Research Article

Nonword Repetition Across Two Dialects of English: Effects of Specific Language Impairment and Nonmainstream Form Density

Janet L. McDonald^a and Janna B. Oetting^b

Purpose: Nonword repetition (NWR) has been proposed as a culturally and linguistically fair measure of children's language abilities that is useful for the identification of specific language impairment (SLI). However, Moyle, Heilmann, and Finneran (2014) suggested that the density of a child's nonmainstream forms also influences NWR in ways that could complicate its use. Using speakers of either African American English (AAE) or Southern White English (SWE), we asked if NWR performance differed in children with SLI and same dialect-speaking typically developing (TD) children and if nonmainstream form density impacted their scores.

Method: The participants were 106 kindergartners (AAE: SLI $n = 35$; TD $n = 35$; SWE: SLI $n = 18$; TD $n = 18$; groups matched for age and IQ) who performed the NWR task of Dollaghan and Campbell (1998). Nonmainstream form

density measures were gathered from listener judgments of conversational samples.

Results: NWR performance differed between those with and without SLI, but the difference was smaller in AAE than in SWE, especially at the longest syllable length. Nonmainstream form density was found to further explain NWR performance beyond the children's SLI status for AAE speakers; density and SLI status were confounded for the SWE speakers, making it harder to disentangle the effects of each in that dialect.

Conclusions: Results indicate the NWR may differ in diagnostic utility between speakers of different dialects. Results also support Moyle et al.'s (2014) finding that density affects NWR. Thus, NWR is more sensitive to dialectal differences than originally assumed.

Standardized tests of mainstream English are frequently used to assess children's language abilities. However, these tests often lack validity when children are from a minority culture or speak a nonmainstream dialect of English (Campbell, Dollaghan, Needleman, & Janosky, 1997; Weismer et al., 2000; Rodekohr & Haynes, 2001). In an effort to reduce cultural and linguistic biases found in standardized tests, researchers have proposed the use of processing-based measures such as nonword repetition (NWR) or verbal working memory; such tasks emphasize processing and de-emphasize the influence of prior language experience (Campbell et al., 1997; Engel, Santos,

& Gathercole, 2008; Laing & Kamhi, 2003; Oetting & Cleveland, 2006; Rodekohr & Haynes, 2001; Washington & Craig, 2004).

The processing-based measure explored in this article, the NWR task, requires children to repeat a series of unfamiliar nonwords, and phoneme errors and omissions are noted. Past research has shown that minority cultural and dialectal groups perform as well on NWR tasks as majority groups (Campbell et al., 1997; Oetting & Cleveland, 2006; Rodekohr & Haynes, 2001; Weismer et al., 2000), and this task distinguishes typically developing (TD) children from those with specific language impairment (SLI; e.g., Bishop, North, & Donlan, 1996; Conti-Ramsden, 2003; Conti-Ramsden & Hesketh, 2003; Dollaghan & Campbell, 1998; Graf Estes, Evans, & Else-Quest, 2007; Montgomery & Evans, 2009; Montgomery & Windsor, 2007; Oetting & Cleveland, 2006; Rodekohr & Haynes, 2001). The current study examines the utility of the NWR task for identifying children with SLI in the context of two nonmainstream dialects—African American English (AAE) and Southern White English (SWE). In addition, it examines the role of nonmainstream form density

^aDepartment of Psychology, Louisiana State University, Baton Rouge

^bDepartment of Communication Sciences & Disorders, Louisiana State University, Baton Rouge

Correspondence to Janet L. McDonald: psmcd@lsu.edu

Editor-in-Chief: Sean Redmond

Editor: Lisa Archibald

Received June 21, 2018

Revision received November 7, 2018

Accepted November 24, 2018

https://doi.org/10.1044/2018_JSLHR-L-18-0253

Disclosure: The authors have declared that no competing interests existed at the time of publication.

on this task, because a recent study of at-risk AAE-speaking children found that NWR scores were negatively correlated to their nonmainstream form densities (Moyle, Heilmann, & Finneran, 2014). The study of AAE and SWE together allows for an expanded range of nonmainstream form densities, and the inclusion of children with SLI allows for an expanded range of language aptitudes.

NWR Studies With Minority Cultural Groups and Nonmainstream English Speakers

Several studies have examined how members of a minority versus majority cultural group perform on NWR. Campbell et al. (1997) examined the performance of 12-year-old boys on a standardized, knowledge-based language test, an NWR task, and two other processing-based tasks. The minority group, which was overwhelmingly African American, scored a standard deviation below the majority Caucasian group on the knowledge-based language test. However, there were no significant differences between the two groups on the NWR task or the other two processing-based tasks. Weismer et al. (2000) tested 8-year-old children on standardized tests of vocabulary and grammar and on a NWR task. Similar to Campbell et al., they found that the African American children scored significantly lower than Caucasian children on the standardized tests of language but that the two groups did not differ on NWR. In both of these studies, no information was provided about the participants' dialects, so we do not know if NWR was dialect neutral. However, we can conclude that NWR seems to be a culturally fairer measure than some standardized tests of language.

Other studies have compared groups based specifically on the dialect spoken. Rodekohr and Haynes (2001) tested 7-year-old children who were either African American and speakers of AAE (as confirmed by language samples) or Caucasian and speakers of a non-AAE dialect (dialect assumed but not confirmed). They found that AAE speakers scored lower than their Caucasian counterparts on a standardized language test, whereas the two groups did not differ on NWR. Oetting and colleagues (Horohov & Oetting, 2004; Oetting & Cleveland, 2006) also showed that 4- to 6-year-old AAE-speaking children did not differ from similarly aged SWE-speaking children on NWR.

Nonmainstream Form Density and the Study of AAE and SWE

In addition to the type of dialect spoken, researchers have looked at the density with which children produce nonmainstream dialect forms and how different densities of these forms impact NWR performance. Nonmainstream forms of English are those that contrast in some way with forms produced in mainstream English varieties (Seymour, Bland-Stewart, & Green, 1998). As reviewed by Oetting (in press), they can include (a) zero marked forms (e.g., *he Ø walking*), (b) overtly marked forms that are specific to one or more nonmainstream dialect (e.g., *I'ma buy it; I ain't wearing that*), (c) overtly marked forms that are

shared across dialects but produced with expanded functions (e.g., *seen* to mark simple past tense as in *I seen it* or *be* to indicate an ongoing or habitual activity as in *She be working*), and (d) overtly marked forms that are shared across dialects but show variation in the rates at which they are produced (e.g., appositive pronouns as in *My sister, she went there*). Density metrics index the relative frequencies of children's nonmainstream forms when they are conversing with others or answering questions, and although different metrics can be found in the literature, all appear correlated to each other (Horton-Ikard & Apel, 2014; Oetting & McDonald, 2002).

One of the first studies to examine the relation between children's nonmainstream form densities and their NWR performance was by Terry, Connor, Petscher, and Conlin (2012). Their participants were 49 first graders of various races and ethnicities, their nonmainstream form densities were calculated using 15 items from Part I of the Diagnostic Evaluation of Language Variation–Screening Test (DELV-ST; Seymour, Roeper, & de Villiers, 2003), and their NWR scores were from six items on Part II of the DELV-ST. Results showed that children who decreased their nonmainstream densities from first to second grade earned higher NWR scores than children who did not. Moyle et al. (2014) also examined the relation between children's nonmainstream form densities and NWR abilities. The participants were 50 African American preschoolers, their nonmainstream form densities were calculated from a 5-min language sample, and their NWR scores were based on 16 items from Dollaghan and Campbell (1998). Results revealed a negative correlation between the children's nonmainstream form densities and their NWR scores and a significant group difference in NWR scores when the children were divided by their densities (high vs. low) using a median split procedure.

Findings from both studies raise issues about NWR, because both show speakers with high nonmainstream form densities to be disadvantaged. Recent findings on a different processing-based measure—size judgment, which measures a child's working memory—have raised the same issue; AAE- and SWE-speaking children's nonmainstream form densities impacted performance on this supposed unbiased measure as well (McDonald, Seidel, Hammarlund, & Oetting, 2018). In addition, McDonald et al. found that, among TD children, AAE speakers had higher nonmainstream form density rates than SWE speakers and that children with SLI had higher density rates than TD children, and the presence of an interaction signaled this effect was particularly strong for AAE speakers. As illustrated by this study, inclusion of AAE and SWE child speakers within a single study leads to greater representation of children with low and high nonmainstream form densities than can be achieved by studying only one of these dialects.

Children With SLI

The evidence is overwhelming that children with SLI have more trouble with NWR tasks than do TD children.

For example, Conti-Ramsden (2003) had English-speaking 5-year-old children with SLI, and an age-matched TD group repeat Gathercole and Baddeley's (1996) nonwords and scored them by number of entire words correctly imitated. She found the children with SLI performed significantly worse (28th percentile) than the TD group (78th percentile). Montgomery and Evans (2009) examined the NWR scores of 6- to 12-year-old English-speaking children with and without SLI using Dollaghan and Campbell's (1998) nonwords. Responses were scored by the percentage of phonemes correctly repeated. They found the children with SLI performed significantly worse ($M = 79\%$) than the TD children ($M = 90\%$). Finally, in a meta-analysis of 23 NWR studies, Graf Estes et al. (2007) found that children with SLI performed on average more than a standard deviation below that of TD children.

Three studies have examined the clinical utility of NWR in nonmainstream English speakers. As mentioned above, these studies have found NWR to be dialect neutral while also showing group differences between those with and without SLI (Horohov & Oetting, 2004; Oetting & Cleveland, 2006; Rodekohr & Haynes, 2001). For example, Oetting and Cleveland (2006) found that children with SLI scored significantly worse than TD children in both AAE and SWE. There was no main effect for dialect type or an interaction between the children's SLI status and dialect type. Rodekohr and Haynes (2001) also found NWR differences between children with and without SLI in their two dialects. However, they also found 2 two-way interactions—one between syllable length and SLI status and one between syllable length and dialect type. Although there was no three-way interaction, they noted that both two-way interactions were related to especially low performance by the SWE-speaking SLI group on the four-syllable stimuli—hinting that this subgroup may be especially vulnerable, especially when the number of syllables to be repeated is high. Note that neither of these prior studies have investigated the children's nonmainstream form densities.

The Current Study

We first compare each group of children with SLI to a matched group of same dialect-speaking TD children to see if their NWR scores are significantly different. Based on previous studies, we predicted differences between those with and without SLI in both dialects without significant dialect differences. However, because of the trend shown in Rodekohr and Haynes (2001), we also examine whether the SWE speakers with SLI might have lower scores than other groups, especially on nonwords with a high number of syllables. We then examine correlations between the children's NWR scores and their nonmainstream form densities in AAE and SWE to see if density accounts for variance in performance, and whether this is in addition to SLI status differences. Based on the findings of Moyle et al. (2014), we expected form density effects on the children's NWR scores. However, Moyle et al. did not compare TD children to

children with SLI, so we do not know if density and clinical status have independent effects on NWR performance; this would be important to know when using this task in research and clinical practice.

Method

The study was approved by the Louisiana State University Institutional Review Board, and parental consent and child assent were obtained from all participants prior to the study.

Participants

The participants were 106 kindergartners who were recruited from public schools located in rural Louisiana (for other studies with these same children, see McDonald et al., 2018; Oetting et al., 2016). There were 52 male and 54 female participants, and they ranged in age from 59 to 74 months ($M = 66.24$; $SD = 3.78$). Based on caregiver report, 65 were classified as African American, 34 as Caucasian, one as Asian, one as American Indian, one as mixed race, and four as not reported (with African American listed by the schools for these children). All participated in the schools' free lunch program and passed a school-based hearing screening. For the 102 caregivers who provided the information, maternal education ranged from sixth grade to age 17+ years (beyond college; $M = 12.57$; $SD = 2.70$).

A blind listener judgment task was initially used to classify the children's dialects as either AAE or SWE. The listener judgment task asked three trained listeners to independently rate the children's dialects after listening to 1 min of conversational speech. As reported in Oetting et al. (2016), at least two of three listeners rated 90% of the African American children (which included the one mixed race child and four children without caregiver race information) as speakers of AAE and 94% of the non-African American children as speakers of SWE. Analysis of longer conversational samples, responses on Part I of the DELV-ST, and the children's race confirmed the AAE and SWE dialects of the other children. In each of these cases, the child's African American versus non-African American status corresponded to their AAE ($n = 70$) versus SWE ($n = 36$) dialect classification.

Fifty-three of the children (AAE $n = 35$; SWE $n = 18$) met the criteria for SLI. These children scored $\leq -1 SD$ on the Syntax subtest of the Diagnostic Evaluation of Language Variation–Norm Referenced (DELV-NR; Seymour, Roeper, & de Villiers, 2005; scores ranged from 1 to 7), as well as $\geq -1 SD$ on the Goldman-Fristoe Test of Articulation–Second Edition (GFTA-2; Goldman & Fristoe, 2000; scores ranged from 89 to 113), and $\geq -1.2 SD$ on the Primary Test of Nonverbal Intelligence (PTONI; Ehrler & McGhee, 2008; scores ranged from 82 to 125). Although not given to determine the SLI status of the children, their vocabulary was measured via the Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4; Dunn & Dunn, 2007; scores ranged from 66 to 111). Thirteen (25%) of

the children were receiving services by a speech-language pathologist.

The 53 children with SLI were matched for dialect, age and IQ, and to the extent possible, maternal education level, with 53 TD (AAE $n = 35$; SWE $n = 18$) children. These children scored > -1 SD on the Syntax subtest of the DELV-NR (scores ranged from 8 to 14), ≥ -1 SD on the GFTA-2 (scores ranged from 92 to 116), and ≥ -1.2 SD on the PTONI (score ranged from 84 to 117). Their PPVT-4 scores ranged from 85 to 117. None of these children were receiving services or had a history of receiving services by a speech-language pathologist.

The use of the -1 SD cutoff on DELV-NR to classify children as clinically impaired versus TD has been shown to have both excellent sensitivity (.95) and specificity (.93; Seymour et al., 2005). This test has also been shown not to yield any significant differences in scores between children speaking AAE and mainstream American English (Seymour et al., 2005).

Demographic and standardized test scores by dialect and clinical group are shown in Table 1. As also reported in Oetting et al. (2016) and McDonald et al. (2018), 2 (dialect) $\times$ 2 (clinical group) analyses of variance (ANOVAs) showed good matching on age and IQ; maternal education had a small effect of clinical group (TD $>$ SLI). Standardized

language tests all showed the expected advantage for TD children over children with SLI. In addition, the vocabulary and articulation measures showed a dialect difference, with the AAE-speaking children having lower scores than the SWE-speaking children. This matches the findings of others who find that standardized tests show differences based on race, culture, or dialect (Campbell et al., 1997; Rodekohr & Haynes, 2001; Weismer et al., 2000). There were no significant interactions on any of the demographics or standardized language tests.

We had three different measures of nonmainstream form density available. Two were based on responses to Part I of the DELV-ST. The standard scoring classified children into one of three categories (low, some, strong variation from mainstream use), while a continuous scoring method was obtained by dividing the children's number of nonmainstream responses by the sum of their mainstream and nonmainstream responses (Terry, Connor, Thomas-Tate, & Love, 2010). The third measure came from the listener judgment task that was used to rate the children's type of dialect from the children's conversational samples. Following Oetting and McDonald (2002), after the listeners rated each child's dialect (SWE, AAE, other), they then rated the children's use of nonmainstream AAE or SWE forms using a 7-point scale (1 = *none* to 7 = *heavy use*).

Table 1. Demographics, standardized test scores, and dialect density of specific language impairment (SLI) and typically developing (TD) groups by dialect.

M/F ratio	AAE				SWE				Significant effects
	SLI		TD		SLI		TD		
	14M/21F		18M/17F		11M/7F		9M/9F		
	M	SD	M	SD	M	SD	M	SD	
Matching variables									
Age (months)	66.94	(3.74)	65.60	(3.55)	65.72	(3.89)	66.61	(4.18)	none
PTONI	93.69	(9.62)	98.09	(8.87)	96.50	(8.35)	98.28	(8.14)	none
Maternal education (years)	11.67	(2.27)	13.27	(2.62)	12.33	(2.87)	13.17	(3.05)	group: $F(1, 98) = 4.96$, $p = .028$, $\eta_p^2 = .05$
Language measures									
DELV-NR Syntax	4.83	(1.01)	10.00	(1.55)	4.78	(1.66)	10.39	(1.72)	group: $F(1, 102) = 328.80$, $p < .001$, $\eta_p^2 = .76$
PPVT-4	82.34	(9.42)	101.06	(9.32)	85.78	(7.01)	105.56	(5.62)	group: $F(1, 102) = 122.32$, $p < .001$, $\eta_p^2 = .55$ dialect: $F(1, 102) = 5.20$, $p = .03$ $\eta_p^2 = .05$
GFTA-2	104.49	(5.72)	107.00	(4.38)	104.78	(4.18)	110.50	(3.09)	group: $F(1, 102) = 18.49$, $p < .001$, $\eta_p^2 = .15$ dialect: $F(1, 102) = 3.92$, $p = .05$ $\eta_p^2 = .04$
Nonmainstream form density measure									
Density rating	5.06	(1.38)	4.44	(1.43)	4.44	(1.16)	2.43	(.65)	group: $F(1, 102) = 25.64$, $p < .001$, $\eta_p^2 = .20$ dialect: $F(1, 102) = 25.41$, $p < .001$, $\eta_p^2 = .20$ interaction: $F(1, 102) = 7.22$, $p = .008$, $\eta_p^2 = .07$

Note. AAE = African American English; SWE = Southern White English; M = male; F = female; PTONI = Primary Test of Nonverbal Intelligence; DELV-NR = Diagnostic Evaluation of Language Variation–Norm Referenced; PPVT-4 = Peabody Picture Vocabulary Test–Fourth Edition; GFTA-2 = Goldman-Fristoe Test of Articulation–Second Edition.

The density ratings of the three raters were significantly correlated, with r 's ranging from .58 to .68, all p 's < .001, with 91% of ratings across judges being within 2 points of each other. The ratings of the three judges were averaged, and this served as the third measure. The density rates from each of the three methods were intercorrelated (DELV-ST standard scoring approach vs. DELV-ST Terry et al. scoring approach $r = .86$, $p < .001$; DELV-ST standard scoring approach vs. listener judgment $r = .43$, $p < .001$; DELV-ST Terry et al. scoring approach vs. listener judgment $r = .55$, $p < .001$). Because Moyle et al. (2014) had computed their nonmainstream form density from language samples (albeit from hand coding the presence of nonmainstream forms rather than by listener judgments), we used our listener judgments from the samples measure. A 2×2 ANOVA on this measure showed main effects of both dialect and clinical group, as well as a significant interaction due to the lower density ratings for SWE TD children. As can be seen in Table 1, the difference in nonmainstream form density between the AAE and SWE speakers was not significant for the children with SLI, $t(51) = -1.60$, $p = .11$, $\eta_p^2 = .05$, but it was for the TD children, $t_{\text{corr}}(51) = -7.03$, $p < .001$, $\eta_p^2 = .39$. In addition, the difference between the children with and without SLI was not significant for the AAE speakers, $t(68) = 1.84$, $p = .07$, $\eta_p^2 = .05$, but it was for the SWE speakers, $t(34) = 6.43$, $p < .001$, $\eta_p^2 = .55$.¹

Stimuli and Procedure

The 16 nonwords were taken from Dollaghan and Campbell (1998). The nonwords ranged in length from one to four syllables, with four nonwords per length. Unlike other NWR tasks (Graf Estes et al., 2007), the Dollaghan and Campbell's stimuli were designed to reduce wordlikeness and the occurrence of real words within the target stimuli. All nonwords followed a consonant vowel consonant format, with no consonant clusters, "late" consonants (Shriberg & Kwiatkowski, 1994), or repetitions of consonants or vowels within a given nonword. In addition, nonwords only featured tense vowels and consonants in unpredictable syllable positions.

The nonword stimuli were recorded by an African American female speaker, uploaded to a PowerPoint presentation, and presented via laptop loudspeakers. Prior to the NWR task, children were read instructions and given three practice items. Each nonword was presented only once; however, in the rare occurrence that an external distraction occurred during testing that obscured the presentation of the stimulus, the experimenter was instructed to repeat the target nonword via the PowerPoint presentation. All responses were audio-recorded and scored off-line. The children completed the NWR task after the administration of

most of the reported standardized tests, a language sample session, and other tasks reported elsewhere (McDonald et al., 2018; Oetting et al., 2016).

Scoring

The NWR task was scored by calculating the children's percentage of phonemes correct (PPC). Following Dollaghan and Campbell (1998), omitted phonemes and target phoneme substitutions were considered errors, whereas phonemes that were added or distorted were not classified as errors. When a child failed to respond to one or more target nonwords, these were not counted as omitted phonemes, but rather these items were eliminated from the scoring (this occurred for only 14 children). A percentage of phonemes correct score was calculated for each participant based on the total number of correctly produced phonemes in their appropriate positions divided by the total phonemes in the words they attempted (maximum = 96).

Reliability

Scoring reliability was examined by having an independent examiner score 20% of the NWR responses across all the children in a larger study. These children included the SLI group, and those selected as part of the matched TD group reported here as well as others who were not selected as matches. Point-by-point agreement was 94%.

Results

Percentage of Phonemes Correct

A 2 (dialect) $\times 2$ (clinical group) between-subjects ANOVA on the percentage of phonemes correct across all lengths showed no main effect of dialect, $F(1, 102) = 2.86$, $p = .09$, $\eta_p^2 = .03$; a significant main effect of clinical group, $F(1, 102) = 46.47$, $p < .001$, $\eta_p^2 = .31$; and an interaction between dialect and clinical group that narrowly missed the conventional level of significance, $F(1, 102) = 3.40$, $p = .068$, $\eta_p^2 = .03$. Means are given in Table 2. We note that the difference in performance between TD children and those with SLI was significant for both AAE speakers, $t_{\text{corr}}(57) = -3.94$, $p < .001$, $\eta_p^2 = .19$, and SWE speakers, $t(34) = -6.59$, $p < .001$, $\eta_p^2 = .56$; the nonsignificant tendency for it to be stronger in the SWE than AAE speakers will be further tested in our analyses by syllable length below.

Analysis at the individual syllable lengths found a main effect of dialect only at the four-syllable length, $F(1, 102) = 5.62$, $p = .02$, $\eta_p^2 = .05$, with the AAE speakers outperforming the SWE speakers. There also was a significant main effect of clinical group at all lengths greater than one syllable: one syllable, $F(1, 102) = 3.36$, $p = .07$, $\eta_p^2 = .03$; two syllables, $F(1, 102) = 23.20$, $p < .001$, $\eta_p^2 = .18$; three syllables, $F(1, 102) = 33.06$, $p < .001$, $\eta_p^2 = .24$; and four syllables, $F(1, 102) = 49.28$, $p < .001$, $\eta_p^2 = .33$. The main effects of dialect and group found at the four-syllable level were further qualified by the interaction between these factors, $F(1, 102) = 8.79$, $p = .004$, $\eta_p^2 = .08$. As shown

¹The other two measures of density also gave lower density scores for the SWE TD children and yielded the two main effects and a significant interaction. Although significance levels for the individual comparisons differed, all measures of nonmainstream form density showed a significant dialect difference for TD speakers and a significant clinical group difference for AAE speakers.

Table 2. Nonword repetition means (and standard deviations) by children with specific language impairment (SLI) and typically developing (TD) children by nonmainstream dialect spoken.

Dialect	SLI	TD
Nonword repetition PPC overall		
AAE	.70 (.13)	.80 (.08)
SWE	.63 (.09)	.80 (.06)
Nonword repetition PPC 1-syllable length		
AAE	.83 (.13)	.91 (.10)
SWE	.86 (.09)	.87 (.10)
Nonword repetition PPC 2-syllable length		
AAE	.82 (.09)	.90 (.07)
SWE	.80 (.09)	.89 (.09)
Nonword repetition PPC 3-syllable length		
AAE	.73 (.13)	.83 (.10)
SWE	.67 (.11)	.83 (.07)
Nonword repetition PPC 4-syllable length		
AAE	.57 (.17)	.69 (.12)
SWE	.42 (.14)	.70 (.11)

Note. PPC = Percentage of Phonemes Correct; AAE = African American English; SWE = Southern White English.

in Table 2, the gap between the two groups of children was significant in both dialects but smaller for the AAE speakers, $t_{\text{corr}}(62) = -3.34$, $p = .001$, $\eta_p^2 = .14$, than for the SWE speakers, $t(34) = -6.74$, $p < .001$, $\eta_p^2 = .57$, due to the especially poor performance by the SWE children with SLI. This low performance also caused the SWE children with SLI to be significantly different from the AAE children with SLI, $t(51) = 3.28$, $p = .002$, $\eta_p^2 = .17$, whereas TD speakers of the two dialects did not differ from each other, $t(51) = 0.51$, $p = .61$, $\eta_p^2 = .005$.

In order to investigate if individual item difficulty differed across our populations, we first ran correlations between the four different dialects by clinical group subsets across the 16 items. As shown in Table 3, all correlations were high, ranging from .72 to .96; all p 's $\leq .002$. In particular, the correlations within each diagnostic group across dialects are both above .90 (SLI group AAE-SWE $r = .91$; TD group AAE-SWE $r = .92$). Thus, the items varied similarly in difficulty for all subgroups. Then we also redid the above 2×2 ANOVA including the individual items as a repeated-measures factor. Individual items on the test varied in difficulty as indicated by a significant item effect at

Table 3. Correlations on item difficulty across diagnostic groups by dialects.

	AAE SLI	AAE TD	SWE SLI	SWE TD
AAE SLI	—			
AAE TD	.96***	—		
SWE SLI	.91***	.88***	—	
SWE TD	.84***	.92***	.72**	—

Note. AAE = African American English; SLI = specific language impairment; TD = typically developing; SWE = Southern White English.

** $p = .002$. *** $p < .001$.

each of the syllable lengths: one syllable, $F(3, 274) = 3.23$, $p = .026$, $\eta_p^2 = .03$; two syllables, $F(3, 291) = 16.30$, $p < .001$, $\eta_p^2 = .14$; three syllables, $F(3, 267) = 30.34$, $p < .001$, $\eta_p^2 = .24$; and four syllables, $F(3, 288) = 3.16$, $p = .025$, $\eta_p^2 = .03$. More interestingly, there were also a few interactions with item, which showed significant group effects or the interaction occurring for only some of the items (usually the more difficult items as indicated by low percentage of phonemes correct) within the various lengths. At the two-syllable level, there was an item by clinical group interaction, $F(3, 291) = 8.97$, $p < .001$, $\eta_p^2 = .08$, which showed significant differences between the clinical groups (SLI < TD) on only two of the four items (*vaefap*, *nortaof*). At the three-syllable level, there was an item by clinical group interaction, $F(3, 291) = 2.84$, $p = .043$, $\eta_p^2 = .03$, and a three-way interaction between item, clinical group, and dialect, $F(3, 291) = 4.03$, $p = .003$, $\eta_p^2 = .05$. All items at this syllable level showed the effect of clinical group, and one item (*dotaovaeb*) also showed a Group $\times$ Dialect interaction similar to the one found at the four-syllable level—that is, that children with SLI who spoke SWE were much worse than the other three groups, including children with SLI who spoke AAE. Thus, the overall picture is that the performance of children with SLI can be differentiated from TD children as the nonwords increase in difficulty due to their length, or even within a length if the item is a more difficult one. In addition, when items are the most difficult (i.e., four-syllable items and one of the three-syllable items), there is a larger difference in performance between the clinical groups for the SWE than AAE speakers.

Finally, two different cut-points were identified that optimized the classification of children into those with and without SLI based on their overall percentage of phonemes correct on the NWR task. A cut-point of .65 classified 75.5% of the children correctly (sensitivity: 53%; specificity: 98%). However, this rate was poorer for AAE speakers (70%: sensitivity: 43%; specificity: 97%) than SWE speakers (86%: sensitivity: 72%; specificity: 100%). The other cut-point was .76, which also classified 75.5% of the children correctly (sensitivity: 77%; specificity: 74%). Again, this was poorer for AAE speakers (70%: sensitivity: 69%; specificity: 71%) than SWE speakers (86%: sensitivity: 94%; specificity: 78%). We can compare these results to Oetting and Cleveland (2006) who also tested AAE- and SWE-speaking children on the NWR. They found an overall classification rate of 81% (sensitivity: 56%; specificity: 92%), which corresponds fairly well with the sensitivity and specificity in our data set with the cut-point of .65.

The results of the current study indicate that the NWR task generally differentiated the TD group from the SLI group in both AAE and SWE. However, there was an interaction between dialect and group at the four-syllable length (and for 1 three-syllable item as well) due to the low performance by the SWE children with SLI. Classification accuracy was also worse for the AAE- than SWE-speaking children, because smaller group differences were found for the AAE speakers than for the SWE speakers. Because dialect density also differed between the children

with and without SLI for SWE speakers, it is possible that this factor exacerbated the difference between these groups. We therefore turn to the effects of nonmainstream form density on NWR.

Nonmainstream Form Density

Table 4 shows the correlations between NWR, nonmainstream form density, and clinical group. Performance on the NWR is highly correlated to both density and group. To see if group and density made independent contributions, an analysis of covariance (ANCOVA) was run with dialect and group as factors, and nonmainstream form density as the covariate. Considering both dialects together, dialect type was significant, $F(1, 101) = 7.18, p = .001, \eta_p^2 = .07$; group was significant, $F(1, 101) = 26.23, p < .001, \eta_p^2 = .21$; and density was also significant, $F(1, 101) = 6.36, p = .013, \eta_p^2 = .06$. When further broken down into the two dialects, an interesting picture emerged. For AAE speakers, the ANCOVA yielded an effect of group, $F(1, 67) = 12.03, p < .001, \eta_p^2 = .15$, and a significant effect of density, $F(1, 67) = 4.96, p = .029, \eta_p^2 = .07$. Thus, nonmainstream form density added significant explanatory variance above clinical group for the AAE speakers. However, for the SWE speakers, the ANCOVA only yielded a significant effect of group, $F(1, 33) = 13.70, p = .001, \eta_p^2 = .29$, and not one for density, $F(1, 33) = .96, p = .33, \eta_p^2 = .03$. The high correlation between group membership and nonmainstream form density ($r = .74$) for the SWE speakers left little variance for density to capture beyond the group factor.²

We also did a further exploration of the nonmainstream form density effect found in the AAE speakers by running the ANCOVA with clinical group as a factor and density as a covariate at each syllable length. These findings showed that the group effect remained significant at all syllable levels, but that the impact of density decreased as the number of syllables increased (one syllable: $F(1, 67) = 8.21, p = .006, \eta_p^2 = .11$; two syllables: $F(1, 67) = 8.86, p = .004, \eta_p^2 = .11$; three syllables: $F(1, 67) = 3.52, p = .065, \eta_p^2 = .05$; four syllables: $F(1, 67) = 3.20, p = .078, \eta_p^2 = .05$). Thus, for the AAE speakers, density effects may be more visible on items with fewer syllables and higher performance.

In addition, because one of the defining phonological characteristics of AAE is the omission of consonants in the word final position, particularly in consonant clusters (Stockman, 2006), we also looked at the effect of density

Table 4. Correlations between nonword repetition scores, nonmainstream form density, and group.

Dialect	NWR PPC	Density
All		
NWR PPC		
Density rating	-.33***	
Group	.53***	-.36***
AAE		
NWR PPC		
Density rating	-.33**	
Group	.43***	-.22
SWE		
NWR PPC		
Density rating	-.63***	
Group	.75***	-.74***

Note. NWR PPC = nonword repetition percentage of phonemes correct; AAE = African American English; SWE = Southern White English. ** $p < .01$. *** $p \leq .001$.

as a covariate on initial versus final phonemes of the nonwords (note that these nonwords did not contain any consonant clusters). For the AAE speakers, there was no significant contribution of density to the performance on the initial phoneme for any syllable length—one syllable: $F(1, 67) = 1.16, p = .29, \eta_p^2 = .02$; two syllables: $F(1, 67) = 3.34, p = .072, \eta_p^2 = .05$; three syllables: $F(1, 67) = 1.24, p = .27, \eta_p^2 = .02$; four syllables: $F(1, 67) = 0.24, p = .62, \eta_p^2 = .004$. However, there were significant contributions of density for the final phoneme for the shorter nonwords (one syllable: $F(1, 67) = 6.97, p = .02, \eta_p^2 = .09$; two syllables: $F(1, 67) = 6.42, p = .014, \eta_p^2 = .09$), but this effect was not there for the longer nonwords (three syllables: $F(1, 67) = 0.49, p = .49, \eta_p^2 = .01$; four syllables: $F(1, 67) = 0.13, p = .72, \eta_p^2 = .002$). Thus, the effect of nonmainstream form density found for AAE speakers may be a reflection of their tendency to omit word final consonants, at least for the shorter items. Similar analyses of the SWE speakers found no effect of dialect density beyond group membership on either the initial or final phoneme at any syllable length. This may indicate that the tendency to omit final phonemes on the NWR may be dialect specific; however, recall that the confound between group membership and dialect density in SWE speakers does not allow any definitive conclusions.

Finally, we were interested if factors other than group and density could account for additional variance in performance on the NWR. We therefore did a hierarchical regression on percentage of phonemes correct to see if vocabulary scores, articulation scores, or IQ scores would add extra explanatory variance beyond clinical group membership and nonmainstream form density. Group and form density were included in the first step, and then PPVT-4 standard scores, GFTA-2 standard scores, and PTONI standard scores were added in the next step (note that DELV-NR syntax scores were not added, as they were already implicitly included in clinical group, which was assigned based on DELV-NR syntax scores). For all children considered

²Similar results occur if the DELV-ST Terry et al. (2010) scoring method for calculating nonmainstream form density is used instead of the listener judgment approach, that is, group remains significant for the children as a whole and for each dialect considered individually. The covariate of density is significant in addition to group membership only when considering the AAE speakers alone. Different results are obtained for the DELV-ST scoring method. Here, group remains significant for all children and each dialect, but the covariate of density is never significant. Thus, the more coarse DELV-ST three-category scoring of nonmainstream form density was not as good as the finer grained scoring method of Terry et al. in finding the effects of density on NWR.

together, group and nonmainstream form density explained 31% of the variance in NWR scores, $F(2, 103) = 22.85$, $p < .001$. The addition of the three standardized tests increased the variance explained to 32%; this was not a significant increase, $F(3, 100) = 0.75$, $p = .53$. For AAE speakers, group and density explained 24% of the variance, $F(2, 67) = 10.68$, $p < .001$, and the addition of the standardized tests increased the explained variance to 28%, which was not significant, $F(3, 64) = 1.00$, $p = .40$. For SWE speakers, group and density explained 57% of the variance, $F(2, 33) = 22.18$, $p < .001$, and the addition of the standardized tests increased the explained variance to 65%, which was not significant, $F(3, 30) = 2.22$, $p = .11$. Thus, scores on these three standardized tests did not significantly increase variance accounted for beyond group and density.

Discussion

We hypothesized that NWR would be equally able to distinguish children with SLI from TD counterparts in two different nonmainstream dialects of English. Performance on the NWR task showed significant differences between those with and without SLI for speakers of both dialects. Thus, NWR performance continues to bear up as a clinical marker of SLI even in nonmainstream dialects, as has been found by others (Oetting & Cleveland, 2006; Rodekohr & Haynes, 2001). However, we also found that NWR, particularly on more difficult items, separated the children with SLI from TD controls better for SWE speakers than for AAE speakers, something not found by these researchers. There was especially low performance by SWE-speaking children with SLI on items with four syllables and one item with three syllables. Similar low performance by a comparable group on these longer nonwords was previously found by Rodekohr and Haynes (2001) indicating a vulnerability in this population. However, in Rodekohr and Haynes, this low performance did not result in a statistically significant interaction containing both dialect and group, perhaps due to their smaller sample sizes. Thus, ours is the first study to find a significant dialect by clinical group interaction.

Our finding that NWR was worse at classifying AAE-speaking children (70%) than SWE-speaking children (86%) also contrasts with the findings of Oetting and Cleveland (2006). They did not find much difference in overall classification rates between their AAE speakers (78%) and SWE speakers (82%). However, a reexamination of their data showed that this was achieved differently in the two groups. Classification of AAE speakers yielded a sensitivity of 67% with a specificity of 83%, whereas that of SWE speakers yielded a sensitivity of 50% and specificity of 90%. Where Oetting and Cleveland found higher sensitivity for AAE speakers, we found higher sensitivity for SWE speakers. Note that Oetting and Cleveland had smaller numbers of children with SLI in their study, making the effect of misclassification of an individual higher.

We also investigated the susceptibility of the NWR task to nonmainstream form density. We found that density

made a contribution to performance above and beyond clinical group membership for AAE speakers. As indicated by the negative correlations in Table 4, AAE speakers with higher density had poorer NWR scores. The finding that density affects NWR replicates that of Moyle et al. (2014), who also investigated AAE speakers. Indeed, the magnitude of the correlation between NWR performance and density found by Moyle et al. ($r = -.44$) and that found in our study for the AAE speakers ($r = -.32$) is not statistically different, $z = -.74$, $p = .46$. However, due to the nature of our sample, we were not able to extend this finding to another dialect, SWE. Although there was a strong correlation between NWR performance and nonmainstream form density for our group of SWE speakers, density was also highly correlated with group membership in our sample. Thus, with group in the ANCOVA, there was little independent variance left for the covariate of density to contribute in explaining NWR performance. The finding that nonmainstream form density can affect a processing-based task such as NWR dovetails with other work in our lab that showed it also affected performance on a size judgment task, a processing-based working memory task (McDonald et al., 2018). Indeed, we checked the magnitude of the correlation between listener judgments of nonmainstream form density and the NWR ($r = -.33$) and between this same density measure and the size judgment task ($r = -.25$), and found these not to be significantly different, $z = -.84$, $p = .40$.

Such findings call into question the commonly held belief that processing-based tasks are immune from linguistic or cultural bias. Although it does appear to be the case that they are useful for distinguishing performance of children with and without SLI in different dialects, the additional influence of nonmainstream form density could influence their diagnostic usefulness both within and across dialects. For example, within any particular dialect, it could be hard to distinguish a child with SLI with low nonmainstream form density from a TD child with high nonmainstream form density if nonmainstream form density is not taken into account. Moreover, across dialects, it may be easier to distinguish children with and without SLI in dialects that have generally lower nonmainstream density than in dialects with high or highly variable densities.

We used listener judgments of nonmainstream form density in language samples in the current article to match in spirit Moyle et al.'s (2014) use of samples to calculate their density rates. However, they did a more laborious coding of the presence of nonmainstream forms in their samples, whereas our listeners rated density on a 7-point scale. Nevertheless, the listener judgments of density showed the same significant relationship to NWR performance in AAE speakers as Moyle et al. found for their coding of dialect density. In addition, we also had density rates available to us from the DELV-ST Part 1, which is more easily administered and scored. We found very similar results to our listener judgments if we used the Terry et al. (2010) scoring method for the DELV-ST—that is, significant extra variance beyond clinical group membership was

accounted for by nonmainstream form density for AAE speakers.

It is interesting to consider why nonmainstream form density may affect NWR. Speakers with high nonmainstream form density are most likely being exposed to different phonological and phonotactic sequences, and these at different frequencies than speakers with lower-density usage. Thus, their phonological representation of their language is probably different from that of a low-density speaker. Indeed, past research has shown that the densities with which nonmainstream forms of a dialect are spoken may influence children's phonological knowledge and lexical access (Brown, 2011; Edwards et al., 2014). Such differences could in turn affect performance on NWR. Indeed, we found that, for the AAE speakers, nonmainstream form density affected performance, particularly on the one- and two-syllable items, and within these items, it was the final phoneme and not the initial phoneme that was affected, relating perhaps to the tendency for speakers of AAE to delete final phonemes (Stockman, 2006). Note that the explanation for the effect of nonmainstream form density on NWR performance is different than that offered for poorer performance on the NWR by children with language impairment. These latter explanations include weaker vocabulary, poor phonological representation of words, trouble accessing sublexical phonological knowledge, difficulty repeating multiple syllables, and reduced phonotactic knowledge (Archibald & Gathercole, 2007; Coady & Evans, 2008; Coady, Evans, & Kluender, 2010; Edwards, Beckman, & Munson, 2004; Jones, Tamburelli, Watson, Gobet, & Pine, 2010; Leclercq, Maillart, & Majerus, 2013; McKean, Letts, & Howard, 2013; Munson, Kurtz, & Windsor, 2005; Tamburelli & Jones, 2013; Tattersall, Nelson, & Tyler, 2015).

Previous research has also shown that phonological knowledge differs by dialect spoken (Pearson, Velleman, Bryant, & Charko, 2009). For example, certain phonemes are acquired earlier in AAE than in mainstream American English, and vice versa. In addition, AAE speakers often reduce consonant clusters at the end of words, and this is also subject to phonological determinants (Stockman, 2006). Thus, any particular set of items used on an NWR task may do an excellent job of differentiating children with and without language impairments in one dialect and an adequate, but less definitive job of differentiating these groups in another dialect.

Conclusions

The NWR task can differentiate children with SLI from well-matched TD controls in nonmainstream dialects, at least those of AAE and SWE. However, differences in the effect sizes of the clinical group difference within the two dialects indicate that NWR may not be as culturally or dialectically unbiased as first thought. In addition, the current results show that a child's SLI status is not the only contributor to NWR performance—nonmainstream form density is also a significant contributor, at least in some dialects. This indicates that the task is also sensitive to

another linguistic characteristic, nonmainstream form density. Indeed, the generally lower nonmainstream form density of the TD SWE speakers in our sample may be what makes the NWR task better able to separate those with or without SLI in this dialect.

The findings also indicate that it is important to get nonmainstream form density measures when administering NWR tasks to nonmainstream dialect-speaking children. One can get a nonmainstream form density measure from a listener judgment task as was done here, an analysis of a language sample as was done by Moyle et al. (2014), or from a standardized test such as the DELV-ST. Any of these measures should help determine if children have a poor NWR score because they have SLI, or because they have high nonmainstream form density, or both.

Acknowledgments

This research was funded by the National Institute on Deafness and Other Communication Disorders Grant R01DC009811, awarded to Janna B. Oetting, Janet L. McDonald, and Michael Hegarty. We would like to thank Christy Seidel, Brittany Harris, Hannah Smitherman, Rebecca Hammarlund, Abby Massey, Julie Town, and a number of others who helped collect the data. We would also like to thank the teachers, families, and children who participated in the study.

References

- Archibald, L. M. D., & Gathercole, S. E. (2007). Nonword repetition and serial recall: Equivalent measures of verbal short-term memory? *Applied Psycholinguistics*, 28(4), 587–606.
- Bishop, D. M., North, T., & Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: Evidence from a twin study. *The Journal of Child Psychology and Psychiatry*, 37(4), 391–403.
- Brown, M. C. (2011). *Dialect and lexical access: An investigation into dialect density, dialect environment and word knowledge* (Unpublished doctoral dissertation). University of Wisconsin, Madison, WI.
- 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(3), 519–525.
- Coady, J. A., & Evans, J. L. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *International Journal of Language & Communication Disorders*, 43(1), 1–40.
- Coady, J. A., Evans, J. L., & Kluender, K. R. (2010). Role of phonotactic frequency in nonword repetition by children with specific language impairments. *International Journal of Language & Communication Disorders*, 45(4), 494–509.
- Conti-Ramsden, G. (2003). Processing and linguistic markers in young children with specific language impairment (SLI). *Journal of Speech, Language, and Hearing Research*, 46(5), 1029–1037.
- Conti-Ramsden, G., & Hesketh, A. (2003). Risk markers for SLI: A study of young language-learning children. *International Journal of Language & Communication Disorders*, 38(3), 251–263.

- Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research, 41*(5), 1136–1146.
- Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test—Fourth Edition (PPVT-4)*. Bloomington, MN: NCS Pearson.
- Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research, 47*(2), 421–436.
- Edwards, J., Gross, M., Chen, J., MacDonald, M. C., Kaplan, D., Brown, M., & Seidenberg, M. S. (2014). Dialect awareness and lexical comprehension of mainstream American English in African American English-speaking children. *Journal of Speech, Language, and Hearing Research, 57*(5), 1883–1895.
- Ehrler, D. J., & McGhee, R. L. (2008). *PTONI: Primary Test of Nonverbal Intelligence*. Austin, TX: Pro-Ed.
- Engel, P. M. J., Santos, F. H., & Gathercole, S. E. (2008). Are working memory measures free of socioeconomic influence? *Journal of Speech, Language, and Hearing Research, 51*(6), 1580–1587.
- Gathercole, S. E., & Baddeley, A. D. (1996). *The Children's Test of Nonword Repetition*. London, United Kingdom: The Psychological Corporation.
- Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation—Second Edition (GFTA-2)*. Circle Pines, MN: AGS.
- Graf Estes, K., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 50*(1), 177–195.
- Horohov, J. E., & Oetting, J. B. (2004). Effects of input manipulations on the word learning abilities of children with and without specific language impairment. *Applied Psycholinguistics, 25*(1), 43–65.
- Horton-Ikard, R., & Apel, K. (2014). Examining the use of spoken dialect indices with African American children in the southern United States. *American Journal of Speech-Language Pathology, 23*(3), 448–460.
- Jones, G., Tamburelli, M., Watson, S. E., Gobet, F., & Pine, J. M. (2010). Lexicality and frequency in specific language impairment: Accuracy and error data from two nonword repetition tests. *Journal of Speech, Language, and Hearing Research, 53*(6), 1642–1655.
- Laing, S. P., & Kamhi, A. (2003). Alternative assessment of language and literacy in culturally and linguistically diverse populations. *Language, Speech, and Hearing Services in Schools, 34*(1), 44–55.
- Leclercq, A.-L., Maillart, C., & Majerus, S. (2013). Nonword repetition problems in children with specific language impairment: A deficit in accessing long-term linguistic representations? *Topics in Language Disorders, 33*(3), 238–254.
- McDonald, J. L., Seidel, C. M., Hammarlund, R., & Oetting, J. B. (2018). Working memory performance in children with and without specific language impairment in two nonmainstream dialects of English. *Applied Psycholinguistics, 39*, 145–167.
- McKean, C., Letts, C., & Howard, D. (2013). Developmental change is key to understanding primary language impairment: The case of phonotactic probability and nonword repetition. *Journal of Speech, Language, and Hearing Research, 56*(5), 1579–1594.
- Montgomery, J. W., & Evans, J. L. (2009). Complex sentence comprehension and working memory in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 52*(2), 269–288.
- Montgomery, J. W., & Windsor, J. (2007). Examining the language performances of children with and without specific language impairment: Contributions of phonological short-term memory and speed of processing. *Journal of Speech, Language, and Hearing Research, 50*(3), 778–797.
- Moyle, M. J., Heilmann, J. J., & Finneran, D. A. (2014). The role of dialect density in nonword repetition performance: An examination with at-risk African American preschool children. *Clinical Linguistics & Phonetics, 28*(9), 682–696.
- Munson, B., Kurtz, B. A., & Windsor, J. (2005). The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetitions of children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research, 48*(5), 1033–1047.
- Oetting, J. B. (2019). Variability within varieties of language: Profiles of typicality and impairment. In T. Ionin & M. Rispoli (Eds.), *Streams of generative language acquisition research* (pp. 59–82). Philadelphia, PA: John Benjamins.
- Oetting, J. B., & Cleveland, L. H. (2006). The clinical utility of nonword repetition for children living in the rural south of the US. *Clinical Linguistics & Phonetics, 20*(7–8), 553–561.
- Oetting, J. B., & McDonald, J. L. (2002). Methods for characterizing participants' nonmainstream dialect use in child language research. *Journal of Speech, Language, and Hearing Research, 45*(3), 505–518.
- Oetting, J. B., McDonald, J. L., Seidel, C. M., & Hegarty, M. (2016). Sentence recall by children with SLI across two nonmainstream dialects of English. *Journal of Speech, Language, and Hearing Research, 59*(1), 183–194.
- Pearson, B. Z., Velleman, S. L., Bryant, T. J., & Charko, T. (2009). Phonological milestones for African American English-speaking children learning mainstream American English as a second dialect. *Language, Speech, and Hearing Services in Schools, 40*(3), 229–244.
- Rodekohr, R. K., & Haynes, W. O. (2001). Differentiating dialect from disorder: A comparison of two processing tasks and a standardized language test. *Journal of Communication Disorders, 34*(3), 255–272.
- Seymour, H. N., Bland-Stewart, L., & Green, L. J. (1998). Difference versus deficit in child African American English. *Language, Speech, and Hearing Services in Schools, 29*(2), 96–108.
- Seymour, H. N., Roeper, T., & De Villiers, J. G. (2003). *DELV: Diagnostic Evaluation of Language Variation™—Screening Test*. San Antonio TX: The Psychological Corporation.
- Seymour, H. N., Roeper, T., & De Villiers, J. G. (2005). *DELV: Diagnostic Evaluation of Language Variation™—Norm Referenced Test*. San Antonio TX: The Psychological Corporation.
- Shriberg, L., & Kwiatkowski, J. (1994). Developmental phonological disorders I: A clinical profile. *Journal of Speech and Hearing Research, 37*(5), 1100–1126.
- Stockman, I. (2006). Alveolar bias in the final consonant deletion patterns of African American Children. *Language, Speech, and Hearing Services in Schools, 37*(2), 85–95.
- Tamburelli, M., & Jones, G. (2013). Investigating the relationship between nonword repetition performance and syllabic structure in typical and atypical language development. *Journal of Speech, Language, and Hearing Research, 56*(2), 708–720.
- Tattersall, P. J., Nelson, N. W., & Tyler, A. A. (2015). Associations among nonword repetition and phonemic and vocabulary awareness: Implications for intervention. *Child Language Teaching and Therapy, 31*(2), 159–171.

-
- Terry, N. P., Connor, C. M., Petscher, Y., & Conlin, C. R.** (2012). Dialect variation and reading: Is change in nonmainstream American English use related to reading achievement in first and second grades? *Journal of Speech, Language, and Hearing Research, 55*(1), 55–69.
- Terry, N. P., Connor, C. M., Thomas-Tate, S., & Love, M.** (2010). Examining relationships among dialect variation, literacy skills, and school context in first grade. *Journal of Speech, Language, and Hearing Research, 53*(1), 126–145.
- Washington, J. A., & Craig, H. K.** (2004). A language screening protocol for use with young African American children in urban settings. *American Journal of Speech-Language Pathology, 13*(4), 329–340.
- Weismer, S. E., Tomblin, J. B., Zhang, X., Buckwalter, P., Chynoweth, J. G., & Jones, M.** (2000). Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech, Language, and Hearing Research, 43*(4), 865–878.