JSLHR

Research Forum

Sentence Recall by Children With SLI Across Two Nonmainstream Dialects of English

Janna B. Oetting,[a] Janet L. McDonald,[a] Christy M. Seidel,[a] and Michael Hegarty[a]

Purpose: The inability to accurately recall sentences has proven to be a clinical marker of specific language impairment (SLI); this task yields moderate-to-high levels of sensitivity and specificity. However, it is not yet known if these results hold for speakers of dialects whose nonmainstream grammatical productions overlap with those that are produced at high rates by children with SLI.
Method: Using matched groups of 70 African American English speakers and 36 Southern White English speakers and dialect-strategic scoring, we examined children's sentence recall abilities as a function of their dialect and clinical status (SLI vs. typically developing [TD]).
Results: For both dialects, the SLI group earned lower sentence recall scores than the TD group with sensitivity and specificity values ranging from .80 to .94, depending on the analysis. Children with SLI, as compared with TD controls, manifested lower levels of verbatim recall, more ungrammatical recalls when the recall was not exact, and higher levels of error on targeted functional categories, especially those marking tense.
Conclusion: When matched groups are examined and dialect-strategic scoring is used, sentence recall yields moderate-to-high levels of diagnostic accuracy to identify SLI within speakers of nonmainstream dialects of English.

Sentence recall (Briscoe, Bishop, & Norbury, 2001), also called *sentence repetition* (Conti-Ramsden, Botting, & Faragher, 2001; Riches, 2012) or *sentence imitation* (Eadie, Fey, Douglas, & Parsons, 2002; Seeff-Gabriel, Chiat, & Dodd, 2010), is a simple task in which participants repeat a sentence they have just heard, yet performance on this task is moderately to highly diagnostic in identifying participants who have specific language impairment (SLI). Indeed, Conti-Ramsden et al. (2001) found that children with SLI performed so poorly on this task that 90% of them fell below the 16th percentile, and 84% of the typically developing (TD) controls fell above this cutoff (for similarly high levels of diagnostic accuracy, see Redmond, Thompson, & Goldstein, 2011). Sentence recall also differentiates those with SLI from those with other disorders, including attention-deficit/hyperactivity disorder (Redmond, 2005; Redmond et al., 2011), consistent speech/phonological disorder (Seeff-Gabriel et al., 2010), developmental coordination disorder (Archibald & Alloway, 2008), and sensorineural hearing loss (Briscoe et al., 2001), and from those identified as resolved late talkers (Petruccelli, Bavin, & Bretherton, 2012). Because of its moderate-to-high levels of sensitivity and specificity, sentence recall has been proposed as a clinical marker of SLI for both children (Conti-Ramsden et al., 2001) and adults (Poll, Betz, & Miller, 2010).

Many of the aforementioned studies have included TD children who are well matched to those with SLI. Matched control groups are an important aspect of these studies because sentence recall has been shown to correlate with a child's age and a number of different psycholinguistic abilities, such as nonverbal IQ (Boyle, Lindell, & Kidd, 2013). Sentence recall has also been shown to be influenced by a child's dialect. Beyer and Hudson Kam (2011) found that 45 first- and second-grade children who spoke African American English (AAE) made significantly more alterations to 10 sentences in a recall task than did children who spoke General American English. In fact, when all alterations were counted as errors, no AAE child speaker earned a sentence recall score that was within 2 *SD* of those who spoke General American English. Charity, Scarborough, and Griffin (2004) reported a similar finding in their study of 217 AAE-speaking children in grades K through 2. Their stimuli included 15 sentences that were scored for 37 phonological or morphosyntactic structures. When all alterations

[a]Louisiana State University, Baton Rouge
Correspondence to Janna B. Oetting: cdjanna@lsu.edu

Editor and Associate Editor: Mabel Rice
Received January 30, 2015
Revision received May 10, 2015
Accepted June 9, 2015
DOI: 10.1044/2015_JSLHR-L-15-0036

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

to these structures were counted as errors, only 34% of the children recalled at least 75% of the items correctly. Last, J. M. Terry, Jackson, Evangelou, and Smith (2010) examined the effects of scoring modifications to allow for AAE-appropriate responses on the sentence recall subtest of the Clinical Evaluation of Language Fundamentals–Third Edition (Semel, Wiig, & Secord, 1995). Although the scoring modifications were found to increase the 45 AAE-speaking children's totals, scores were still lower than expected on items that contained verbal –s marking (e.g., *likes*, *lives*, *helps*), a structure that is infrequent in AAE. From this finding, the authors argued that an added cognitive load is present within tests when children are given sentences that contain grammatical structures infrequent in their dialect.

In the current work, we asked if sentence recall is a good diagnostic tool for English dialect speakers whose nonmainstream grammatical productions overlap with those that are produced at high rates by children with SLI. To do this, we selected matched groups of children with and without SLI who spoke one of two nonmainstream dialects of English, AAE or Southern White English (SWE), and implemented dialect-strategic scoring. As background, we review previous studies that show structural overlap between AAE, SWE, and the grammatical condition of SLI. Next, we look at studies that reveal important rate-based differences of the structures presenting the overlap. Findings from these studies suggest that rate-based differences of grammatical structure help differentiate AAE from SWE and, within these dialects, those with SLI from those without. Findings from these studies provide the motivation for controlling for a child's dialect and implementing dialect-strategic scoring when evaluating the diagnostic usefulness of sentence recall.

It is important to expand sentence recall research to nonmainstream dialect speakers. Although many of the sentence recall studies have focused on monolingual English language learners in the United States, studies have been conducted in a number of different countries, such as Australia (Petruccelli et al., 2012), Canada (Archibald & Joanisse, 2009), Ireland (Frizelle & Fletcher, 2014a, 2014b, 2015), and the United Kingdom (Briscoe et al., 2001; Conti-Ramsden et al., 2001, Riches, 2012), with languages other than English, such as French (Thordardottir et al., 2011), Cantonese (Stokes, Wong, Fletcher, & Leonard, 2006), and Czech (Smolík & Vávrů, 2014), and with bilingual language learners, including Spanish–English bilinguals (Ebert, 2014) and Swedish–Finnish bilinguals (Westman, Korkman, Mickos, & Byring, 2008). Noticeably absent are sentence recall studies (and clinical marker research in general) that have been designed for, and tested with, children and adults who speak different nonmainstream dialects of various languages. This is unfortunate because children and adults who speak nonmainstream dialects, especially if they live in disadvantaged areas, often present with less access to quality education and health care. Limited information about the clinical markers of SLI within these communities further increases the risk for misdiagnosis and long-term negative outcomes.

Structural Overlap Between AAE, SWE, and the Grammatical Condition of SLI

As with other nonmainstream dialects of English, AAE and SWE are similar in that they allow grammatical structures to be marked in three ways. Using forms of *be* as an example, these marking options include mainstream overt expressions (i.e., *I am*, *you are*, *she is*, *I was*, *they were*), nonmainstream overt expressions (i.e., *they is*, *we was*), and nonmainstream zero expressions (e.g., *he Ø*, *you Ø*). In the AAE and SWE varieties studied here, and as shown by the examples, the nonmainstream overt expressions are limited to *are* and *were* contexts, and the nonmainstream zero marked expressions are limited to *is* and *are* contexts (Roy, Oetting, & Moland, 2013).[1] On the basis of these findings, AAE and SWE can be described as presenting structural overlap that makes it impossible to determine if utterances such as *They was driving* or *You Ø coming with me* were produced by an AAE or SWE speaker.

Structural overlap in AAE and SWE is not limited to forms of *be* or to a small number of grammatical structures. Instead, almost all of the structures that allow nonmainstream marking in AAE also appear in SWE (for evidence with 35 different nonmainstream grammatical structures, see Oetting & McDonald, 2002). Significant amounts of structural overlap are also well attested when other dialects are compared with each other (for British vs. American English, see Tottie, 2009; for Canadian vs. American French, see Dubois, 2002). In other words, structural overlap is a universal property of dialect variation.

Overlap also exists between the structures produced within nonmainstream dialects of English, such as AAE and SWE, and the structures produced by children with SLI. This structural overlap was recognized more than 30 years ago by the American Speech-Language-Hearing Association (1983), and it is still identified as a major barrier when determining if a child presents a language difference (i.e., dialect difference) or a language disorder (Seymour, Bland-Stewart, & Green, 1998; see also recent test development by Seymour, Roeper, & de Villiers, 2003, 2005). Indeed, a number of structures that are produced in AAE and SWE, especially those involving zero marked forms (i.e., regular and irregular past, regular and irregular verbal –s, *be* and *do*), are also well known to be characteristic of children with SLI (for AAE and SWE, see Oetting & McDonald, 2002; Oetting & Pruitt, 2005; for children with SLI, see Leonard, 2014). Again, this structural overlap makes it impossible to determine if utterances such as *You Ø coming*, *He walkØ*, and *Where Ø I put this* were

[1]Patterns of *be* marking in the AAE and SWE varieties studied here align with many others that have been documented for AAE- and SWE-speaking adults. For English dialects in the United States that do not show these same patterns of *be* marking, such as Gullah/Geechee-influenced child AAE in South Carolina and elderly adult speakers of Pamlico Sound English in North Carolina, see Richardson-Berry and Oetting (2013) and Wolfram and Thomas (2002), respectively.

produced by an AAE or SWE TD speaker or by a child with SLI.

Rate-Based Differences Between AAE, SWE, and the Grammatical Condition of SLI

Although speakers of AAE and SWE produce many of the same nonmainstream grammatical structures, they differ in the rates at which they produce these structures. For example, Roy et al. (2013) found rates of zero *is* and zero *are* for 6-year-olds to average 43% and 69%, respectively, within AAE and 6% and 28%, respectively, within SWE. Rate-based differences between AAE and SWE are also not limited to forms of *be* or a small number of grammatical structures. On the contrary, the rate-based differences across structures are so robust that with no other information, Oetting and McDonald (2001) accurately classified 97% of their sample as speakers of either AAE or SWE.

Within AAE and SWE, children with and without SLI also differ in the frequency at which they produce nonmainstream grammatical structures. Using no information other than rate-based measures of grammatical structure, Oetting and McDonald (2001) accurately classified 90% of their sample as members of either the SLI or TD group. Within that study, the children with SLI were also found to produce some of the nonmainstream structures more frequently than the TD controls, omitting auxiliary *do* and zero marking irregular past tense three times more often than the controls and zero marking *be* twice as often.

At least five other studies have revealed rate-based differences between AAE- and/or SWE-speaking children with and without SLI for some (but not all) nonmainstream structures examined (Cleveland & Oetting, 2013; Garrity & Oetting, 2010; Oetting & Garrity, 2006; Oetting & Newkirk, 2008; Seymour et al., 1998). Within these studies, rates of use have been calculated as the percentage of overtly marked forms, which shifts the number of nonmainstream zero marked forms to the denominator (i.e., overtly marked forms/[overtly marked forms + zero marked forms]). Some of the structures showing clinical group differences in AAE and/or SWE include regular past tense (AAE SLI = 50% < TD = 91%; Seymour et al., 1998), auxiliary forms of *am*, *is*, and *are* (AAE SLI 24% < TD 47%; Garrity & Oetting, 2010), regular and irregular verbal –s (SWE but not AAE SLI = 64% < TD = 89%; Cleveland & Oetting, 2013), and relative clause markers (AAE and SWE SLI = 63% < TD = 82%; Oetting & Newkirk, 2008).

From these findings, we view the rate-based differences between dialects as reflections of typical language variation and the rate based-differences between children with and without SLI within dialects as reflections of atypical or impaired variation. As such, we reasoned that a sentence recall task would need to control for a child's dialect and involve dialect-strategic scoring, crediting only those nonmainstream responses that are not sensitive to grammatical impairment. We also hypothesized that in AAE and SWE, those with SLI would make more dialect-inappropriate errors than TD controls when recalling linguistically complex sentences. Examples of these types of errors might include omissions or substitutions of content words, such as nouns and verbs, and ungrammatical alterations of words within clauses (e.g., *new rap a song* for the target *a new rap song*). Previous sentence recall studies have shown that children with SLI produce fewer grammatical recalls than controls, making errors they don't always produce in conversation and with errors spanning content words, function words, and inflections (Frizelle & Fletcher, 2014b; Riches, 2012; Seeff-Gabriel et al., 2010; Smolík & Várnů, 2014; but see also Abel, Rice, & Bontempo, 2015, who found similar proportions of grammatical recalls for children with SLI and controls when stimuli involved simple syntax). The current study was designed to further test this hypothesis in the context of AAE and SWE.

Research Questions

Using matched SLI and TD groups of AAE and SWE speakers and dialect-strategic scoring, the research questions guiding the study were (a) do children with SLI earn lower sentence recall scores than controls, (b) are the sentence recalls of children with SLI less grammatical than those of controls, and (c) when children with SLI produce grammatical errors, what is the nature of their error relative to those produced by controls?

Method

Participants

Participants were 106 children (52 boys, 54 girls) who completed a sentence recall task as part of a larger study about children's grammar. All of the participants lived in a rural area in southeastern Louisiana and attended a public kindergarten with 90% or more of the enrolled students receiving free lunch. Seventy participants were classified as African American (AA), and 36 non–AA (34 White [W], one Asian, one American Indian) on the basis of school records. Caregiver report confirmed the race classifications for 99 children. Those whose race was not confirmed by a caregiver included four whose race was unspecified and one whose race was reported as mixed. The participants' ages ranged from 59 to 74 months, and their maternal education levels ranged from 6 to 17+ years (i.e., sixth grade to beyond a college degree).

Verifying Dialect Status

All of the AA children were classified as AAE speakers, and all of the non–AA children were classified as SWE speakers on the basis of two measures, the dialect portion of the Diagnostic Evaluation of Language Variation–Screener Test (DELV-ST; Seymour et al., 2003) and blind listener judgments of conversational speech (Oetting & McDonald, 2002) with the children's full language samples also explored in a few cases. The DELV-ST and listener judgments to classify and describe a child's dialect have been used in multiple studies (e.g., DELV-ST: Champion, Rosa-Lugo, Rivers, & McCabe, 2010; N. P. Terry, Connor, Petscher, & Conlin, 2012; listener judgments: Cotrell, Williams, Talley, & Taran, 2012; Garrity & Oetting, 2010; Horton-Ikard & Weismer, 2005; Pruitt & Oetting, 2009) and led to complementary information about a child's dialect (Horton & Apel, 2014).

The dialect portion of the DELV-ST includes 15 items and allows a child's English dialect to be classified as mainstream or nonmainstream. A percentage of nonmainstream responses out of the total scorable responses on the DELV-ST can also be calculated for each child. Listener judgments allow a child's English dialect to be classified by type (AAE vs. SWE vs. other). The listener judgment task involves three trained raters who classify each child's dialect after listening to a 1-min excerpt of conversational speech from a language sample. During the task, the raters are blind to the race and clinical status of the children, and the excerpts are void of any content (e.g., names, skin color) that could influence their dialect judgments.

All but three children (two W girls, one W boy) produced at least one nonmainstream response on the DELV-ST. These three children were retained as SWE speakers because they were identified as SWE speakers on the listener judgment task, and they produced a nonmainstream grammatical structure in more than 10% of their utterances within their full language samples. The percentage of nonmainstream responses produced by the AA and W children on the DELV-ST averaged 83 (SD = 20) and 56 (SD = 30), respectively, $F(1, 104) = 29.78$, $p < .001$, $\eta^2 = .22$.

In addition, the dialects of 97 (92%) of the 106 children were classified by two of three raters as AAE or SWE, and these dialect classifications aligned with the children's AA versus non-AA status. Those whose dialects were not classified consistently with their AA or non-AA status were seven AA and two W children, all but one of whom produced a high (>50%) percentage of nonmainstream responses on the DELV-ST (M = 71, SD = 18, range = 23–100). These children were also retained as AAE and SWE speakers, respectively. The 92% rate of dialect classification from a 1-min listener judgment task is identical to what was reported by Oetting and McDonald (2002) for another group of AAE and SWE child speakers. This high, albeit not perfect, rate of agreement reflects both the robust perceptual differences that exist between child AAE and SWE and the inherent variability that exists within any community of speakers.

Determining Clinical Status

Fifty-three of the children were classified as SLI, and the others served as TD controls. All passed a hearing screening conducted by the school during the fall semesters when data were being collected. The participants' clinical status was determined through a review of standardized test scores and family/school histories. The tests, in the order of their importance for classification purposes, were the syntax portion of the Diagnostic Evaluation of Language Variation–Norm Referenced (DELV-NR; Seymour et al., 2005), Primary Test of Nonverbal Intelligence (PTONI; Ehrler & McGhee, 2008), Goldman-Fristoe Test of Articulation–Second Edition (GFTA-2; Goldman & Fristoe, 2000), and the Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4; Dunn & Dunn, 2007). Fifty-two of the children also completed the Test of Language Development–Primary: Fourth Edition (TOLD-P:4; Newcomer & Hammill, 2008), which was added in the final years of the study. Children classified as SLI earned standard scores ≤ −1 SD on the DELV-NR, ≥ −1.2 SD on the PTONI, and > −1 SD on the GFTA-2. Their standard scores ranged from −2.27 to 0.73 SD on the PPVT and −1.40 to 0.40 SD on the TOLD-P:4. Children classified as TD earned scores > −1 SD on the DELV-NR and scores ≥ −1.07 SD on all of the other tests. In other words, the DELV-NR, as the only norm-referenced test that has been designed to be dialect neutral, was considered first with the others used for further confirmation in the case of the PTONI and GFTA-2 or further description in the case of the PPVT-4 and TOLD-P:4.

Thirteen (25%) of the SLI group and none of the TD group were receiving services by a speech-language pathologist, and 24 (45%) of the SLI group and eight (15%) of the TD group presented a positive family history of speech, language, or reading impairment. Also, the 53 classified as SLI reflected 6% of the 834 kindergartners within the schools and 8% of the 669 kindergartners with signed consent forms during the 5 years of data collection. These rates of identification and prevalence are not inconsistent with others in the SLI literature (cf. for prevalence and rate of identification studies, see Tomblin et al., 1997; for family history studies, see Pruitt, Garrity, & Oetting, 2010; Rice, Haney, & Wexler, 1998).

Matching of the SLI and TD groups was completed for each dialect separately, first using age, then the PTONI, and then maternal education. Group participant profiles are presented in Table 1. A series of 2 (Dialect: AAE vs. SWE) × 2 (Clinical Status: SLI vs. TD) analyses of variance (ANOVAs) revealed no effects for age or PTONI with clinical status effects for maternal education, $F(1, 98) = 4.96$, $p = .028$, partial $\eta^2 = .05$; DELV-NR, $F(1, 102) = 328.8$, $p < .001$, partial $\eta^2 = .76$; GFTA-2, $F(1, 102) = 18.49$, $p < .001$, partial $\eta^2 = .15$; PPVT-4, $F(1, 102) = 122.32$, $p < .001$, partial $\eta^2 = .55$; and TOLD-P:4, $F(1, 48) = 118.61$, $p < .001$, partial $\eta^2 = .71$. Although a clinical status effect for maternal education was not desired, the magnitude of the effect was small, and this variable was considered less important than the children's PTONI scores, which did not differ between the groups. A dialect effect (AAE < SWE) was also documented for the GFTA-2, $F(1, 102) = 3.92$, $p = .05$, partial $\eta^2 = .04$, and PPVT-4, $F(1, 102) = 5.20$, $p = .03$, partial $\eta^2 = .05$. In both cases, the magnitude of the effect was small.

Materials

Each child completed a sentence recall task with 36 items. Although our concern here was whether a recall task is diagnostically useful within AAE and SWE, the task was also designed to examine different types of sentences by manipulating the number and type of functional categories within them. *Functional categories* are syntactic categories in which reside the closed-class morphemes expressing

Table 1. Participant profiles by dialect and clinical status.

	AAE (*n* = 70)				SWE (*n* = 36)			
	SLI (*n* = 35)		TD (*n* = 35)		SLI (*n* = 18)		TD (*n* = 18)	
Variables	***M* (*SD*)**	**Range**	***M* (*SD*)**	**Range**	***M* (*SD*)**	**Range**	***M* (*SD*)**	**Range**
Age[a]	66.94 (3.74)	61–74	65.60 (3.55)	60–71	65.72 (3.89)	60–71	66.61 (4.18)	59–74
Maternal education[b]	11.67 (2.27)	6–16+	13.27 (2.62)	8–16+	12.33 (2.87)	8–16+	13.17 (3.05)	6–18+
DELV-NR syntax[c]	4.83 (1.01)	3–7	10.00 (1.55)	8–14	4.78 (1.67)	1–7	10.39 (1.72)	8–14
PTONI[d]	93.69 (9.62)	82–125	98.09 (8.87)	84–117	96.50 (8.35)	84–112	98.28 (8.14)	84–114
GFTA-2[e]	104.49 (5.72)	89–113	107.00 (4.38)	92–113	104.78 (4.18)	98–112	110.50 (3.09)	105–116
PPVT-4[f]	82.34 (9.42)	66–111	101.06 (9.32)	85–117	85.78 (7.01)	72–101	105.56 (5.62)	89–113
TOLD-P:4[g]	79.74 (6.48)	70–94	104.85 (7.66)	94–119	80.92 (5.39)	70–88	109.00 (9.54)	100–119

Note. AAE = African American English; SWE = Southern White English; SLI = specific language impairment; TD = typically developing; DELV-NR = Diagnostic Evaluation of Language Variation–Norm Referenced; PTONI = Primary Test of Nonverbal Intelligence; GFTA-2 = Goldman-Fristoe Test of Articulation–Second Edition; PPVT-4 = Peabody Picture Vocabulary Test–Fourth Edition; TOLD-P:4 = Test of Language Development–Primary: Fourth Edition.

[a]Reported in months. [b]Reported in years of schooling (i.e., 12 = high school graduate, 16+ = college graduate or more, with data missing for four children). [c]Standardized scores for the syntax portion of the DELV-NR (normative *M* = 10, *SD* = 3). [d]Standardized scores for the PTONI (normative *M* = 100, *SD* = 15). [e]Standardized scores for the GFTA-2 (normative *M* = 100, *SD* = 15). [f]Standardized scores for the PPVT-4 (normative *M* = 100, *SD* = 15). [g]Standardized scores for the TOLD-P:4 (normative *M* = 10, *SD* = 3, with data missing for 54 children).

functional content, such as tense and agreement. The functional categories convey clausal phrase structure above the level of the verb phrase (for examples of syntactic models that recognize functional category structure in child language development, see Guilfoyle & Noonan, 1992; Hegarty, 2005; Radford, 1990; Vainikka, 1993). To do this, the stimuli involved equal numbers of items with one, two, and three functional categories, and the functional categories included tense, negation, and complementizer. Tense and complementizer are included in most models that recognize functional categories; negation is included in some but not all models (for some models that include negation, see Déprez & Pierce, 1993; Hegarty, 2005; Pollock, 1989, 1997).

The development of the stimuli was as follows. Twelve base sentences were created with equal numbers featuring one of four auxiliary *be* verbs (*is*, *are*, *was*, *were*); names of one, two, or three well-known characters (i.e., *Big Bird*, *Ernie*, *Bert*, *Mickey*, *Minnie*); and common nouns (e.g., *ball*, *cookie*, *dog*) and verbs (e.g., *run*, *bake*, *watch*). The complexity of the sentences was then manipulated to control for the number and type of possible functional categories per sentence. This yielded five different sentence types for each base sentence, resulting in a pool of 60 (5 × 12) different sentences. For each base sentence, sentence types featured either the functional category of tense (e.g., *Minnie was jumping on the big bed last night*); two functional categories of tense and negation (e.g., *Big Bird is not driving to the store downtown today*); or three functional categories of tense, negation, and complementizer (e.g., *Why were Bert and Ernie not cooking?*) (see Table 2). Sentence types with one and two functional categories were both nine words in length; sentences with three functional categories, however, were further broken into three different subtypes with two subtypes remaining at nine words in length (e.g., *Ernie wonders if Big Bird is not driving downtown today*, *Ernie wonders who is not driving to the store*) and the third subtype shortened to seven words (e.g., *Who is not driving to the store?*).[2] Differences between the three functional category sentences, aside from length, stem from the type and placement of the complementizer. Then, to reduce the possibility of fatigue by the children, three analogous versions of 36 sentences were drawn from the pool of 60. Sentences across versions were counterbalanced such that each version contained equal numbers of each sentence type (i.e., 12 with tense; 12 with tense and negation; 12 with tense, negation, and complementizer, with each participant getting four of each subtype of complementizer). The sentences were recorded by a southern AA female and presented through a PowerPoint slide show that played from a laptop computer.

Procedures

After institutional review board approval and parental consent were secured, testing began at the children's schools in an unoccupied area. Children were administered the sentence recall task across three separate days. For each day, two practice sentences and 12 experimental sentences were presented over the laptop's internal loudspeakers. Sentences were presented in sets of three, following a slide in which the main characters within those sentences were introduced (e.g., slide presented pictures of Mickey and Minnie while examiner said, "Let's try sentences with Mickey and Minnie"). The children's recalls were digitally recorded for later transcription. Each recall was transcribed by two examiners; a third examiner relistened to the recall if the two transcriptions did not match.

Scoring

The data were first analyzed using a dialect-strategic modification of the 2-point scoring system that was created

[2]*Big Bird* was counted as one word.

Table 2. Example items from the sentence recall task.

Category	Length	Example items
One Functional Category	9 words	
		Minnie is cleaning the dirty dishes in the sink.
		Bert and Ernie are singing a new rap song.
		Minnie was jumping on the big bed last night.
		Yesterday, Bert and Ernie were cooking a big hamburger.
Two Functional Categories	9 words	
		Minnie is not cleaning the dishes in the sink.
		Bert and Ernie are not singing a new song.
		Yesterday, Minnie was not jumping on the bed.
		Bert and Ernie were not cooking a big hamburger.
Three Functional Categories	7–9 words	
		Mickey wonders if Minnie is not cleaning the dishes.
		Big Bird wonders when Bert and Ernie are not singing.[a]
		Why were Bert and Ernie not cooking?

Note. Underlined text indicates the targeted functional categories within the sentences.

[a]*Big Bird* was counted as one word.

by Archibald and Joanisse (2009) and later utilized by Redmond et al. (2011). With this method, each sentence receives a score on the basis of the number of errors relative to the target. If a child repeats the target exactly, a score of 2 is given. If one to three errors are observed, a score of 1 is given. If four or more errors from the target are noted or if a child fails to respond, the sentence receives no credit. Errors include morpheme or word additions, substitutions, or omissions; alterations of phonology are ignored.

However, given that the children were speakers of either AAE or SWE, three dialect-strategic modifications were made to the scoring, and these included dialect-appropriate productions of *is* for third person plural present progressive verbs (e.g., *Bert and Ernie is singing a new rap song*), *was* for third person plural past progressive verbs (e.g., *Yesterday, Bert and Ernie was cooking a hamburger*), and zero marking of verbal –s (e.g., *Big Bird wonderØ when Bert and Ernie are not singing*). These three nonmainstream productions were selected for modification because they are frequently produced and have not been shown to be robust clinical markers of SLI in one or both of these dialects. Evidence to support this claim for nonmainstream productions of *was* can be found in Oetting and Garrity (2006). Evidence to support this claim for zero marking of verbal –s within AAE can be found in Cleveland and Oetting (2013). Although zero marking of verbal –s is infrequent in SWE and it has been shown to be a robust clinical marker of SLI in this dialect, it is not a robust marker in negative contexts involving the word *doesn't* (i.e., zero marking in SLI ~ 67% vs. controls ~ 56%; Cleveland & Oetting, 2013). In the current study, the 70 AAE-speaking children produced 584 tokens of nonmainstream *is*, nonmainstream *are*, and zero verbal –s, which corresponds to an average of 8.34 productions per child, and the 36 SWE-speaking children produced 174, which corresponds to an average of 4.83 per child. No other dialect modifications were made to the scoring.

Next, and with the dialect-strategic modification maintained, the children's responses were classified as exact recall, nonexact grammatical recall, ungrammatical recall, unscorable, or missing. The unscorable items included recalls with unintelligible content, which precluded judgments about grammaticality, and the missing items included no responses from the children and a very small number of items that had to be thrown out due to examiner error. An example of a grammatical recall was *Ernie and Bert are singing a song* for the target *Bert and Ernie are singing a new rap song*. Examples of ungrammatical recalls for this same target sentence were *Bert Ernie a rap song* and *Bert is not singing a new song and bird*. For each ungrammatical recall, the errors were then classified as related to either the targeted functional categories (i.e., tense, negation, complementizer) or other content. For this level of coding, two examiners classified the recalls, and disagreements were resolved through consensus. Examples of errors for each functional category include tense: *He Ø not jumping on the bed* for the target *Minnie was not jumping on the big bed*; negation: *They not are not singing a new song* for the target *Bert and Ernie are not singing a new song*; complementizer: *Mickey wonders Ø is not washing the dirty dishes* for the target *Mickey wonders who is not cleaning the dirty dishes*; and other content: *He's not jumping on Ø bed* for the target *Yesterday, Minnie was not jumping on the bed*. Note that two of the ungrammatical examples above (i.e., *He Ø not jumping on the bed* and *Mickey wonders Ø is not washing the dirty dishes*) are perfectly acceptable in adult and child AAE and SWE; however, because these examples included zero marked forms, which are produced more frequently by children with SLI than by TD controls within AAE and SWE, we counted them as errors.

Reliability

Reliability of the 2-point scoring system was evaluated by having a second examiner independently score the recalls of 40 children. Disagreements occurred 51 times out of 1,440 (40 × 36 sentences) opportunities for agreement, resulting in a 96% agreement rate. The prevalence of disagreements between the two examiners who scored the recalls for grammaticality and error category was evaluated by examining the two sets of coded recalls from 40 children. Disagreements between the two sets of coded recalls occurred on 58 sentences out of 1,440 (40 × 36 sentences) opportunities for agreement, resulting in a 96% agreement rate.

Results

Accuracy

Using the 2-point system, the maximum possible score on the sentence recall task was 72 (36 × 2 points; see Table 3). A 2 (dialect) × 2 (clinical status) ANOVA yielded a main effect of clinical status, $F(1, 102) = 123.33$, $p < .001$, partial $\eta^2 = .55$. The TD group ($M = 49.98$, $SD = 10.46$) had higher sentence recall scores than the SLI group ($M = 27.06$, $SD = 10.69$). There was no main effect of dialect, $F(1, 102) = 0.69$, $p = .41$, partial $\eta^2 = .01$, nor was there a significant interaction, $F(1, 102) = 2.27$, $p = .13$, partial $\eta^2 = .02$.

Considering all the participants and taking the maximum of Youden's J index gives a cut point of 40 to maximize the number of people correctly classified as SLI or TD using the sentence recall task. This yielded a sensitivity of .91 and specificity of .85, with 88% of the children correctly classified by their clinical status. It is interesting and important for looking at this task in the context of AAE and SWE that this same cut point also maximized the number of children correctly classified as SLI or TD within each dialect. For AAE speakers, a cut point of 40 yielded a sensitivity of .89 and specificity of .86, and for SWE speakers, the same cut point yielded a sensitivity of .94 and specificity of .83.

Grammaticality

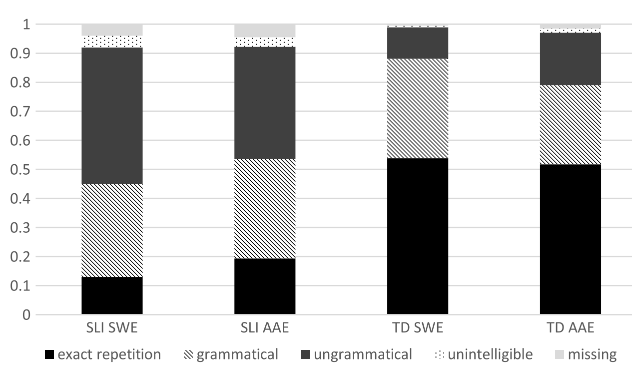
The 2-point scoring system is based on the number of errors made in the sentence recall task, but it does not address the grammaticality of the children's recalls. To do this, we classified the children's recalls as exact, not exact but grammatical, ungrammatical, unscorable, or missing. Recall that dialect-strategic scoring was maintained for this analysis, and the unscorable items included recalls with unintelligible content that precluded judgments about grammaticality. Figure 1 shows the proportion of each type of recall for the four groups. Exact recalls were higher for the TD groups than the SLI groups; grammatical nonexact recalls were approximately equal across the groups; and ungrammatical recalls tended to be greater for the SLI groups than the TD groups, as was also the case for the unscorable recalls.

Table 3. Sentence recall accuracy by dialect and clinical status (maximum is 72).

Clinical status	AAE	SWE	Total
SLI	28.77 (9.75)	23.72 (11.90)	27.06 (10.69)
TD	49.49 (10.84)	50.94 (9.90)	49.98 (10.46)
Total	39.13 (14.61)	37.33 (17.52)	38.52 (15.60)

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

Figure 1. Proportion of response types by dialect and clinical status. SLI = specific language impairment; SWE = Southern White English; AAE= African American English; TD = typically developing.



To examine these data, we calculated the proportion of ungrammatical recalls over all scorable recalls (i.e., exact recalls, grammatical recalls, and ungrammatical recalls). A 2 (dialect) × 2 (clinical status) ANOVA showed a main effect for clinical status, $F(1, 102) = 84.73$, $p < .001$, partial $\eta^2 = .45$, with the SLI group ($M = 0.45$, $SD = 0.21$) producing higher proportions of ungrammatical recalls than the TD group ($M = 0.16$, $SD = 0.12$). There was also an interaction between dialect and clinical status, $F(1, 102) = 6.46$, $p = .013$, partial $\eta^2 = .06$. A breakdown of the data into the two dialects showed that the clinical status group difference held for both the AAE speakers, $F(1, 68) = 34.03$, $p < .001$, partial $\eta^2 = .33$, and SWE speakers, $F(1, 34) = 48.40$, $p < .001$, partial $\eta^2 = .59$. The interaction reflects the fact that the difference tended to be smaller for the AAE speakers (SLI: $M = 0.41$, $SD = 0.19$ vs. TD: $M = 0.19$, $SD = 0.13$) than the SWE speakers (SLI: $M = 0.51$, $SD = 0.23$ vs. TD: $M = 0.11$, $SD = 0.08$).

We also examined the sensitivity and specificity of using these alternative scoring methods to differentiate children with and without SLI. If sensitivity and specificity are calculated on the basis of the number of exact repetitions, instead of the 2-point scoring system, very similar results occur. In fact, using an optimal cut point of 12 or fewer exact repetitions for classification of SLI, one more TD AAE speaker was correctly classified, increasing specificity to .87 for the AAE group. In addition, if sensitivity and specificity are calculated using the proportion of ungrammatical responses, an optimal cut point of .24 yielded a sensitivity of .89 and a specificity of .83 when considering the speakers of both dialects together. When speakers of the two dialects are considered separately, this cut point yielded slightly lower scores for the AAE speakers (sensitivity of .86 and a specificity of .80) than for SWE speakers (sensitivity of .94 and specificity of .89), but these values are still moderately high and consistent with those of other sentence recall studies. Thus, different types of scoring methods for the sentence recall task yield relatively good classification of the children, with the exact repetitions doing better than the proportion of ungrammatical responses and about the same as the 2-point system. In all cases, we see that the optimal cut point is the same across the dialects, indicating that group differences in clinical status are stable across dialects and scoring methods.

Nature of Ungrammatical Errors

To further examine the children's ungrammatical recalls, we noted whether the ungrammaticality was due to errors in one of the functional categories of tense, negation, or complementizer or to an error outside of these functional categories (see Table 4). A 2 (dialect) × 2 (clinical status) ANOVA on percentage of errors attributable to function categories yielded a main effect of clinical status, $F(1, 102) = 10.32$, $p = .002$, partial $\eta^2 = .09$. The SLI group ($M = 0.53$, $SD = 0.16$) made a higher proportion of functional category errors than the TD group ($M = 0.37$, $SD = 0.30$), which also means, conversely, that the TD group made more other errors ($M = 0.59$, $SD = 0.31$) than the SLI group ($M = 0.47$, $SD = 0.16$). There was no main effect for dialect and no interaction.

Further investigation into the functional category errors showed that there was a main effect for clinical status for tense errors, $F(1, 102) = 13.87$, $p < .001$, partial $\eta^2 = .12$. The SLI group ($M = 0.44$, $SD = 0.17$) made a higher proportion of tense errors than the TD group ($M = 0.27$, $SD = 0.27$). There was also a main effect for negative errors, $F(1, 102) = 5.14$, $p = .025$, partial $\eta^2 = .05$. Again, the SLI group ($M = 0.04$, $SD = 0.06$) made a higher proportion of negative errors than the TD group ($M = 0.01$, $SD = 0.04$). However, there were no differences between the SLI group ($M = 0.05$, $SD = 0.07$) and TD group ($M = 0.09$, $SD = 0.16$) in terms of proportion of complementizer errors, $F(1, 102) = 2.18$, $p = .14$, partial $\eta^2 = .02$.

Next, we further investigated the functional category errors by breaking them down by sentence type (one functional category: tense only; two functional categories: tense and negation; and three functional categories: tense, negation, and complementizer) in order to compare the relative number of each error type across the items when given an equal number of opportunities for error (see Table 5). For one functional category sentences, we ran a 2 (dialect) × 2 (clinical status) ANOVA on the proportion of tense errors.

Table 4. Proportion of functional category errors versus other errors by dialect and clinical status.

	AAE		SWE	
Errors	**SLI**	**TD[a]**	**SLI**	**TD[a]**
Functional category errors	.53 (.15)	.37 (.26)	.53 (.18)	.37 (.37)
Other errors	.47 (.15)	.60 (.27)	.47 (.18)	.57 (.39)

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

[a]One AAE TD speaker and one SWE TD speaker produced no ungrammatical utterances and had zeroes entered for the proportion of functional category errors and other errors.

Table 5. Proportion of error type for ungrammatical recalls by functional category, dialect, and clinical status.

	AAE		SWE	
Functional category	**SLI**	**TD**	**SLI**	**TD**
One functional category				
Tense	.38 (.27)	.10 (.24)	.38 (.29)	.03 (.12)
Two functional categories				
Tense	.40 (.35)	.22 (.33)	.40 (.38)	.08 (.26)
Negation	.01 (.03)	.02 (.08)	.004 (.02)	.00 (.00)
Three functional categories				
Tense	.42 (.16)	.29 (.29)	.47 (.20)	.24 (.33)
Negation	.05 (.10)	.02 (.04)	.06 (.09)	.01 (.05)
Complementizer	.09 (.10)	.10 (.13)	.10 (.13)	.12 (.25)

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

Here there was a main effect for clinical status, $F(1, 102) = 40.87$, $p < .001$, partial $\eta^2 = .29$, with the SLI group ($M = 0.38$, $SD = 0.27$) making a higher proportion of tense errors than the TD group ($M = 0.08$, $SD = 0.21$).

For two functional category sentences, we ran a mixed ANOVA with dialect and clinical status as between-subjects measures and functional category (tense and negation) as the repeated measure. Here there was a main effect of clinical status, $F(1, 102) = 12.05$, $p < .001$, partial $\eta^2 = .11$, with more errors made by the SLI group than the TD group, and a main effect of functional category, $F(1, 102) = 60.29$, $p < .001$, partial $\eta^2 = .37$, with more errors on tense than negation. Both of these main effects were qualified by an interaction with clinical status and functional category, $F(1, 102) = 13.89$, $p < .001$, partial $\eta^2 = .12$. The interaction arises because, although the SLI group made more errors on tense ($M = 0.40$, $SD = 0.36$) than did the TD group ($M = 0.17$, $SD = 0.31$), they did not differ on errors of negation.

For three functional category sentences, again run as a mixed ANOVA with functional category (tense, negation, and complementizer) as the repeated measure, there was a main effect of clinical status, $F(1, 102) = 14.35$, $p < .001$, partial $\eta^2 = .12$, with more errors made by the SLI group than the TD group. With Greenhouse-Geisser correction used due to lack of sphericity, there was also a main effect of functional category, $F(1.4, 146) = 82.67$, $p < .001$, partial $\eta^2 = .44$. A post hoc Bonferroni test showed there were more errors on tense ($M = 0.35$, $SD = 0.26$) than on complementizers ($M = 0.10$, $SD = 0.15$), which in turn had more errors than negation ($M = 0.03$, $SD = 0.08$). These main effects were qualified by an interaction with clinical status and functional category, $F(1.4, 146) = 7.48$, $p < .001$, partial $\eta^2 = .07$. Although both tense (SLI: $M = 0.44$, $SD = 0.17$; TD: $M = 0.27$, $SD = 0.30$) and negation errors (SLI: $M = 0.05$, $SD = 0.09$; TD: $M = 0.01$, $SD = 0.05$) were higher for the SLI group than the TD group, complementizer errors did not differ between the groups (SLI: $M = 0.09$, $SD = 0.11$; TD: $M = 0.11$, $SD = 0.18$). Thus, across all three types of sentences, ungrammaticality due to tense errors was statistically higher for the children with SLI than for the TD controls. Ungrammaticality due to negation errors only differed between the SLI and TD groups for sentences with three functional categories.

Last, we examined if the children's tense and negation errors were influenced by the number of functional categories present in the sentences. For tense errors (which were present across all three levels of functional categories), there was a main effect of number of functional categories, $F(2, 204) = 7.29$, $p < .001$, partial $\eta^2 = .07$. Post hoc Bonferroni tests showed that children made more errors when there were three functional categories than when there was one. Although there was no interaction with clinical status, separate analyses of these groups found this effect to hold for only the TD group. For negative errors (which were present across two levels of functional categories), there was again a main effect of number of functional categories, $F(1, 102) = 7.71$, $p = .007$, partial $\eta^2 = .07$, and an interaction between number of categories and clinical diagnosis, $F(1, 102) = 6.57$, $p = .012$, partial $\eta^2 = .06$. Errors increased with more categories for the SLI but not the TD group.

Discussion

Studies have repeatedly shown sentence recall to differentiate children with SLI from TD controls and from children who present with other types of clinical conditions, but these studies have not included speakers of dialects whose nonmainstream grammatical productions overlap with those that are produced at high rates by children with SLI. The current study was designed to evaluate the diagnostic utility of a sentence recall task in AAE and SWE, two dialects of English that show this overlap. To examine sentence recall within the context of these dialects, we followed the methods of previous sentence recall studies that have selected well-defined groups of children with SLI and well-matched groups of TD controls. We did this by not only selecting and matching children on the basis of their ages, sociodemographic/educational backgrounds, and psycholinguistic testing profiles, but also by selecting TD controls who spoke the same dialect as the children with SLI. We also implemented a dialect-strategic scoring system that recognized differences between dialect-appropriate productions that are sensitive to the grammatical condition of SLI and dialect-appropriate productions that are not.

Using previous child studies of AAE and SWE as a guide, scoring modifications for the children's dialects were strategically limited to three types of dialect-appropriate productions that have not been shown to be robust markers of SLI within these dialects. These productions included *is* for third person plural present progressive verbs, *was* for third person plural past progressive verbs, and zero marking of verbal –s. No other dialect modifications were made because we wanted the scoring system to be sensitive to rate-based grammar differences between children with SLI and TD controls within AAE and SWE.

Results from a number of analyses showed sentence recall to be diagnostically useful for differentiating children with SLI from TD controls within the two dialects studied. In both AAE and SWE, a difference between the children with SLI and the TD controls was found when the 2-point scoring system was implemented, and the same clinical group difference was found when we examined the children's proportion of ungrammatical recalls. For the 2-point system, scores of the children with SLI were lower than controls, and for the proportion of ungrammatical recalls, their scores were higher. In both cases, effects sizes (i.e., partial eta squared values) of the group differences were large, ranging from .45 to .55. These effect sizes can be interpreted as showing clinical status as a variable accounting for ~50% of the variance in the children's sentence recall scores that was not accounted for by the variance explained by the children's dialects or their interaction.

Sentence recall was also found to be diagnostically accurate when empirically derived cut points were used to classify the children as SLI or TD. Overall diagnostic accuracy was 88% for both the 2-point system and the number of exact repetitions with cut points of 40 and 12, respectively, and it was 86% for the proportion of ungrammatical utterances with a cut point of .24. These rates of diagnostic accuracy compare favorably with those reported in other studies in which total percentage of children correctly classified by a sentence recall task has ranged from 87% to 90% (Archibald & Joanisse, 2009; Conti-Ramsden et al., 2001; Poll et al., 2010; Redmond et al., 2011).

For studies conducted within the context of dialect diversity, it is important to point out that the diagnostic accuracy of the sentence recall task, with accompanying values of sensitivity and specificity, were highly similar for the AAE and SWE speakers. Of the three scoring methods (i.e., 2-point system, number of exact repetitions, proportion of ungrammatical recalls), we recommend either the 2-point system or the number of exact repetitions because they led to the highest levels of diagnostic accuracy. These two scoring systems are also easier to calculate than the proportion of ungrammatical recalls. Nevertheless, across all three scoring methods, the empirically derived cut points were the same for the two dialects, and across both dialects and scoring methods, levels of sensitivity and specificity were at or above .80.

The final set of analyses examined the types of errors made by the children when their recalls were ungrammatical. Results from these analyses indicated that those with SLI, regardless of their dialect, made proportionally more errors with structures involving the functional categories of tense, negation, and complementizer as compared with the TD controls. Of the three types of functional categories, those with SLI, regardless of their dialect, produced proportionally more errors with grammatical structures involving tense than the others. This finding held when we examined all of the sentences together and when we examined the relative frequencies of each type of error on the basis of the number of opportunities to make an error within the sentences.

Errors involving negation also led to differences between the children with SLI and the TD controls, but this finding was less stable than the findings for tense, surfacing only when all of the sentences were considered together and when negation was examined in sentences that contained three functional categories. When the sentences contained only two functional categories (i.e., tense and negation), a clinical group difference for negation was not observed. Although not reviewed here, grammatical errors with negation are rarely if ever reported in the SLI literature when language sample data are analyzed (for review of studies, see Leonard, 2014). Recall also that other sentence recall studies by Frizelle and Fletcher (2014b), Riches (2012), Seeff-Gabriel et al. (2010), and Smolík and Várnů (2014) have documented grammatical errors by children with SLI that have not always been documented in language sample data. Given this, future studies are needed to fully flesh out the types of errors children with SLI and TD controls make on sentence recall tasks and the effects that different types of sentence complexity manipulations have on the frequency and nature of their errors. Methods used within studies by Frizelle and Fletcher (2014a) and others can be used to guide these studies.

In the future, we plan to examine possible trade-offs children make when they spontaneously produce utterances and/or are asked to recall sentences with grammatical structures from multiple functional categories. According to at least one model of functional category acquisition, that by Hegarty (2005), the expression of multiple functional categories within clauses may stress immature or impaired language systems in ways that do not always lead to omissions or errors of one type of functional category over others. Instead, systematic trade-offs may be evident such that the production of one or two functional categories within a clause may lead to omissions or errors of others and vice versa. To complete this type of work with sentence recall data, grammatical omissions will need to be treated differently than errors of commission. We will also have to examine children's nonexact, grammatical recalls to account for simplification strategies that reduce the need for complex syntax. The current coding for exact repetition and grammaticality of nonexact repetition does not necessarily capture if a functional category is omitted (e.g., omission of a negative and simplification of a complex sentence with a complementizer still results in a grammatical nonexact repetition). Thus, the data will need to be recoded to test such hypotheses. Again, work by Frizelle and Fletcher (2014a) may be useful for guiding this type of study.

Although the focus of this research forum article is on the diagnostic usefulness of sentence recall for children who speak nonmainstream dialects of English, it is important to note that sentence recall measures children's abilities in both language and cognitive processing. Thus, it is not clear if a child's inability to recall sentences is due to deficits in language, cognitive processing, or both. In the current study, deficits related to language may help explain why the functional category of tense was more difficult for the SLI group than the TD group. However, deficits in cognitive processing may help explain why the children's tense and negation errors increased as the number of functional categories increased within the sentences. Consonant with this latter finding, past work has shown that children's sentence recall scores have been correlated to measures of short-term memory (forward digit recall: Frizelle & Fletcher, 2015; Riches, Loucas, Baird, Charman, & Simonoff, 2010), working memory involving verbal and nonverbal, auditory materials (competing language processing: Poll et al., 2013; listening span: Frizelle & Fletcher, 2015; Riches, 2012; nonverbal, auditory recall: Ebert, 2014), phonological short-term memory (nonword repetition: Ebert, 2014; Riches, 2012), and speed of processing (Poll et al., 2013). Thus, sentence recall may be diagnostically useful for identifying children with SLI because it is sensitive to children's deficits in both language and cognitive processing.

Conclusion and Clinical Implications

When matched groups are examined and dialect-strategic scoring is used, sentence recall yields moderate-to-high levels of diagnostic accuracy to identify SLI in AAE and SWE, two nonmainstream dialects of English that are spoken in the United States. These levels of diagnostic accuracy are similar to what has been reported for experimentally controlled studies of children who speak mainstream varieties of English and who live in a number of different countries.[3] The clinical implication of this finding is significant because the field lacks diagnostic tools and measures that can be administered to children who speak nonmainstream dialects of English in these countries.

An important aspect of the current study was the use of well-matched groups of same dialect–speaking SLI and TD groups. Although the current findings need to be replicated with larger groups of children, larger, community-based studies are needed for all experimentally controlled studies of SLI. Findings from the current study indicate that these larger, community-based studies should not be limited to mainstream dialect speakers. Instead, these studies should be expanded to include all children with added layers of description and control to accommodate the dialect diversity that exists within communities.

In the current work, the use of dialect-strategic scoring was also critical for ensuring that children were not penalized for producing dialect-appropriate productions that are not sensitive to the grammar weaknesses of children with SLI. Given the success of our dialect-strategic scoring system, we recommend that this type of system be seriously studied and considered to replace those that treat all nonmainstream dialect productions as correct. As demonstrated by the literature review and the results of the current study, some nonmainstream dialect productions, especially those

[3]Our assumption is that the varieties of English that have been spoken by the participants in previous sentence recall studies have been relatively mainstream for the communities and countries in which they have been recruited; however, the dialects of these participants have not been fully described by the authors. Dialect information within studies would facilitate the cross-linguistic/cross-dialectal study of SLI and lead to more rigorous tests of hypotheses about this clinical condition.

related to functional categories, are sensitive to the grammatical condition of SLI. Treating these types of productions as universally correct, which is typically recommended in many commercially available language tests, most likely reduces the power to identify children with SLI within the context of a nonmainstream dialect. Dialect-strategic scoring offers an alternative to these existing scoring systems.

Acknowledgments

Funding for this study was provided through National Institute on Deafness and Other Communication Disorders Grant RO1DC009811, awarded to Janna B. Oetting, Janet L. McDonald, and Michael Hegarty. Appreciation is extended to Jessica Berry, Rebecca Hammarlund, Kyomi Gregory, Ryan Lee, Karmen Porter, Andrew Rivière, Tina Villa, and a number of current and past MA students who helped create the stimuli and collect the data. We would also like to thank the teachers, families, and children who agreed to participate in the study.

References

Abel, A. D., Rice, M. L., & Bontempo, D. E. (2015). Effects of verb familiarity on finiteness marking in children with SLI. *Journal of Speech, Language, and Hearing Research, 58,* 360–372.

American Speech-Language-Hearing Association. (1983). *Social Dialects and Implications of the Position on Social Dialects: Position Statement*. Retrieved from http://www.asha.org/policy

Archibald, L. M. D., & Alloway, T. P. (2008). Comparing language profiles: Children with specific language impairment and developmental coordination disorder. *International Journal of Language & Communication Disorders, 43,* 165–180.

Archibald, L. M. D., & Joanisse, M. F. (2009). On the sensitivity and specificity of nonword repetition and sentence recall to language and memory impairments in children. *Journal of Speech, Language, and Hearing Research, 52,* 899–914.

Beyer, T., & Hudson Kam, C. (2011). First and second graders' interpretation of Standard American English morphology across varieties of English. *First Language, 32,* 365–384.

Boyle, W., Lindell, A. K., & Kidd, E. (2013). Investigating the role of verbal working memory in young children's sentence comprehension. *Language Learning, 63,* 211–242.

Briscoe, J., Bishop, D. V. M., & Norbury, C. F. (2001). Phonological processing, language, and literacy: A comparison of children with mild-to-moderate sensorineural hearing loss and those with specific language impairment. *The Journal of Child Psychology and Psychiatry, 42,* 329–340.

Champion, T. B., Rosa-Lugo, L. I., Rivers, K. O., & McCabe, A. (2010). A preliminary investigation of second- and fourth-grade African American students' performance on the Gray Oral Reading Test–Fourth Edition. *Topics in Language Disorders, 30,* 145–153.

Charity, A. H., Scarborough, H. S., & Griffin, D. M. (2004). Familiarity with school English in African American children and its relation to early reading achievement. *Child Development, 75,* 1340–1356.

Cleveland, L. H., & Oetting, J. (2013). Verbal –s marking by dialect and clinical status. *American Journal of Speech-Language Pathology, 22,* 604–614.

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment. *The Journal of Child Psychology and Psychiatry, 42,* 741–748.

Cotrell, E., Williams, S., Talley, K., & Taran, V. (2012, April). *Comparing DELV scores with listener judgments of AAE*. Paper presented at the convention of the Black Association for Speech-Language and Hearing, Raleigh, NC.

Déprez, V., & Pierce, A. (1993). Negation and functional projections in early grammar. *Linguistic Inquiry, 24,* 25–67.

Dubois, S. (2002). French language's status and preservation in Louisiana, USA, and in the Maritime Provinces, Canada. In W. Zacharasiewicz & P. Kirsch (Eds.), *Aspects of interculturality–Canada and the United States: A comparison* (pp. 123–137). Vienna, Austria: ISL-Verlag.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test–Fourth Edition*. Bloomington, MN: The Psychological Corporation.

Eadie, P. A., Fey, M. E., Douglas, J. M., & Parsons, C. L. (2002). Profiles of grammatical morphology and sentence recall in children with specific language impairment and Down syndrome. *Journal of Speech, Language, and Hearing Research, 45,* 720–732.

Ebert, K. D. (2014). Role of auditory non-verbal working memory in sentence repetition for bilingual children with primary language impairment. *International Journal of Language & Communication Disorders, 49,* 631–636.

Ehrler, D. J., & McGhee, R. L. (2008). *Primary Test of Nonverbal Intelligence*. Austin, TX: Pro-Ed.

Frizelle, P., & Fletcher, P. (2014a). Profiling relative clause constructions in children with specific language impairment. *Clinical Linguistics & Phonetics, 28,* 437–449.

Frizelle, P., & Fletcher, P. (2014b). Relative clause constructions in children with specific language impairment. *International Journal of Language & Communication Disorders, 49,* 255–264.

Frizelle, P., & Fletcher, P. (2015). The role of memory in processing relative clauses in children with specific language impairment. *American Journal of Speech-Language Pathology, 24,* 47–59.

Garrity, A. W., & Oetting, J. B. (2010). Auxiliary BE production by African American English-speaking children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research, 53,* 1307–1320.

Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation–Second Edition*. Circle Pines, MN: AGS.

Guilfoyle, E., & Noonan, M. (1992). Functional categories and language acquisition. *Canadian Journal of Linguistics, 37,* 241–272.

Hegarty, M. (2005). *A feature-based syntax of functional categories: The structure, acquisition and specific impairment of functional systems*. Berlin, Germany: Mouton de Gruyter.

Horton, 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,* 448–460.

Horton-Ikard, R., & Weismer, S. E. (2005). Distinguishing African-American English from developmental errors in the language production of toddlers. *Applied Psycholinguistics, 26,* 597–620.

Leonard, L. (2014). *Children with specific language impairment*. Cambridge, MA: MIT Press.

Newcomer, P. L., & Hammill, D. D. (2008). *Test of Language Development–Primary: Fourth Edition*. Austin, TX: Pro-Ed.

Oetting, J. B., & Garrity, A. W. (2006). Variation within dialects: A case of Cajun/Creole influence within child SAAE and SWE. *Journal of Speech, Language, and Hearing Research, 49,* 16–26.

Oetting, J. B., & McDonald, J. L. (2001). Nonmainstream dialect use and specific language impairment. *Journal of Speech, Language, and Hearing Research, 44,* 207–223.

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,* 508–518.

Oetting, J. B., & Newkirk, B. L. (2008). Subject relatives by children with and without SLI across different dialects of English. *Clinical Linguistics & Phonetics, 22,* 111–125.

Oetting, J. B., & Pruitt, S. L. (2005). Use of Southern African American English across groups. *International Journal of Multicultural Communication Disorders, 3,* 136–144.

Petruccelli, N., Bavin, E. L., & Bretherton, L. (2012). Children with specific language impairment and resolved late talkers: Working memory profiles at 5 years. *Journal of Speech, Language, and Hearing Research, 55,* 1690–1703.

Poll, G. H., Betz, S. K., & Miller, C. A. (2010). Identification of clinical markers of specific language impairment in adults. *Journal of Speech, Language, and Hearing Research, 53,* 414–429.

Poll, G. H., Miller, C. A., Mainela-Arnold, E., Adams, K. D., Misra, M., & Park, J. S. (2013). Effects of children's working memory capacity and processing speed on their sentence imitation performance. *International Journal of Language & Communication Disorders, 48,* 329–342.

Pollock, J. (1989). Verb movement, universal grammar, and the structure of IP. *Linguistic Inquiry, 20,* 365–424.

Pollock, J. (1997). Notes on clause structure. In L. Haegeman (Ed.), *Elements of grammar* (pp. 237–279). Dordrecht, the Netherlands: Kluwer Academic.

Pruitt, S. L., Garrity, A. W., & Oetting, J. B. (2010). Family history of speech and language impairment in African American children. *Topics in Language Disorders, 30,* 154–164.

Pruitt, S. L., & Oetting, J. B. (2009). Past tense marking by African American English-speaking children reared in poverty. *Journal of Speech, Language, and Hearing Research, 52,* 2–15.

Radford, A. (1990). *Syntactic theory and the acquisition of English syntax: The nature of early child grammar of English*. Oxford, England: Oxford University Press.

Redmond, S. M. (2005). Differentiating SLI from ADHD using children's sentence recall and production of past tense morphology. *Clinical Linguistics & Phonetics, 19,* 109–127.

Redmond, S. M., Thompson, H. L., & Goldstein, S. (2011). Psycholinguistic profiling differentiates specific language impairment from typical development and from attention-deficit/hyperactivity disorder. *Journal of Speech, Language, and Hearing Research, 54,* 99–117.

Rice, M. L., Haney, K. R., & Wexler, K. (1998). Family histories of children with SLI who show extended optional infinitives. *Journal of Speech, Language, and Hearing Research, 41,* 419–432.

Richardson-Berry, J., & Oetting, J. (2013, November). *Examining the performance of child AAE and Gullah speakers on the DELV-ST*. Paper presented at the annual American Speech-Language-Hearing Association Convention, Chicago, IL.

Riches, N. G. (2012). Sentence repetition in children with specific language impairment: An investigation of underlying mechanisms. *International Journal of Language & Communication Disorders, 47,* 499–510.

Riches, N. G., Loucas, T., Baird, G., Charman, T., & Simonoff, E. (2010). Research report: Sentence repetition in adolescents with specific language impairments and autism: An investigation of complex syntax. *International Journal of Language & Communication Disorders, 45,* 47–60.

Roy, J., Oetting, J. B., & Moland, C. (2013). Linguistic constraints on children's overt marking of BE by dialect and age. *Journal of Speech, Language, and Hearing Research, 56,* 933–944.

Seeff-Gabriel, B., Chiat, S., & Dodd, B. (2010). Sentence imitation as a tool in identifying expressive morphosyntactic difficulties in children with severe speech difficulties. *International Journal of Language & Communication Disorders, 45,* 691–702.

Semel, E., Wiig, E., & Secord, W. A. (1995). *Clinical Evaluation of Language Fundamentals–Third Edition*. San Antonio, TX: The Psychological Corporation.

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

Seymour, H. N., Roeper, T., & de Villiers, J. G. (2003). *Diagnostic Evaluation of Language Variation Screening Test*. San Antonio, TX: The Psychological Corporation.

Seymour, H. N., Roeper, T., & de Villiers, J. G. (2005). *Diagnostic Evaluation of Language Variation: Norm-Referenced Test*. San Antonio, TX: The Psychological Corporation.

Smolík, F., & Várnů, P. (2014). Sentence imitation as a marker of SLI in Czech: Disproportionate impairment of verbs and clitics. *Journal of Speech, Language, and Hearing Research, 57,* 837–849.

Stokes, S. F., Wong, A. M-Y., Fletcher, P., & Leonard, L. B. (2006). Nonword repetition and sentence repetition as clinical markers of specific language impairment: The case of Cantonese. *Journal of Speech, Language, and Hearing Research, 49,* 219–236.

Terry, J. M., Jackson, S. C., Evangelou, E., & Smith, R. L. (2010). Expressive and receptive language effects of African American English on a sentence imitation task. *Topics in Language Disorders, 30,* 119–134.

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,* 55–69.

Thordardottir, E., Kehayia, E., Mazer, B., Lessard, N., Majnemer, A., Sutton, A., . . . Chilingaryan, G. (2011). Sensitivity and specificity of French language and processing measures for the identification of primary language impairment at age 5. *Journal of Speech, Language, and Hearing Research, 54,* 580–597.

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research, 40,* 1245–1260.

Tottie, G. (2009). How different are American and British English grammar? And how are they different? In G. Rohdenburg & J. Schlüter (Eds.), *One language, two grammars? Differences between British and American English* (pp. 341–363). New York, NY: Cambridge University Press.

Vainikka, A. (1993). Case in the development of English syntax. *Language Acquisition, 3,* 257–325.

Westman, M., Korkman, M., Mickos, A., & Byring, R. (2008). Language profiles of monolingual and bilingual Finnish preschool children at risk for language impairment. *International Journal of Language & Communication Disorders, 43,* 699–711.

Wolfram, W., & Thomas, E. (2002). *The development of African American English*. Malden, MA: Blackwell.