoveling Snow

Prompt	Response	Score
What this? (points to snowman)	a snowman	3
	snow	2
	he holdin' a shovel	1
What he doing? (points to man shoveling)	shoveling, shoveling the snow	3
	throwing the snow	3
	digging, picking up snow	2
	when it's snow outside	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	2
	in the garage	1
Who this? (points to the mother)	the mother, the mama	3
	the woman, lady, girl	2
	in the house, wavin'	1
How many cars in the picture?	two	3
	one, ten (any other number)	2
	they driving, here	1
Where this? (points to garage)	garage, in the garage	3
	in the driveway/street	2
	it in there, car	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
Why he shovelin' here? (points to man shoveling)	because the car is stuck	3
	to get under the car	2
	that the shovel, the man	1
How far he goin'? (points to man with snowblower)	to end of the driveway	3
	little far away, short	2
	all the way gettin' the snow	1
How he movin' the snow? (points to man with snowblower)	with a snowblower	3
	with a lawnmower/vacuum cleaner	2
	he tryin' to drive the car	1

How often they hafta shovel? (points to people)	every time it snows	3
	five, seven (any number)	2
When this happenin'? (points to picture)	winter, in the wintertime	3
	Christmas, now	2
	five	1

B. Picture: Barbeque

Prompt	Response	Score
What this? (points to kite)	a kite	3
	a thing that fly in the sky	2
	they boy lookin' up	1
What he doin'? (points to boy with camera)	takin' a picture	3
	lookin' through a hole	2
	that's the camera	1
Who ball this? (points to ball)	the boy's, his (while pointing)	3
	the baby's, grandma's	2
	up in the sky	1
Who this? (points to grandma)	the grandma, a grandma	3
	a lady	2
	she standin' up	1
How many glasses on the table?	two	3
	three, one (any number)	2
	that the table	1
Where this? (points to ball)	up in the air, sky	3
	up, right there (pointing)	2
	a ball, the boy	1
How long it take to cook the food? (points to man near grill)	thirty minutes, a long time	3
	ten, two, long	2
	they cookin' the food	1
Why he standin' here? (points to grandfather)	cause he's gettin' his picture tooken	3
	so he can look at the baby	2
	the baby	1
How far he throwin'? (points to boy with ball)	real far, far away, 4 feet	3
	a little, long	2
	the ball	1
How he cookin' the food? (points to man near grill)	he's barbequing with a grill	3
	with a pan	2
	hamburgers, carrots	1
How often they barbeque? (points to people)	always in the summertime, a lot	3
	ten, seven (any number)	2
	at a park, picnic	1
When this happenin'? (points to picture)	in the summertime, on a hot day	3
	now	2
	at a picnic	1

Appendix B. The 30 prompts for the Reversible Sentences Task.

Only the first active voice trial in each set was used to calculate the total score although all trials were presented. The order of presentation of each trial was randomized, as were the 10 sets of trials.

1. The boy hit the car
   The car hit the boy
   The car was hit by the boy

2. The dad touch the baby
   The baby touch the dad
   The baby was touched by the dad

3. The boy pull the girl
   The girl pull the boy
   The girl was pulled by the boy

4. The dog pat the cat
   The cat pat the dog
   The cat was patted by the dog

5. The bird bite the fish
   The fish bite the bird
   The fish was bit by the bird

6. The woman help the man
   The man help the woman
   The man was helped by the woman

7. The girl push the boy
   The boy push the girl
   The boy was pushed by the girl

8. The dog chase the cat
   The cat chase the dog
   The cat was chased by the dog

9. The mom hug the baby
   The baby hug the mom
   The baby was hugged by the mom

10. The bird splash the fish
    The fish splash the bird
    The fish was splashed by the bird