ELSEVIER

Journal of Communication Disorders
34 (2001) 255–272

Journal of Communication Disorders

# Differentiating dialect from disorder A comparison of two processing tasks and a standardized language test

Rachel K. Rodekohr, William O. Haynes*

*Department of Communication Disorders, 1199 Haley Center, Auburn University, Auburn, AL 36849, USA*

Received 10 August 2000; received in revised form 9 January 2001; accepted 4 February 2001

**Abstract**

Previous research has indicated that norm-referenced language assessment protocols are often biased against dialectal speakers. Recently, the use of processing tasks has emerged as one possible means of reducing this bias in language testing. Processing tasks measure a child's ability to process and manipulate language rather than tap previous linguistic knowledge. The present study utilized 40 subjects between the ages of 7;0 and 7;3 in the following equal groupings: White normal language, White language impaired, African American normal language, African American language impaired. The subjects were administered the Test of Language Development-2P (TOLD-2P), the Nonword Repetition Task (NRT), and the Competing Language Processing Task (CLPT). Results indicated that all three measures differentiated normal-language and language-impaired subjects from one another. With regard to cultural group, confirmed speakers of African American English (AAE) with normal language scored significantly lower on the TOLD-2P compared to White normal-language subjects. Scores of the AAE-speaking subjects with normal language on the NRT and CLPT, however, did not differ significantly from the White normal-language subjects. These results suggest that AAE speakers with normally developing language (LN) may be at a disadvantage on tests of prior language knowledge and that processing tasks may be a useful tool in combination with other assessment measures to make less biased clinical decisions. Educational objectives: As a result of this activity, the reader will (1) be able to determine the utility of processing tasks in culturally unbiased language assessment. (2) The reader will be able to discriminate the difference

* Corresponding author. Tel.: +1-334-844-9600; fax: +1-334-844-4585.
*E-mail address*: haynewo@auburn.edu (W.O. Haynes).


PII: S0021-9924(01)00050-8

between the results of a standardized language test and processing tasks on speakers of AAE. 

*Keywords:* Culturally unbiased language assessment; African American English; Processing tasks; Dialect

## 1. Introduction

Research into the area of African American English (AAE) has led to a heightened awareness of the need for culturally unbiased assessment procedures. This awareness is based on the knowledge that AAE is a legitimate dialect with roots that can be traced back to African languages, coupled with the knowledge that most standardized language tests do not reflect the validity of AAE in their scoring methods. The result of this discrepancy is the possible misdiagnosis of normally developing African American children as language impaired when, in reality, they may just be speaking AAE. The American Speech–Language–Hearing Association's (ASHA) position on this issue has stated that it is important to, "... distinguish between those aspects of linguistic variation that represent the diversity of the English language from those that represent speech, language and hearing disorders" (Battle et al., 1983, p. 24).

Attempts at distinguishing between diversity and disorder have, for the most part, met with many challenges. Although options ranging from developing revised standardized test norms to language sample analysis have been examined, even a decade ago, it was clear that most assessment alternatives have myriad disadvantages (Bailey & Harbin, 1980; Cole & Taylor, 1990; Taylor, Payne, & Anderson, 1987; Vaughn-Cooke, 1983). Research has continued in the area of differentiating dialectal variation from disorder, and the most recent attempts have involved improved standardization (Washington & Craig, 1999), dynamic assessment (Fagundes, Haynes, Haak, & Moran, 1998; Lidz & Pena, 1996), and detailed linguistic analyses (Craig & Washington, 2000; Seymour, Bland-Stewart, & Green, 1998; Stockman, 1996). While all of these methods have shown promise in differentiating dialectal variation from disorder, many of the techniques are still in the development stage, some have remaining deficiencies and others are so time intensive that routine clinical implementation would be difficult.

One area that has recently been explored in culturally nonbiased language assessment is the use of certain "processing-dependent" measures (Campbell, Dollaghan, Needleman, & Janowsky, 1997). According to Campbell et al., unlike traditional standardized tests, processing-dependent measures assess language through means that do not rely heavily on the subject's prior knowledge of events, vocabulary, or language structure. Rather, they test the subject's ability to learn and process new information. Because all of the subjects performing a

processing-dependent task are equally familiar with the words/ideas used in the test, the playing field is leveled somewhat for those children who do not have the advantage of intimate knowledge of the culture that produced the test. Repeating novel phonetic nonword sequences, for example, would not involve prior exposure to vocabulary. It would, however, involve metalinguistic ability and some attentional and memory skills.

One study of processing-dependent measures (Campbell et al., 1997) involved testing the language of groups the authors called "majority" and "minority" children using three processing measures and one traditional knowledge-based test. The subjects were 156 boys with a mean age of 12.6 (range 11–14 years). It is important to note that all of the subjects in this study were normally developing in terms of language ability. Racially, the subjects consisted of 31% majority and 69% minority students, 67% of whom were African American. No mention was made in this study regarding the use of AAE by the subjects. We cannot, however, assume that membership in the African American culture insures that the person will be a dialect speaker. All subjects were administered one knowledge-dependent test (the Oral Language Scale from the Woodcock Language Proficiency Battery-Revised, Woodcock, 1991). This test includes five subtests (memory for sentences, picture vocabulary, oral vocabulary, listening comprehension, and verbal analogies), all of which are strongly dependent on specific language knowledge. One of the three processing tasks administered was the Nonword Repetition Task (NRT), which involved repeating nonsense words of one, two, three, and four syllable lengths (Campbell et al., 1997). According to Campbell et al. (1997, p. 521), the NRT "... was designed to evaluate phonological working memory storage." The second processing task was the Competing Language Processing Task (CLPT), which taps the subject's ability to process and store information simultaneously. Based on a model by Daneman and Carpenter (1980) known as the Reading Span Task, this procedure involves judging the veracity of a series of sentences while holding the last word of each sentence in memory. The sentences are presented in six groups that represent levels of increasing difficulty. The third processing task was the shortened form of the Revised Token Test (Arvedson, McNeil, & West, 1985). It was based on activities involving the manipulation of colored geometric shapes. Before beginning the test, subjects are familiarized with all of the vocabulary that would be used so that participants are relatively similar in their exposure to the items. The findings of the Campbell et al. (1997) study supported the notion that processing measures were less biased than the knowledge-based measure in assessing language. When comparing minority vs. majority scores on the three processing tasks, there was very little difference between the two groups. However, minorities scored much lower than majority children on the knowledge-based test.

A second study of processing dependent measures (Dollaghan & Campbell, 1998) examined the ability of the NRT to differentiate children with language impairment (LI) from those with normal language. Subjects in this experiment

ranged in age from 6.0 to 9.9, and were divided into two racially mixed groups of 20 children. One of these groups contained children with LI (12 African American, 5 White, and 3 "other"). The second group was made up of 13 African American children, 4 White, and 3 "other" all of whom had normal-language abilities. Again, the African American children who participated in this study were not confirmed to be speakers of AAE. The procedure involved administration of the NRT to both groups of children. Several knowledge-based tests of language were also administered (Test of Language Development-Primary-2; Peabody Picture Vocabulary Test-Revised; Test of Nonverbal Intelligence-Revised). Both the processing task and the knowledge-based tests differentiated the group with LI from the normal-language group. However, the processing task differentiated the two groups with no overlap between performances of language-impaired and normal-language subjects on the three-syllable words, four-syllable words, and on combined NRT scores. The experimenters interpreted the findings as supporting the notion that the NRT differentiated individuals with LI with a greater degree of accuracy as compared to the knowledge-based tests. It is important to note that the Dollaghan and Campbell (1998) study was not specifically designed to use processing tasks and knowledge-based tests to differentiate speakers of AAE from children with LIs. The study does, however, support the idea that processing tasks appear to be useful in determining the existence of language disorder and when coupled with the earlier study (Campbell et al., 1997) suggests exciting possibilities for research on nondiscriminatory language assessment.

Other studies have shown that the NRT can differentiate children with LI from those with normal-language abilities (Bishop, North, & Donlan, 1996; Ellis-Weismer et al., 2000). The study by Ellis-Weismer et al. (2000) included 581 subjects in a population based sample of school-age children. The average age of the subjects in this sample was 95 months and included 493 Caucasian and 74 African American children. Again, this study did not confirm whether or not the subjects were dialectal speakers nor did the study specifically address a comparison of language impaired groups vs. normal-language groups from the different cultures. Indeed, the authors state: "The present study used a population-based sample of children in which presence and absence of language problems were not equally represented and children from majority cultural backgrounds far outnumbered minority children" (p. 875). The results of this study demonstrated that children with LI "exhibited deficient nonword repetition skills compared to normal-language controls" (p. 865) and suggested that the NRT was a "culturally nonbiased measure of language processing."

The present study was designed to further examine the use of the NRT in comparison to a knowledge-based language test in differentiating AAE speakers from those with LI. Several important variables have been considered in the present study. First, all African American subjects in the present study were confirmed to be speakers of AAE through analysis of a brief conversational

language sample. Prior research involving processing tasks had not controlled for this factor. Second, the present study used two different processing tasks, the NRT and the CLPT, in an effort to determine if one of these tasks is more efficient than the other in differentiating dialect from disorder. Both of these processing tasks were effective in the Campbell et al. (1997) study of normal-language majority and minority subjects, however, the CLPT was not studied in the Dollaghan and Campbell (1998) experiment with subjects having LI. Third, the processing task study specifically designed to differentiate between "majority and minority" groups (Campbell et al., 1997) used subjects ranging in age from 11 to 14 years. The present study involved 7-year-old children since it is more likely that clinicians will have to distinguish between dialect and disorder at a younger age. The present study was designed to determine (1) whether a knowledge-based test (TOLD-2P) and processing tasks (NRT; CLPT) can differentiate between children with normally developing language (LN) and those with language disorder and (2) to determine the performance of AAE speakers with normal-language skills on the two types of assessment tasks.

## 2. Methodology

### *2.1. Subjects*

The study utilized 40 subjects divided into four groups. The four groups were matched as closely as possible for age and sex. The mean age for the four groups was 7.2 (range 7.0–7.3). Two groups were African American (10 children with LI children and 10 children with normal language who spoke AAE). The other two groups were White (10 children with LI and 10 children with normal language who spoke Southern English). As in Dollaghan and Campbell (1998), subjects with LI were targeted through their enrollment in language therapy in the public school system rather than exclusively through specific standardized test scores. This use of intervention status was referred to as "the gold standard" for identifying LI subjects by Dollaghan and Campbell. Dollaghan and Campbell used enrollment in treatment as a primary selection criterion for several reasons. First, standardized tests may not be appropriate, nonbiased tools for targeting LI children in multicultural populations. Second, choosing children who are enrolled in treatment as the LI group provides a number of nondiscriminatory influences to counterbalance the influence of test scores. For example it takes into account the "concern that speech–language pathologist, teachers, and parents have about an individual child's language skills" (p. 1138), because an agreement must be reached between all of these parties in addition to standardized test scores in order to enroll a child in therapy. In the present study, the standard ($Z$) score for the Spoken Language Quotient (SLQ) on the TOLD-2P for the group with LI was $-1.57$ (S.D. = 0.60) and the $Z$ score for the group with normal language was $-0.05$ (S.D. = 0.81). When subjected to a $t$ test, these means were significantly

different ($t = 6.66$, $df = 38$, $P < .0001$). These mean standard scores on the TOLD-2P were highly similar to those (LI = – 1.59; NL = – 0.02) reported by Dollaghan and Campbell. The finding that the TOLD-2P scores differed significantly does not confirm that all of the subjects in the language disordered group had impairments. Rather, enrollment in treatment was used in the present study and Dollaghan and Campbell to be the primary criterion for the existence of impairment. The 20 subjects with LN were selected from a pool of students who returned informed consents based on age and gender restrictions dictated by the 20 subjects with LI.

A short (5 min) conversational language sample was gathered from each African American subject. The purpose of the sample was to confirm the use of AAE in these children and the specific features used are listed in Table 1 following a similar procedure implemented by Washington and Craig (1998). While this was only a brief sample, it was long enough to document the use of many features of AAE in the African American subjects. Some caveats would be in order regarding this point. First, the sample was small and may not have detected certain features. Second, as would be predicted in any group, there was considerable variability in the use of AAE features in the subjects that could have been the result of style shifting with a White experimenter or genuine use of the acrolect of AAE. Finally, Southern Dialect and AAE have some features that overlap (e.g., allophones of /r/ being substituted for by a schwa vowel) and since the African American subjects lived in the South, they may exhibit a particular feature due to dialect importation. On the other hand, some features (e.g., use of "be" as habitual, final cluster reductions, substitution of f/th, omission of copula/auxiliary) are peculiar to AAE. Table 1 shows that most African American subjects exhibited some of the nonoverlapping features of AAE.

### *2.2. Materials/equipment*

The procedure used to develop the stimulus tape for the NRT (Appendix B) was adapted from Dollaghan and Campbell (1998). The stimulus tape was recorded through a head mounted microphone in a soundproof booth. As in the Dollaghan and Campbell study, the stimulus nonwords were "spoken by a trained adult female speaker ... [who has] previously practiced producing each nonword at a consistent rate, assigning primary stress to the second syllable of the four-syllable nonwords, and to the first syllable of all others."(p. 1139). The nonwords are listed in Appendix C. These 16 groups of phonemes are the same ones used in the Dollaghan and Campbell study and were developed in such a way that articulation errors and vocabulary knowledge should not interfere with performance. For instance, the arrangement of the phonemes avoided consonant clusters, later-developing phonemes, and lax vowels (which may be reduced to schwa). They were also arranged in such a way that the CVC combinations of the syllables did not form true words. After being recorded, these nonwords were transcribed by

Table 1
Frequency of use of selected features of AAE as noted in 5-min language samples from 20 African American children

| | African American subjects | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AAE forms | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| Regular past tense deletion | 1 | | 1 | | | | 2 | 5 | 2 | | | | 2 | | | 4 | | | | |
| Irregular verb, seen/saw | 1 | | | | | | 1 | | | 1 | | | | | | 1 | | | | |
| Third person -s deletion | | | | 1 | | 3 | 2 | 1 | 1 | 2 | 1 | | 1 | | | 2 | | | | 1 |
| Future verb tense "I'ma goin" | | | 1 | | | | | | | | | | | | | | | | 3 | 1 |
| Possessive -s deletion | | | | | | | 1 | | | 1 | | | | | | | | | | |
| Plural -s deletion | 3 | 3 | | 1 | | 2 | 1 | | | 33 | | | | | | 2 | | | | |
| Double negatives | 1 | | | | | | | | | | | | | 2 | 1 | | | | | |
| Regularized reflexive Pronoun (hisself) | | | 1 | | | | | | | | | | | | | | | | | |
| Pronominal apposition "My brother he big" | | | 1 | 1 | | | | 1 | | | | | | 2 | | | | | | |
| Have/has and do/does | | | | | 2 | | | | 1 | | | | | | | | | | | 1 |
| Indirect questions "I wonder was he walking" | | | | | | | | | | 1 | 1 | | 1 | | | | 2 | | | |
| Deletion of be forms in contractions "He goin" | 3 | 2 | | 1 | 4 | | 1 | | 16 | 1 | | | | | | 2 | 1 | | 11 | 2 |
| Use of be habitual "He be working" | | | | | | | | | | 1 | 1 | | | | | | | | | |
| Consonant cluster reduction | 18 | 10 | 15 | 3 | 2 | 1 | 4 | 8 | 19 | 12 | 14 | 4 | 13 | 8 | 9 | 34 | 3 | 11 | 3 | 13 |
| Final consonant deletion | 6 | 14 | 2 | 5 | 1 | 15 | 11 | 9 | 12 | 11 | 2 | 1 | 5 | 13 | 3 | 3 | | | 1 | |
| /r/ → /O/, /↔/ | | 10 | | 2 | 3 | 1 | 7 | 1 | 4 | 4 | 5 | 2 | | 2 | | 3 | | 2 | 1 | 2 |
| /Π/ → /t/, /f/ | 1 | | | 1 | | | 1 | 1 | 1 | 2 | 2 | | 3 | 1 | 1 | 1 | | | | 1 |
| /δ/ → /d/, /v/ | 9 | 7 | 10 | 2 | 10 | 3 | 5 | 7 | 14 | 14 | 25 | 9 | 10 | 5 | 9 | 12 | | 15 | 2 | 26 |

two research assistants independent of each other. The agreement between the resultant transcriptions was used as a measure of reliability. An agreement of greater than 90% was required for the stimulus tape. A final measurement of the tapes was completed through the Computerized Speech Laboratory (Kay Elemetrics) (CSL: Computerized Speech Laboratory Model 4300B, 1994). In this case, the length of the stimulus was measured. Dollaghan and Campbell felt that this measurement was important in order to insure that stimuli were spoken at a consistent rate. Rates comparable to the prior research were found in the present study for one-syllable words ($M = 667$ ms; S.D. = 179 ms), two-syllable words ($M = 1.435$ ms; S.D. = 371 ms), three-syllable words ($M = 2.424$ ms; S.D. = 47 ms) and four-syllable words ($M = 2.589$; S.D. = 770 ms). The NRT stimulus tape was recorded on a Marantz PMD 101 Professional tape recorder with a Shure Prologue 14 H dynamic microphone.

The CLPT (Appendix C) stimulus tape was developed utilizing female live voice. To maintain consistency, it contained the 42 sentences that had been used in previous studies of this task (Gaulin & Campbell, 1994). As in Gaulin and Campbell (1994), the sentences were read at a speed of approximately 160 wpm; a 3–4-s pause followed each sentence. Following each group of sentences (grouped into levels of difficulty), the tape asked, ''What was the last word of the sentence?'' The subject's responses were audio-recorded on a second tape recorder (Marantz PMD 101) for later analysis.

Although a new version of the Test of Language Development, TOLD-3P, is available, the previous form of the test, TOLD-2P, was used in the present study mainly because it was the knowledge-based measure used in the earlier study by Dollaghan and Campbell (1998). In this way, our results could more easily be compared to the findings of prior research. The SLQ composite score from the TOLD-2P for each subject was calculated according to the guidelines provided in the examination manual. It is stated in the test manual (Newcomer & Hammill, 1988, p. 29), ''The SLQ is comprised of the standard scores of the seven TOLD-2P subtests. As such, it is the best, most comprehensive estimate of a persons' overall language ability.'' According to the examination manual, the TOLD-2P has adequate psychometric qualities of validity and reliability and has been used extensively in research since its original development in 1977. The examination tests picture vocabulary, oral vocabulary, grammatical understanding, sentence imitation, grammatical completion, word discrimination, and word articulation.

### *2.3. Procedure*

Hearing screening, language sampling, administration of the TOLD-2P, and processing-dependent tasks (NRT; CLPT) were all carried out in a quiet room within the subjects' school. The experimenter and subject sat perpendicular to each other at a rectangular table used for all activities. The tape recorder and a set of frequency matched Koss TD-80 earphones used to present the NRT and

CLPT were located in the center of the table for the processing-dependent activities. A second tape recorder used to record the subjects' responses to the NRT and CLPT was also located on the table during the processing tasks. The subject wore the headphones while listening to the stimulus tapes and a stand microphone was located within approximately 8–12 in. from the subject to record the child's responses.

First, all subjects underwent a hearing screening of the frequencies 500, 1000, 2000, and 4000 HZ at 25 dB ANSI using a portable audiometer. Second, a 5-min language sample was gathered for each subject interacting with the experimenter conversing about topics of interest to the child. Books, toys, and other items were provided as necessary to stimulate conversation. The samples were audio-recorded for later analysis. Third, all subjects were administered the TOLD-2P (Newcomer & Hammill, 1988).

The first processing-dependent task (NRT) was carried out in the same manner as Dollaghan and Campbell (1998). This task was administered to subjects individually by the experimenter. The tape recorder provide the following instructions, "Now I will say some made-up words. Say them after me exactly the way that I say them." After the instructions, the subject heard the four sets of nonwords, progressing from one to four syllable lengths. Each stimulus was presented one time. The subject's attempt at repeating the nonwords was recorded using an external microphone and tape recorder.

The CLPT was administered using procedures prescribed by earlier processing studies (Campbell et al., 1997; Gaulin & Campbell, 1994). The task involved subjects listening to the 42 sentences recorded on audiotape. The subjects were asked to judge the veracity of each statement while holding the last word of the statement in memory for later recall. This process was explained to the subjects through the same instructions used by Gaulin and Campbell (1994, p. 60):

> I am going to read you some true and false sentences. After each one I want you to say 'yes' or 'no.' After we have done a group of sentences I will ask you to tell me the last word of each sentence in that group. Don't worry about getting them in the right order. As we go on, the groups will have more sentences. It will get hard and you won't be able to ask any questions, but I want you to keep on trying to do the best you can. Remember to say 'yes' or 'no' to each sentence. Then, when I ask you, please say the last word of each sentence you have just heard. Do you understand? Let's try some for practice.

The groups of sentences referred to in these instructions were arranged in levels of increasing difficulty. Each level consisted of two groups, which contained increasing numbers of sentences. For instance, the two groups in Level 1 each contained one sentence; the two groups in Level 2 each contained two sentences, etc. The number of sentences contained in each group increased until the subject reached Level 6 (the two groups in Level 6 each contain six sentences). For each group of sentences, the subject is

required to make the appropriate number of true/false judgments (depending on the number of sentences in the group), and then repeat the last word of every sentence in the group.

### 2.4. Data analysis

The procedures specified by Dollaghan and Campbell (1998) for scoring the NRT were followed. Specifically:

> Each phoneme (consonant or vowel) was scored as correct or incorrect in relation to its target phoneme. Phoneme substitutions and omissions were scored as incorrect; distortions of a phoneme were scored as correct. Phoneme additions were not counted as errors . . . In those cases in which a subject did not recreate the syllable structure of the nonword (adding or omitting one or more syllables), individual phoneme scoring proceeded after aligning the syllable sequence produced by the subject as nearly as possible to that of the target, using vowels repeated as syllable anchors to maximize the subjects score.

### 2.5. Reliability

Reliability was calculated in accordance with procedures used by Dollaghan and Campbell (1998). Twenty percent of the tapes containing subjects' NRT were randomly selected and transcribed independently by a trained research assistant. A phoneme by phoneme percentage of agreement was calculated for correct judgments at each syllable level, as well as total percent correct and the average percentage of agreement was 93% (range 90–96%). Agreement on the number of words recalled on the CLPT was 100%.

## 3. Results

The dependent variables in the present study were scores on the NRT, percent of words recalled on the CLPT and the SLQ on the TOLD-2P. Data were analyzed using a series of between-subjects univariate $2 \times 2$ analyses of variance. One factor was cultural group (White/African American) and the other factor was language status (normal/impaired).

### 3.1. TOLD-2P

Table 2 shows the means and standard deviations for African American and White children with and without LIs. TOLD-2P scores were subjected to a two factor analysis of variance. The ANOVA revealed a significant main effect for race ($F = 30.99$; $df = 1, 36$; $P < .0001$) with the White subjects ($M = 94.8$; S.D. = 16.13) scoring significantly higher than the African American subjects

Table 2
Means and standard deviations (in parentheses) for SLQ composite scores on the TOLD-2P, percent words recalled on the CLPT, percent total phonemes correct on the NRT-T, percent phonemes correct on the NRT for one-, two-, three-, and four-syllable levels in normal-language and language-impaired African American and White subjects

| | African American subjects | | White subjects | |
|---|---|---|---|---|
| Variable | Language-normal | Language-impaired | Language-normal | Language-impaired |
| SLQ | 89.4 (9.66) | 71.7 (6.53) | 108.6 (6.83) | 81 (8.90) |
| CLPT | 36.3 (14.25) | 31.5 (13.48) | 48.1 (11.29) | 30.4 (16.72) |
| NRT-T | 84.7 (3.95) | 77.6 (12.36) | 85 (6.00) | 74 (10.94) |
| NRT-1 SYLL | 95 (7.12) | 90.1 (9.51) | 100 (0) | 94.1 (11.11) |
| NRT-2 SYLL | 97.2 (3.08) | 96.5 (5.29) | 99.5 (1.58) | 99.5 (1.58) |
| NRT-3 SYLL | 89.2 (6.54) | 76.7 (15.16) | 96.3 (6.39) | 81.3 (14.67) |
| NRT-4 SYLL | 72.6 (10.16) | 62.5 (20.29) | 62.1 (12.55) | 45.3 (13.52) |

($M = 80.55$; S.D. = 12.12). A significant main effect for language status was also found ($F = 78.29$; $df = 1, 36$; $P < .0001$) with scores for children with LI ($M = 76.35$; S.D. = 8.98) being significantly lower than those with LN ($M = 99$; S.D. = 11.21). No significant race by language status interaction effect was found ($F = 3.74$; $df = 1, 36$; $P = .061$).

### 3.2. Competing Language Processing Task (CLPT)

The percentage of words recalled on the CLPT was obtained using procedures outlined by Campbell and Dollaghan (1997). Table 2 depicts the means and standard deviations for the CLPT scores for subjects with LN and language disorders from African American and White groups. The results of the analysis of variance showed no main effect for race ($F = 0.4511$; $df = 1, 36$; $P = .5061$). The mean score for the African American group was 33.9 (S.D. = 13.73) and the White group was 39.25 (S.D. = 16.6). There was a significant main effect for language status ($F = 5.58$; $df = 1, 36$; $P = .024$) with the normally developing subjects ($M = 42.2$; S.D. = 13.91) scoring higher than the children with language disorders ($M = 30.95$; S.D. = 14.8). The race by language status interaction was not significant ($F = 3.55$; $df = 1, 36$; $P = .067$).

### 3.3. NRT total score (NRT-T)

The NRT requires the subject to imitate stimulus lengths ranging from one to four syllables. The total score on the NRT represents the percent phonemes correct across all four syllable levels. This measurement was also used by Campbell and Dollaghan (1997). Table 2 shows the means and standard deviations for NRT-T for African American and White subjects across language status groups. The ANOVA indicated no main effect for race ($F = 0.3378$; $df = 1, 36$; $P = .565$). Scores for the African American group averaged 81.15 (S.D. = 9.65) compared with the mean score of the white group which was

79.5 (S.D. = 10.28). There was a significant main effect for language status ($F = 10.29$; $df = 1, 36$; $P = .002$). In this case, subjects with LI ($M = 75.8$; S.D. = 11.51) scored significantly lower than subjects without LI ($M = 84.85$; S.D. = 4.94). The race by language status interaction was not significant ($F = 0.4393$; $df = 1,36$; $P = .5117$).

### 3.4. NRT: syllable level scores (NRT-S)

In order to determine if syllable length affected nonword repetition performance, an analysis of separate scores for syllables was performed. A separate percent phonemes correct was calculated for each of the four syllable lengths. Means and standard deviations for the racial groups, language groups and syllable levels are shown in Table 2.

A three-factor ANOVA was performed with race and language as between subject variables and syllable length (one to four) as a within-subjects factor. The analysis of variance showed no main effect for race ($F = 0.1719$; $df = 1, 36$; $P = .6809$) with the mean score of the African American subjects at 84.98 (S.D. = 16.01) and the mean score of the White subjects of 84.76 (S.D. = 21.41). A significant main effect was found for language status ($F = 14.05$; $df = 1, 36$; $P = .0006$). Subjects with LN ($M = 88.32$; S.D. = 14.87) scored higher than those with delayed language development ($M = 80.75$; S.D. = 21.43). There was no significant race by language status interaction effect ($F = 0.3301$; $df = 1, 36$; $P = .5692$). Results of the ANOVA revealed a significant main effect for syllable length ($F = 140.50$; $df = 3, 108$; $P < .0001$) with mean scores decreasing as syllable length increased from one syllable ($M = 94.8$; S.D. = 8.59), to two syllables ($M = 98.18$; S.D. = 3.42), to three syllables ($M = 85.88$; S.D. = 13.41), to four syllables ($M = 60.63$; S.D. = 18.65). There was also a significant race by syllable length interaction ($F = 11.48$; $df = 3, 108$; $P < .0001$) and a significant language status by syllable length interaction ($F = 6.33$; $df = 3, 108$; $P < .0001$). Post hoc testing using the Newman–Keuls procedure showed that the extremely low scoring of the White children with LI was the primary factor creating both interactions.

## 4. Discussion

The results of the present study support the notion that both knowledge-based language tests (TOLD-2P) and processing-based tasks (NRT; CLPT) can differentiate between children with language disorder and those with LN. Although, the TOLD-2P was able to discriminate between language-impaired and normal-language subjects, it was noted that children with normal language who were speaking AAE scored significantly lower on the test than their White peers. In fact, White subjects with normal-language abilities earned a mean SLQ of 109 while the African American subjects with normal language averaged 89. This discre-

pancy supports prior research which suggests that African American children who speak AAE and exhibit LN may be at a disadvantage on such standardized tests based on a Standard English model. It is important to note that all of the AAE-speaking children did not earn scores that would be viewed as signifying LI on the TOLD-2P. A mean SLQ of 89 corresponds to approximately the 25th percentile. The White subjects mean of 109 corresponds to approximately the 75th percentile. The fact that the two groups were significantly different on the ANOVA does not indicate the African American children were language disordered, only that they performed at a significantly lower level on the TOLD-2P. The present results, of course, cannot be generalized to newer editions of the TOLD or other knowledge-based language tests. The TOLD-2P, however, is similar to many other language tests that focus on comprehension, production and imitation of language forms. The important point in the present study and that of Campbell et al. (1997) is that normally developing African American subjects exhibit a larger performance gap when compared to White children on the knowledge-dependent language tests and less of a gap on processing tasks.

In contrast to the knowledge-based test, the CLPT adequately differentiated between LI and LN subjects without biasing against AAE speakers. LI subjects scored significantly lower on this task than did LN subjects, however, the scores of those subjects speaking AAE did not differ significantly from the scores of subjects speaking SAE. The African American group averaged a score of 34 on the CLPT while the White group had a mean of 39. Campbell and Dollaghan found a similar pattern in their 1997 study. Although they did not analyze the performance of LI vs. LN subjects, they did examine scores of "minority" vs. "majority" children. As in the current study, no significant difference was found between the two groups. In Campbell et al. (1997), the mean scores on the CLPT were considerably higher than those found in the present study, however, this could be due to the disparity in age groups of the children studied. Campbell et al. (1997) studied children between 11 and 14 years while the present investigation included children between 6 and 9 years. The present study also included children with language disorders who were included in the main effect means for race when language status was collapsed.

The total scores on the NRT (NRT-T) exhibited a similar pattern to that described for the CLPT. The group with LI and the normal-language group earned significantly different mean scores (LI = 75.8; LN = 84.8), however, the scores of the racial groups were highly similar (African American = 81.15; White = 79.5). The NRT-T did not reveal a pattern of scores similar to the TOLD-2P in which the normal-language subjects who spoke AAE scored significantly lower than the normal-language subjects who spoke Southern English. In fact, the mean scores for the two normal-language groups on the NRT-T were almost identical (African American = 84.7; White = 85).

Scores on the NRT at the various syllable levels showed an inverse relationship; that is, as syllable level increased from one- to four-syllable words,

NRT scores decreased. This relationship was also found in Campbell and Dollaghan (1998) where the three-syllable, four-syllable, and NRT-T scores were the most significant discriminators between children with LI and those with LN.

### *4.1. Clinical implications*

The results of the present study suggest that knowledge-based tests such as the TOLD-2P may not reflect the true language abilities of AAE speakers. The AAE dialect speakers with LN scored significantly lower (20 points) than the normally developing children in the White group on this knowledge-based test. Yet, these same normally developing groups earned highly similar scores on the two processing tasks in the present study. This study supports Campbell et al. (1997), however, it adds to this previous research by demonstrating the utility of processing tasks with documented speakers of AAE rather than "minorities" without verification of their use of dialectal variation. Because processing tasks are not able to pinpoint specific areas of deficit (e.g., linguistic errors in syntax, semantics, morphology, etc.), they cannot be used alone for diagnostic purposes. However, they may be useful as an adjunct to standardized testing and other traditional diagnostic methods. The NRT is an especially efficient screening method since it takes less than 10 min to administer and score and several investigations have now shown that it is highly sensitive to differentiating children with language disorder from those who are normally developing (Bishop et al., 1996; Dollaghan & Campbell, 1998; Ellis Weismer et al., 2000). Processing tasks are particularly useful in identifying LI because they involve metalinguistic activities which are known to be difficult for children with even mild LI (Daneman & Carpenter, 1980). Because knowledge based tests such as the TOLD-2P focus primarily on learned aspects of language, mild LI may be incorrectly attributed to use of a dialect.

Another clinical implication of the present study is in choosing a processing task for use in screening. The present study compared the CLPT and the NRT on the ability to identify language disorder in children. The results suggest that the NRT seems to hold the greatest practical value as a screening tool. First, the differences among the scores of the normally developing subjects on the CLPT were larger than on the NRT. Second, due to the complexity of the CLPT, it may not be appropriate for younger age groups. The task requires relatively extensive instruction and training when used with younger children. Finally, the NRT is far more time efficient in comparison to the CLPT.

The results of the present study should be viewed with caution for several reasons. First, the number of subjects in each group was small. Second, the groups differed in their socioeconomic levels. A majority of the White normal-language subjects came from upper middle class families, while a majority of African American and White subjects with LI came from working class families. Third, it

is possible that scoring on the TOLD-2P may have been influenced by a ceiling effect. The upper age limit for the test is 8 years, 11 months, and the present subjects and those used in Campbell et al. (1997) had mean ages of 7.2 and 7.10, respectively. Perhaps studying the performance of younger subjects may have made a difference. These differences may have had an effect on the variance in scoring of the groups. The use of processing tasks as an adjunct to traditional standardized testing appears to have significant potential in nonbiased assessment of language ability. Further research using processing tasks on children representing other cultural groups and different types of LI has the potential to increase our ability to identify children with language disorders while maintaining culturally fair practices.

## Appendix A. Continuing education

1. Standardized language tests:
   (a) are usually culturally unbiased language measures.
   (b) often penalize speakers of AAE.
   (c) are usually based on a Standard English model.
   (d) both b and c.
   (e) usually have separate norms for different cultural groups.

2. The NRT
   (a) can discriminate well between children with language disorder and those with normal language.
   (b) does not penalize speakers of AAE.
   (c) takes about one hour to administer.
   (d) both a and b.
   (e) involves making true-false judgments.

3. Earlier studies involving the NRT:
   (a) controlled for the use of dialect in the subjects studied.
   (b) did not control for dialect usage.
   (c) used mainly Hispanic subjects.
   (d) both a and c.
   (e) included some real words in their stimuli.

4. The CLPT
   (a) was used for the first time in the present study.
   (b) did not differentiate between subjects with language disorders and those with normal language.
   (c) was biased against speakers of AAE.
   (d) was not biased against speakers of AAE.
   (e) both b and c.

5. The present study recommends:
   (a) using the NRT as a diagnostic test.
   (b) using the NRT as a screening task in concert with other testing.
   (c) using standardized language tests with speakers of AAE.
   (d) using the CLPT in preference to the NRT.
   (e) using the CLPT only with younger children.

**Appendix B. NRT stimuli**

| One syllable | Two syllables | Three syllables | Four syllables |
|---|---|---|---|
| /naIb/ | /teIvak/ | /tʃinItaUb/ | /veitatʃaIdIp/ |
| /voUp/ | /tʃoUvæg/ | /naItʃoUveIb/ | /dævoUnItʃig/ |
| /taUd3/ | /vætʃalp/ | /dItaUvæb/ | /naItʃItaUvub/ |
| /dIf/ | /ncItaUf/ | /teIvItʃaIg/ | /tævatʃinaIg/ |

**Appendix C. CLPT items**

| Group | | Correct response | Child's response | Word recall |
|---|---|---|---|---|
| | Practice items | | | |
| A | Children can play. | Y | Y N | _____ |
| | Apples are black. | N | Y N | _____ |
| B | Ice is hot. | N | Y N | _____ |
| | Mice eat cheese. | Y | Y N | _____ |
| | Level 1 | | | |
| 1 | Trees have leaves. | Y | Y N | _____ |
| 2 | Trains can fly. | N | Y N | _____ |
| | Level 2 | | | |
| 1 | Pumpkins are purple. | N | Y N | _____ |
| | Busses have wheels. | Y | Y N | _____ |
| 2 | Boys can eat. | Y | Y N | _____ |
| | Bananas are blue. | N | Y N | _____ |
| | Level 3 | | | |
| 1 | Carrots can dance. | N | Y N | _____ |
| | Water is dry. | N | Y N | _____ |
| | Sugar is sweet. | Y | Y N | _____ |
| 2 | Buckets tell jokes. | N | Y N | _____ |
| | Horses have tails. | Y | Y N | _____ |
| | Milk is white. | Y | Y N | _____ |

| | | | | |
|---|---|---|---|---|
| | Level 4 | | | |
| 1 | Feathers can tickle. | Y | Y N | _____ |
| | Babies drive trucks. | N | Y N | _____ |
| | Birds can fly. | Y | Y N | _____ |
| | Cars build bridges. | N | Y N | _____ |
| 2 | Snails have shells. | Y | Y N | _____ |
| | Chairs eat cake. | N | Y N | _____ |
| | Giants are small. | N | Y N | _____ |
| | Balloons can float. | Y | Y N | _____ |
| | Level 5 | | | |
| 1 | Shoes have ears. | N | Y N | _____ |
| | Fire burns paper. | Y | Y N | _____ |
| | Robins eat worms. | Y | Y N | _____ |
| | Cars can race. | Y | Y N | _____ |
| | Hotdogs can bark. | N | Y N | _____ |
| 2 | Horses have feet. | Y | Y N | _____ |
| | Dishes can whistle. | N | Y N | _____ |
| | Fish pull wagons. | N | Y N | _____ |
| | Roses have thorns. | Y | Y N | _____ |
| | Cats can talk. | N | Y N | _____ |
| | Level 6 | | Y N | _____ |
| 1 | Apples are square. | N | Y N | _____ |
| | Rabbits read books. | N | Y N | _____ |
| | Houses can jump. | N | Y N | _____ |
| | Pencils eat candy. | N | Y N | _____ |
| | Airplanes can fly. | Y | Y N | _____ |
| | Balls are round. | Y | Y N | _____ |
| 2 | Fish can swim. | Y | Y N | _____ |
| | Clouds wear slippers. | Y | Y N | _____ |
| | Sheep eat lions. | N | Y N | _____ |
| | People have eyes. | Y | Y N | _____ |
| | Dogs can run. | Y | Y N | _____ |
| | Lemons are yellow. | Y | Y N | _____ |

Percentage correct: True/False _____

Percentage correct: Word Recall Score _____

## References

Arvedson, J., McNeil, M., & West, T. (1985). Prediction of revised token test overall, subtest, and linguistic unit scores by two shortened versions. *Clinical Aphasiology*, *15*, 57–63.

Bailey, D., & Harbin, G. (1980). Nondiscriminatory evaluation. *Exceptional Children*, *46*, 590–596.

Battle, D., Aides, M., Grantham, R., Halfond, M., Harris, G., Morgenstern-Lopez, N., Smith, G., Terrell, S., & Cole, P. (1983). Position paper on social dialects. *Journal of the American Speech Language Hearing Association*, *25* (9), 23–24.

Bishop, D., North, T., & Donlan, C. (1996). Nonword repetition as a behavioral marker for inherited language impairment: evidence from a twin study. *Journal of Child Psychology and Psychiatry*, *36*, 1–13.

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

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

Craig, H., & Washington, J. (2000). An assessment battery for identifying language impairments in African American children. *Journal of Speech, Language and Hearing Research*, *43* (2), 366–379.

CSL: Computerized Speech Laboratory Model 4300B (1994). Pine Brook, NJ: Kay Elemetrics.

Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466.

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

Ellis-Weismer, S., Tomblin, J., Zhang, X., Buckwalter, P., Chynoweth, J., & Jones, M. (2000). Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech, Language and Hearing Research*, *43*, 865–878.

Fagundes, D., Haynes, W., Haak, N., & Moran, M. (1998). Task variability effects on the language test performance of Southern lower socioeconomic class African American and Caucasian five-year-olds. *Language, Speech and Hearing Services in Schools*, *29*, 148–157.

Gaulin, C., & Campbell, T. (1994). Procedure for assessing verbal working memory in normal school-age children: some preliminary data. *Perceptual and Motor Skills*, *79*, 55–64.

Lidz, C., & Pena, E. (1996). Dynamic assessment: the model, its relevance as a nonbiased approach and its application to Latino American preschool children. *Language, Speech and Hearing Services in Schools*, *27* (4), 367–372.

Newcomer, P., & Hammill, D. (1988). *Test of language development 2-primary*. Austin, TX: Pro-Ed.

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

Stockman, I. (1996). The promises and pitfalls of language sample analysis as an assessment tool for linguistic minority children. *Language, Speech and Hearing Services in Schools*, *27*, 355–366.

Taylor, O., Payne, K., & Anderson, N. (1987). Distinguishing between communication disorders and communication differences. In: *Seminars in speech and language*, (*8*, pp. 415–427). New York, NY: Thieme Medical Publishers.

Vaughn-Cooke, F. (1983). Improving language assessment in minority children. *Journal of the American Speech–Language–Hearing Association*, *25* (9), 29–33.

Washington, J., & Craig, H. (1998). Socioeconomic status and gender influences on children's dialectal variations. *Journal of Speech, Language, Hearing Research*, *41* (3), 618–626.

Washington, J., & Craig, H. (1999). Performances of at-risk, African American preschoolers on the Peabody Picture Vocabulary Test-III. *Language, Speech and Hearing Services in Schools*, *30*, 74–82.

Woodcock, R. (1991). *Woodcock language proficiency battery-revised*. Allen, TX: DLM Teaching Resources.