



University of
Massachusetts
Amherst

Dialect-neutral indices of narrative cohesion and evaluation.

Item Type	article;article
Authors	Burns, Frances A;de Villiers, Peter A;Pearson, Barbara Zurer;Champion, Tempii B
DOI	http://DOI: 10.1044/0161-1461(2011/10-0101)
Download date	2026-04-13 17:56:01
Link to Item	https://hdl.handle.net/20.500.14394/552

Article

Dialect-Neutral Indices of Narrative Cohesion and Evaluation

Frances A. Burns,^a Peter A. de Villiers,^b Barbara Z. Pearson,^a and Tempii B. Champion^c

AQ1 Purpose: This study compared the development of essential elements of narrative skill in children from African American English (AAE)- and general American English (GAE)-speaking communities using an innovative elicitation and evaluation protocol consisting of four key indices of narrative language: (a) reference contrasting, (b) temporal expressions, (c) mental state descriptions, and (d) understanding of behavior based on false belief.

Method: Participants were 291 AAE speakers and 238 GAE speakers, 4 to 9 years of age. Approximately one-third of both dialect groups were identified as having language impairments. Children generated 2 stories based on short picture sequences. Their stories were coded for the 4 key indices of narrative

AQ2 language. Analyses of variance were performed with subsets

of the measures and a composite index with all measures combined as outcomes; and with age, dialect group, and clinical status as predictors.

Results: Age and clinical status had statistically significant effects on the subset measures and the composite score. Variation between AAE and GAE dialect was not a significant factor.

Conclusion: By focusing on dialect-neutral elements of narratives—creating links across sentences and providing mental state interpretations—this study adds to our knowledge of development and impairment in narrative production among both AAE- and GAE-background children.

Key Words: dialect, narrative, AAE, assessment, cohesion

The narrative abilities of children in prekindergarten and elementary school have been recognized as important indicators of their language level generally and as a key to their school readiness (Justice, Bowles, Pence, & Gosse, 2010; Pankratz, Plante, Vance, & Insalaco, 2007; Snow, Tabors, Nicholson, & Kurland, 1995). Telling stories is a developmentally appropriate and naturalistic activity (Schraeder, Quinn, Stockman, & Miller, 1999) that is found in all cultures (Campbell, 1949). Still, its particular forms are dependent on culturally specific social contexts (Champion, 1997; Gutierrez-Clellen & Quinn, 1993; Michaels, 1981) and can be influenced by different literacy socialization practices (Heath, 1983; Matsuyama, 1983; Melzi, 2000). Therefore, stories from children who are growing up in different cultures may differ and may require assessment

that is culturally sensitive. Although some recent work has focused on narrative development in children with Hispanic and African American (AA) backgrounds (Champion, 2003; Fiestas & Peña, 2004; Gorman, Fiestas, Pena, & Clark, 2011; Gutierrez-Clellen & Quinn, 1993), most developmental research on narratives has been conducted on populations from European American (EA) backgrounds (Gee, 1986; Hester, 1996). Best practices have not yet been established for assessing narratives from a broader range of children, including AAs (Champion, 2003). In this article, we take a step toward developing those best practices by presenting an overview of current narrative assessment practices; selecting candidate measures for efficient, dialect-neutral assessment; and then testing their effectiveness with a large sample of general American English (GAE)-speaking and African American English (AAE)-speaking children.

Linguistically/Culturally Neutral Versus Linguistically/Culturally Specific Aspects of Narrative Language

The study of the narrative skills of AA children has typically focused on thematic coherence, or organization at the *macro* level of the overall plot of stories (Burns, 2004; Champion, Seymour, & Camarata, 1995; Hyon & Sulzby, 1994; Michaels, 1981, 1986). Conceptual frameworks focused

^aUniversity of Massachusetts, Amherst

^bSmith College, Northampton, MA

^cLong Island University, Brooklyn, NY

Correspondence to Frances A. Burns: fburns@comdis.umass.edu

Editor: Marilyn Nippold

Associate Editor: Amy Weiss

Received November 17, 2010

Revision received January 24, 2011

Accepted August 2, 2011

DOI: 10.1044/0161-1461(2011/10-0101)

on plot structure, such as *story grammars* (Labov, 1972; Mandler, 1982) and *high-point analysis* (Peterson & McCabe, 1983; Stein & Glenn, 1979, among others), are well-established methods for analyzing the hierarchical organization of spoken narratives. However, these discourse-level structures, which describe the story as a whole, have been shown to vary across cultures and languages (Bliss, Covington, & McCabe, 1999; Gee, 1986; Matsuyama, 1983), and some researchers advise using alternative and/or additional methods for assessing the narrative skills of culturally and linguistically diverse (CLD) groups (Champion, 1997; Crago, Eriks-Brophy, Pesco, & McAlpine, 1997; Fiestas & Peña, 2004).

Recent work by P. de Villiers (2004; P. de Villiers et al., 2010) suggests that analysis at the *micro* level of narrative may be linguistically and culturally neutral. Micro-level analysis looks *within* individual sentences for elements that establish local relationships between words or elements of the discourse (Berman & Slobin, 1994; Justice et al., 2010). Using Halliday and Hasan's (1976) basic framework for analysis of types of linguistic cohesion in narratives, P. de Villiers et al. (2010) focused on the two essential categories of cohesion that are observed most often in children's narratives: (a) *reference relations*, which are used when introducing, contrasting, and maintaining the separate identities, especially of characters in the discourse; and (b) the expression of causal and *temporal conjunctive links* between events that tie clauses together, whether within or across sentences (Bamberg, 1986; Berman, 1988; Berman & Slobin, 1994; Hickmann, 2003).

Reference relations as cohesive devices. Reference contrasting is a very basic form of cohesion. Typically, the narrator begins by orienting or "setting the scene" for the listener, introducing the characters and identifying the location and time of events (Hickmann, 2003; Labov, 1972). In order for the unfolding action to be comprehensible, the narrator must maintain clear links throughout the story to distinguish the different characters from each other by using proper names or other noun-phrase elaborations like contrasting adjectives (e.g., *the big girl* and *the little girl*) or prepositional phrases (e.g., *the boy with the train*).

Temporal conjunctive links as cohesive devices. Temporal conjunctive links also play a key role in narratives. Constructing a narrative involves describing a series of events that are related in time (Dasinger & Toupin, 1994; Labov, 1972). By relating two or more actions with an adverbial conjunction that foregrounds one and subordinates the other, the narrator can highlight some actions and place others in the background (Hickmann, 2003). Adverbial conjunctions can be single-word sequencers, like *later* or *next*, or full clauses, like *when the girl came back*. (See examples associated with the sample stories in Appendix B.)

AQ3

Horton-Ikard (2009) noted that few studies assessing cohesion have been undertaken with CLD populations. Horton-Ikard also cautioned that certain aspects of the AAE

linguistic system may impact the measures that are used in the assessment. To avoid the potential negative impact of particular AAE grammatical forms like those identified by Horton-Ikard, P. de Villiers (2004) proposed that assessments focused on *functions* can direct attention away from what might be considered nonstandard speech from the point of view of academic GAE, for example, the kind of AAE structures discussed in Wolfram (1991) or Washington and Craig (1998). P. de Villiers suggested evaluating whether cohesion has, in fact, been achieved—whether characters are successfully contrasted and events are ordered in time—but ignoring the actual pronouns used, such as *hims* rather than *his*, for example, and giving the same credit for a relative clause with or without a complementizer *that* or *which*.

Narrative "evaluation" as a component of micro-level analysis. However, as P. de Villiers (2004) pointed out, even focusing on function rather than linguistic forms, micro-level cohesive structures alone may be too limited to distinguish story quality in children's productions across a wide age range. He suggested that a broader perspective on narrative maturity for both younger and older elementary school-children may be gained by pairing micro-level analysis of cohesion with a focus on the child's appreciation of mental state information relevant to the story—what its characters say, want, intend, or think. Such mental state information is a subset of what Labov (1972) referred to as narrative evaluation and what Bruner (1986) called "the landscape of consciousness."

AQ4

Evaluative comments tell about what is going on in the minds of the characters and provide interpretations of the events from outside the line of the plot. That is, clauses that contain evaluation in this sense of the word stand outside the main narrative clauses in what Labov (1972) called "free clauses." The evaluative information is key to making sense of the landscape of action, but it cannot be portrayed directly in picture prompts as can objects and actions in the scene. Rather, elements of evaluation must be inferred by the narrator from clues in the picture that highlight them for the viewer or are supplied by narrators from their own understanding of how people relate to each other and to objects and events (Berman & Slobin, 1994). Children's appreciation of the thoughts and motives of the characters—their theory of mind (ToM; Astington, 1993)—indicates a deeper level of understanding of the events in the story and the needs of their listeners. Importantly, as discussed in the Berman and Slobin (1994) volume, elements of evaluation demonstrate that narrators are reflecting on their own stories.

AQ5

Combining elements of cohesion and evaluation. The key role of cohesion and evaluation in children's developing narratives was recognized by Pellegrini (1984) and again by Westby (1991, 1999), who included them in her construct of "literate language." Westby bundled mental and linguistic verbs—used in evaluation—with micro-level cohesive elements like noun-phrase elaboration, adverbs, and conjunctive ties, and proposed a composite measure that she called

AQ6

AQ7

literate language as an indicator of story quality. Westby's literate language has been a useful framework for helping people understand essential aspects of developing narratives outside of story grammar organization (e.g., Curenton & Justice, 2004; Greenhalgh & Strong, 2001). Such an assessment is also much simpler to perform than macro-structure analysis. Micro-level structures can be analyzed phrase by phrase from within sentences, without reference to how they relate to the text as a whole. The evaluation components of literate language, which are found in free clauses outside the main organization of the plot, can be recognized without having to analyze the many elements that make up the plot of the story.

AQ8 Following Westby's proposal to combine Labov-type evaluation with noun-phrase elaboration and conjunctive links, P. de Villiers and his colleagues, Seymour, Roeper, and J. de Villiers (reported in P. de Villiers, 2004) streamlined the assessment process even further and devised a narrative protocol that specified just four of the elements Westby proposed, namely, two micro-structural elements—reference contrasting and temporal expressions—and two aspects of evaluation—mental state descriptions and a standard task involving a false belief. The protocol P. de Villiers devised with his colleagues included stimuli that improved the elicitation of these elements and permitted them to be noted down while the story was being told. It could be used by people with minimal knowledge about narrative structures, an important feature also pointed out by Justice et al. (2010). P. de Villiers (2004) argued that an assessment based on these four indices together exhibits the three crucial qualities of a satisfactory measure: development, discrimination, and dialect neutrality. The present study reports a test of those claims.

Previous Studies Showing Development, Discrimination, and Dialect Neutrality

FN1 of Cohesion in Narratives¹

Development in cohesion. In young children's stories, referential links are often *exophoric*, or deictic links, in which the referring expression points outside the discourse to the world or social situation (e.g., *that* or *this*) (Halliday & Hasan, 1976). As children's stories become more mature and decontextualized, deictic links are replaced by *anaphoric* expressions, which refer backward or forward to another linguistic element within the conversation or text. Anaphors achieve referential cohesion through the use of proper names or syntactic elements such as simple noun phrases, and especially through elaborated noun phrases containing pronouns, definite and indefinite articles, adjectives (the *tall* girl), prepositional phrases (*on the sofa*), and relative clauses (The girl *who was eating a cracker*) (Dasinger & Toupin,

1994). Westby (1991) demonstrated that these more complex anaphoric links, especially in elaborated noun phrases with prepositional phrases or relative clauses, are seen increasingly in the narratives of children as they get older.

Similarly with conjunctions, Berman and Slobin (1994) demonstrated in their extensive cross-linguistic, cross-sectional study of narratives based on picture support that, for the most part, children younger than 5 years old did not structure their stories temporally and failed to create explicit temporal connections between the pictures. Instead, each picture frame had equal weight, with no distinction in the importance of each event. In the narratives of older children, there was a clear progression in how the children expressed temporal relationships between the events, from no time information before age 5, building up to the emergence among the 9-year-olds of subordinated adverbial time clauses (e.g., *before he came* or *while he was eating*) (Berman, 1988; Berman & Slobin, 1994). Like Berman (1988), Hickmann (2003) demonstrated how both single adverbs and adverbial phrases and clauses that specify the setting and sequence of events contributed to the increasing well-formedness of children's stories.

Discriminating impairment with cohesion. Liles (1985, 1987) was one of the first researchers to explore the effect of language impairment (LI) on cohesion in two empirical studies of 7- to 10-year-olds, half of whom were typically developing (TD) and half who had LI. She adapted Halliday and Hasan's (1976) classification system to examine the children's references to characters and events in the children's spoken retellings of a short movie. The children with LI used significantly fewer correct anaphoric pronouns (those linking back to previous words in the discourse) compared to their TD peers. Children in the LI group were also more likely than TD children to substitute lexical repetition and demonstrative or deictic forms (e.g., *that*). Similarly, TD children's conjunctive links were significantly more accurate across sentences and episodes of the story than those of the LI children. In general, the children with LI produced less effective cohesive devices such that listeners could not easily comprehend the semantic ties that were being attempted between elements of the narrative.

Greenhalgh and Strong (2001) also compared cohesion in stories from children who differed in clinical status. In their study, fifty-two 7- to 10-year-olds with LI and 52 age-matched TD children listened to stories the authors wrote for four of Mayer's wordless storybooks (Mayer, 1967, 1969, 1974; Mayer & Mayer, 1975). Children followed the stories in the picture books and then, without the pictures in front of them, retold the stories to an uninformed listener. Analyzing the frequency of particular narrative features based on Westby's construct of literate language in the children's story retellings, Greenhalgh and Strong found that the children with LI produced significantly fewer elaborated noun phrases and fewer adverbial conjunctions per clause than the TD children did.

AQ9

¹A summary table of references in this section can be found in Appendix A.

Dialect neutrality and cohesion. Cohesion has been shown to be neutral between AAE and GAE. Curen-ton and Justice (2004) reported nearly identical mean rates of occurrence of literate language features in the story transcripts of AA and EA children in their sample of 3- to 5-year-olds. Elaborated noun phrases and conjunctions, especially, occurred relatively frequently in the productions of both ethnic groups. In another study with older elementary school-children, Horton-Ikard (2009) found that AA children used the same categories of cohesive ties as their EA peers. Although some AA children’s referential cohesive ties could be identified morphosyntactically as “AAE features” (Wolfram, 1991), they performed their cohesive function regardless of their form. Similarly, as part of a curriculum intervention study with preschoolers, P. de Villiers et al. (2010) compared the narratives of AA 3- to 5-year-olds who were learning AAE as a first dialect with those of EA peers who were learning GAE as a first dialect. The participants were administered two short narrative tasks adapted from the same protocol that was used in this study (as described in P. de Villiers, 2004). Children completed the task at the beginning of the school year and again ~7 months later. P. de Villiers et al.’s results showed that the dialect difference between the two groups was not a significant factor in performance outcomes on the narrative tasks.

Development, Discrimination, and Dialect Neutrality in Free Clauses, or Evaluation

Development in evaluation. Just as for reference contrast-ing and temporal expressions, a developmental progres-sion has been established for the evaluation component in narrative, especially for ToM and false beliefs. A child’s understanding of a person’s actions as they relate to that person’s desires, thoughts, or mental states—their ToM—is a crucial development for improving children’s narratives (Tager-Flusberg & Sullivan, 1995). According to Bartsch and Wellman (1995), children’s earliest references to the mental states are about simple emotions or desires (e.g.,

happiness, sadness, fear, and anger)—what they want and like. In studies by Perner, Leekam, and Wimmer (1987) and Wellman and Bartsch (1994), TD children were shown to talk about the beliefs of others by age 4. Likewise, in the various language populations and ages studied by Berman and Slobin (1994), younger children typically failed to relate detailed information about the cognitive and emotional states of characters. Only later, between ages 7 and 9, did chil-dren consistently talk about the characters’ *cognitions* (i.e., what they believed, thought, or knew).

In addition to children’s descriptions of mental states, which are a hallmark of their ToM, children’s ability to appreciate and talk about false beliefs also shows develop-ment in this age range. Unlike the expression of true beliefs, false beliefs are an unambiguous test of children’s ability to separate their own beliefs from those of the characters (Bartsch & Wellman, 1995; J. de Villiers, Roeper, Bland-Stewart, & Pearson, 2008). The development of understand-ing false beliefs has been studied in the context of predicting a character’s response to a complex *wh*-question (J. de Villiers et al., 2008; Pearson & Ciolli, 2004). For example, one type of complex sentence embeds a false clause inside a true one, as in, *She said she bought bananas* (when in fact she bought a new bicycle). To answer *What did she say she bought?* a key item in the study reported in J. de Villiers et al. (2008), one must put aside one’s own knowledge of the truth and report on what was said.

Because true beliefs represent actual facts, they are gen-erally shared by both the narrator and the protagonist, and so one cannot easily identify true beliefs as belonging to a character, as opposed to the speaker. By contrast, because false beliefs of the story characters are almost always distinct from the child’s own thoughts and reactions (as in the exam-ple with bananas and a bicycle), they are thus more infor-mative about a narrator’s developmental stage than are true beliefs. Pearson and Ciolli (2004, Table 1) reported a loose progression in typical development. At age 4, almost half of the children in their study responded that a character who heard the complex sentence would answer with their own

Table 1. Participants by age, dialect, and clinical status.

Dialect	Age in years	Clinical status						Total	% Female	Mean PED
		4	5	6	7	8	9			
AAE	LI	16	29	33	8	12	7	105	36	1.8
	TD	45	42	69	8	11	11	186	58	1.9
GAE	LI	17	24	23	3	12	8	87	30	2.0
	TD	33	42	42	11	11	12	151	58	2.2
Total sample		110	137	167	30	46	38	529	49	2.01

Note. PED = parent education; AAE = African American English speaker; GAE = General American English speaker; LI = language impaired; TD = typically developing. PED level scale: 1 = <12 years of schooling, 2 = high school degree, 3 = some college.

knowledge (*She bought a bicycle*). By age 6, they found that 80% of the TD children subordinated the buying event to the saying event, and they responded with what she said. Likewise, 90% of 8-year-olds responded, *She said "bananas,"* revealing a developmental pattern.

AQ15 **Discrimination and evaluation.** Deficits in the evaluative component of narratives in children with LI have also been reported in the literature (Baron-Cohen, Leslie, & Frith, 1985; Bruner & Feldman, 1993; Tager-Flusberg & Sullivan, 1995). Notably, like narratives of younger children, those of children with LI typically pay little attention to the characters' thoughts (Bamberg & Damrad-Frye, 1991; Berman & Slobin, 1994). As suggested by J. de Villiers and Pyers (1997, 2002), the developmental lag in ToM and evaluation more generally in children with LI may stem from difficulties with the syntax required to express other people's thoughts. The child with LI also may have decreased facility in conversation (Leonard, 1998), which Harris (1996) suggested is a context where children can develop their understanding of the viewpoints of others. According to Harris, informational exchanges during conversation, which include requests for information, rejecting or denying, informing, and others, promote ongoing predictions about what the partner will understand. These exchanges also promote repairs and clarifications when the predictions are wrong. Harris proposed that even children without other impairments but who do not have strong experiences of conversation will show a deficit in ToM.

An example of such a TD group with reduced opportunity for conversation is children who are deaf whose parents have normal hearing. Peterson and Siegal (2000) corroborated Harris' (1996) claim with a meta-analysis of 11 studies of children with no other clinical impairment except deafness. The otherwise TD children who were deaf demonstrated considerably delayed ToM development. According to Peterson and Siegal, deafness alone did not predict delays in ToM development. Their study showed that other non-hearing children who grew up in homes with native sign-language speakers had fluent conversational partners (i.e., in sign language), and they did not experience delay on ToM tasks.

Greenhalgh and Strong (2001) also examined aspects of evaluation with their LI and TD groups. However, they found no significant differences between the two groups in their measure, which looked at mental state verbs (e.g., *want*, *think*, or *know*) and linguistic verbs (e.g., *say* and *tell*). The authors speculated that the failure to find differences in the measure that encompassed both types of verbs was due to the nature of the stories they used, which did not lend themselves to consideration of the characters' mental states, even by TD children. (See also Pearson, 2001, for a similar conclusion about one of the same frog stories Greenhalgh and Strong used.) Another possible reason the measure Greenhalgh and Strong used did not discriminate between the TD and LI groups may be that only *linguistic* verbs showed no

difference between Greenhalgh and Strong's groups. They did not test the mental state verbs separately. The less challenging linguistic verbs may have masked a potential difference in the more demanding mental state verbs (Bartsch & Wellman, 1995), and so mental state verbs would have needed to be evaluated separately.

Dialect neutrality and evaluation. Barring LI or an impoverished social environment that restricts access to the universal activity of conversation (Stivers et al., 2009), the ability to take the perspective of another person should not be bound to any particular culture, language, or dialect. Nonetheless, past research has suggested that AA children from low-income backgrounds use fewer mental state verbs as compared to their EA counterparts (Holmes, Black, & Miller, 1996). If this is accurate, evaluation also might be culturally bound and would place AA children at a disadvantage if mental state verbs are used as a part of a narrative assessment tool.

More recent analyses, however, of AA children's and parents' use of mental verbs revealed no differences in the usage rates compared to EA families (Allen, de Villiers, & François, 2001; Burns, 2004). Allen et al. (2001) followed the suggestion found in J. de Villiers and Pyers (1997) that only mental and linguistic verbs *with tensed complements* (e.g., *He thought she was going downtown*) provide a linguistic route to a ToM, and so they reanalyzed a large corpus of AA and EA professional and working class spontaneous speech samples (Hall, 1984). Raw counts of mental and linguistic verbs showed the expected advantage of the non-AA groups for overall use of these verbs (Allen et al., 2001, Figure 1). However, when fixed expressions and routinized opinion markers like *I don't know* or *I think I'll have an apple* were eliminated, the comparison changed substantially. Allen et al. found that AA and EA families used similar numbers of tensed complements with these verbs, and some of the comparisons showed AA families, both parents and children, using more of them than EA families. This study indicates, then, that observations of mental verb understanding and use have the potential to be a part of dialect-neutral language assessment for AA children.

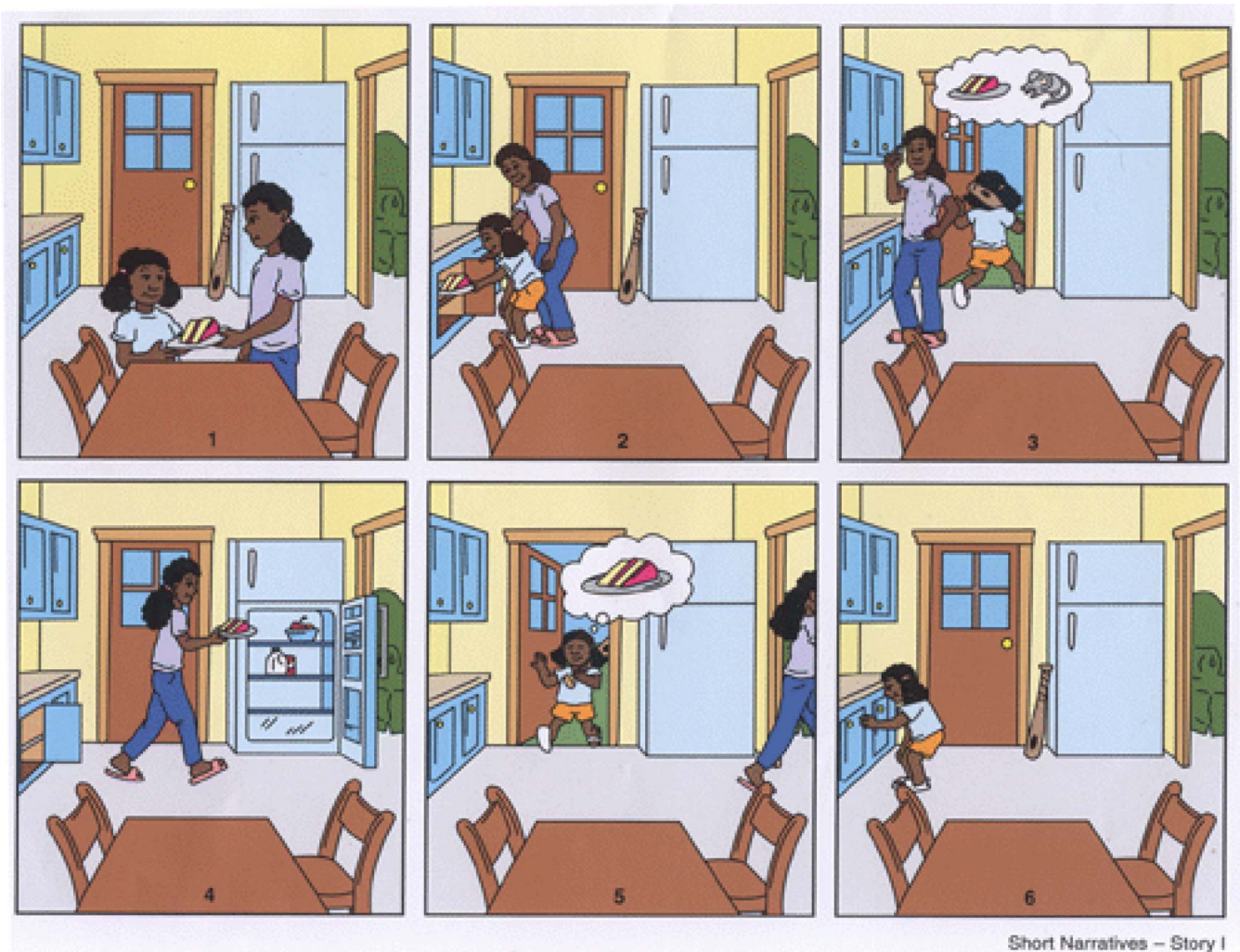
Choice of Indices for Assessing Narrative Cohesion and Evaluation

Especially since the landmark work of Berman and Slobin (1994), several schemas have been proposed for standardizing narrative assessment, most notably, The Test of Narrative Language (Gilliam & Pearson, 2004), the Edmonton Narrative Norms Instrument (Schneider, Dube, & Hayward, 2005), the Index of Narrative Micro-Structure (INMIS; Justice et al., 2006), and very recently, the Narrative Assessment Protocol (NAP; Justice et al., 2010). All of these measures provide valuable information about children's narratives, but except for the NAP, they all require the time and ability to record and listen to or transcribe the multiple stories involved. Furthermore, dialect neutrality has not

AQ16

FIG1

AQ17

Figure 1. Sample six-picture narrative sequence from the Dialect-Sensitive Language Test (Seymour, Roper, & de Villiers, 2000).

Note. Data analysis results from the Dialect-Sensitive Language Test. Copyright 2000 by NCS Pearson, Inc. Reproduced with permission. All rights reserved.

been established for these tests; thus, their applicability to CLD populations is limited.

AQ18 As described earlier, the four elements in the protocol devised by P. de Villiers in conjunction with Seymour, Roper, and J. de Villiers (described in P. de Villiers, 2004) are all likely candidates for efficient, dialect-neutral narrative assessment. P. de Villiers and his colleagues pointed to elaborated noun phrases that serve to make the character and object descriptions more explicit; conjunctions, like *while* and *when*, that tie sentences together and establish the sequence of events; single adverbs that specify such features as the time of the action (e.g., *later*); and mental state verbs (e.g., *wanted*, *thought*, or *decided*) that express the protagonists' motivations and cognitions.

P. de Villiers' (2004) elements share many similarities with Westby's (1991) construct of literate language. Like

Westby's literate language, cohesion and evaluation together in oral narratives highlight discourse elements that are associated more with literacy than oracy. That is, literate language is not necessarily written (Gee, 1985, 1986), as long as the oral discourse contains elements that are characteristic of written texts (Perera, 1986). Thus, the metric described by P. de Villiers (2004) and illustrated in Appendix B would also have utility for classroom teachers and literacy coaches.

The Present Study

The purpose of the present study was to compare the development of essential elements of narrative skill in children from AAE- and GAE-speaking communities. Using an innovative elicitation and evaluation paradigm, we aimed to discover to what extent the path of development of the

four elements in P. de Villiers' (2004) protocol was the same for the two dialect groups, and to what extent AAE-speaking children with LI differed from AAE TD and GAE TD speakers on those elements. Our research questions targeted the following three factors:

- *Development.* Will robust developmental growth in narrative cohesion and evaluation be observed across the age range of 4 to 9 years?
- *Discrimination.* Will children with LI be distinguished from TD children on these measures across this age range among both AAE and GAE speakers?
- *Dialect neutrality.* Will growth and discrimination patterns for these linguistic and evaluative elements be similar in children who are learning AAE as a first dialect and those who are learning GAE as a first dialect?

AQ19

METHOD

Participants

A nationwide sample of 529 children, ages 4 to 9 years, participated in this study. The children were a subset of the sample that took part in the field testing of the Dialect Sensitive Language Test (DSLST; Seymour, Roeper, & de Villiers, 2000), which is a precursor to and superset of the Diagnostic Evaluation of Language Variation tests (Seymour, Roeper, & de Villiers, 2003, 2005). (The tests are described below under Materials.) Children were recruited by certified speech-language pathologists (SLPs) to fulfill the following criteria: They had normal vision and hearing and were within the normal range on standard tests of intelligence. Some of the participants were AAE speakers, and others were GAE speakers. Both dialect groups contained TD children and children with LI, as defined below. To ensure that the sample represented a broad selection of AAE speakers and comparable GAE-speaking peers, the selection criteria were very stringent. Before enrolling a child in the study, SLPs collected demographic and linguistic information from parents and teachers, which was reviewed by a recruitment supervisor who was also a certified SLP.

Clinical status. To be included as TD, there were two criteria that had to be satisfied: The children were categorized by the testing clinicians as TD and were not receiving speech-language services in their schools or communities; and their performance on the diagnostic section, Part 2, of the DELV—Screening Test (DELV—ST, Seymour et al., 2003) fell in the *very low* or *low risk* categories.²

FN2

In order to be classified as LI, there were two criteria: Children had been categorized by the testing clinicians as LI

²The elements analyzed for this criterion are independent of the narrative items that are the focus of this article.

through the standard methods of their practice, and they were receiving speech-language services in their schools or communities; and as confirmation of the clinicians' designation, the children's performance on the diagnostic section, Part 2, of the DELV—ST fell into the *highest risk for language disorder* category. Children who fulfilled these criteria were then categorized by dialect or language variety.

Language variation status (LVS). Participants in this study were from two linguistic communities, AAE-first-dialect and GAE-first-dialect speakers. All children in the AAE group met these three criteria: (a) They were AA by parent report; (b) they lived in a community of predominantly AAs by census report (U.S. Bureau of the Census, 2000); and they were characterized by their performance on the dialect identifier section, Part 1, of the DELV—ST as having *some* or *strong* variation from GAE.

Children in the GAE-first-dialect group met analogous criteria: (a) They were predominantly EA background by parent report; (b) they lived in communities that did not have a predominance of AA residents; and (c) they were characterized by their performance on the dialect identifier section, Part 1, of the DELV—ST as *having no variation from GAE*. Thus, the AAE group included only AA participants, whereas the GAE group included 12 AA children who were categorized by the DELV LVS section as GAE speakers and who did not live in communities with a predominance of AAs.

Parental education (PED) level. The PED level of at least one, and where possible both, parents was categorized by years of education, which ranged from 8 to 16 years. However, because the use of AAE is thought to vary inversely with PED level in the U.S. AA population (Washington & Craig, 1998), the lower PED levels were oversampled in order to have the most stringent test of the effect of dialect. Only 23% of the parents had schooling beyond a high school degree; only 2% had 4-year college degrees. In order to use the variable in the analysis of variance (ANOVA), years of school was categorized into three levels: 1 = *<12 years*, 2 = *high school degree*, and 3 = *more than high school*. Despite efforts to match PED levels of the GAE group to those of the AAE-speaking children, the average PED level of the GAE-first children tended to be slightly higher than that of the AAE speakers: 2.2 versus 1.9 for the TD subgroup and 2.0 versus 1.8 for the subgroup with LI, as indicated in Table 1. This difference was tested statistically before the principal analyses, as reported in Results under "control variables."

AQ21

Gender. Although the ratio of girls to boys for the entire sample was 49:51, the clinical status subgroups were not equally balanced by gender, consistent with demographic statistics on the relationship between gender and LI (Leonard, 1998). That is, there were more girls than boys in the TD groups and more boys than girls in the LI groups, with slightly more boys in the GAE group than in the AAE group. Therefore, the effect of gender was also tested before the

principal analyses. Table 1 shows gender distribution for the two dialect and two clinical status groups.

Region. To ensure the representativeness of the sample, participants came from all over the United States in proportions that approximated the regional distribution of the AA population in the latest census report (U.S. Bureau of the Census, 2000). However, regions thought to have higher overall dialect density (Charity, 2007) were oversampled, resulting in the following percentages: South (60%, instead of 51% as in the census), North Central (28% instead of 24%), Northeast (9% instead of 16%), and West (3% instead of 9%).

The distribution of AAE-speaking and GAE-speaking children in each age group who were classified by the criteria above is given in Table 1.

Materials

Screening materials. As described above in Participants, the children were selected for the study by reference to the certified SLPs' judgments of the children's dialect and clinical status, which were confirmed by the results of the **AQ22** DELV-ST ("Screener," Seymour et al., 2003). The Screener is a criterion-referenced test composed of two parts. One set of items, Part 1, consists of five phonological and 10 morphosyntax items that contrast sharply between AAE and GAE (Green, 2008). Part 1 provides criteria to identify two dialect categories of LVS for ages 4 to 12 years: *GAE* and *some or strong difference from GAE*. The second set of items, Part 2 of the DELV-ST, consists of seven morphosyntactic, four syntactic, and six nonword repetition items that are the same, or *noncontrastive*, across dialect groups. Part 2 provides criteria to identify four diagnostic risk categories: *very low*, *low*, *medium*, or *high* risk for disorder. Part 1 was used to confirm the SLPs' judgments of the children's LVS; Part 2 was used to confirm the SLPs' judgments of the children's clinical status.

Psychometric analyses of the DELV-ST (Seymour et al., 2003) revealed strong developmental growth on the diagnostic items between the ages of 4 and 9 years. Children at these ages who were a priori classified by experienced SLPs as language impaired made significantly more errors on its items than did children of the same ages who were identified by the SLPs as TD. However, there were no differences in performance on the diagnostic items between AAE- and GAE-speaking children at any of the ages.

Narrative materials. The narrative task was one of 14 subtests in the DLST. The sixth subtest, Short Narrative, consisted of two 6-picture sequences, one of which was subsequently incorporated into the DELV—Norm Referenced (DELV-NR, Seymour et al., 2005). Figure 1 shows an example of one of the picture sequences (the one that was not subsequently used in the DELV-NR).

As illustrated in Figure 1, the picture stories possessed three fundamental features that provided very strong

pragmatic motivation for the expression of the chosen indices of narrative cohesion and evaluation in order to elicit them without having to model them first, as one would do in a story retelling paradigm (P. de Villiers, 2004). First, there were two characters of the same gender interacting throughout the scenario who needed to be referred to contrastively as the focus of the events shifted from one to the other. Because the characters were of the same gender, simple pronouns would not suffice to contrast them. Second, there was an explicit temporal relationship between the events in the pictures, both within one picture and between pictures. Finally, the pictured scenario was directly based on a standard test of ToM and false belief reasoning in which a desired object is moved from one place to another without the major character observing the change of location (Perner et al., 1987). A "thought balloon" depicted the mental state of the primary character at a crucial point of the story, and the character's subsequent action was based on a false belief about an object in the story.

Procedure

Certified SLPs recruited children according to the dialect and clinical status criteria of the experimental design. More than 400 SLPs participated, most of whom were of EA background, reflecting the primary ethnic composition of the profession (American Speech-Language-Hearing Association [ASHA], 2005). The DSLT Project Coordinator at The Psychological Corporation, who was herself a certified SLP, designed the training materials for administering the narrative task and supervised the examiners.

Children were tested individually in a quiet room at their schools. They were administered all 14 subtests of the DSLT. Younger children were typically tested in two sessions. Then, for the purposes of this study, the items that became the DELV-ST were extracted from the database of responses for each child, as described in the following paragraphs.³

Narrative elicitation. All children took the subtests of the DSLT in the same order, with the Short Narrative subtest administered sixth. For this project, responses to the Short Narrative subtest were extracted from the DSLT database and were evaluated with the rubrics described in Coding and Scoring.

The picture sequences were each presented on a single page of the tented DSLT stimulus book facing the child and away from the examiner. All six pictures were on a single page so that there were no memory demands. The child was reminded that the examiner could not see the pictures and was told to look carefully at each picture to see what happened in the story. Then the child was asked to start at the

³We recognize that this procedure does not strictly follow the standardized procedure for administration of the DELV-ST, but the scores derived in this manner were used only for confirmation purposes. The primary judgments concerning dialect and clinical status were made in the settings of the children's schools and communities.

beginning and tell the whole story to the examiner. While the child was telling the story, the examiner listened for the two indices of narrative cohesion—reference contrasting and temporal expressions—and marked them down on a pre-printed record form as they occurred.

After the child told the story, the examiner pointed back to the panel in the sequence depicting the character with the thought balloon and asked the child to say again what was happening in that picture (Probe Question 1, ToM). Finally, the examiner pointed to the last panel of the story sequence, which depicted a standard test for children's ability to explain a person's behavior in accordance with a false belief (Bartsch & Wellman, 1989). In the picture, the major protagonist was looking for the desired object in the location in which he or she had first placed it, even though it had subsequently been moved by the other character to a different location. The examiner then asked the child why the protagonist was looking for the object in that location (Probe Question 2, False Belief). For example, for the cake story in Figure 1, the examiner said: "The little girl is looking for the cake in the cabinet. Why is she looking *there*?" (with the stress on the word *there*). Examiners wrote down the child's responses to the probe questions verbatim, usually a sentence or two. In addition, stories from 55 children were audio-recorded so they could be analyzed for reliability.

Coding and Scoring


"Online" coding of the cohesive elements (with subsequent scoring). Evaluation of the targeted cohesive language elements in the children's stories was streamlined so that examiners could listen for them while the child was speaking and code them, without the need for transcription or a second listening. The preprinted record form listed a number of alternatives, with examples of the most frequent responses in prepiloting, and the examiner noted the presence or absence of the relevant forms on the record form but did no actual scoring. Assigning points was done by the authors from the examiners' record form (or in the case of reliability analyses, from their own record forms made while listening to the audio-recorded session). Coding and scoring examples are provided in Appendix B.

Reference contrasting. Examiners were instructed to mark off on the record form when the child explicitly contrasted the two characters so that the listener could tell them apart. Subsequently, each story was scored with 1 point if *both* characters were specifically identified. No points were awarded if only one character was identified, as in the story examples A and B in Appendix B. To get this point, the child could use one or more linguistic devices such as proper names, adjectives, prepositional phrases, or relative clauses: for example, *the big girl and the little girl*, or *the boy with the train and the other guy*. Because characters for each story were the same gender, *she* and *the girl* were not sufficient to distinguish them. Counting scores for both stories, each

child received a 0, 1, or 2 for the reference contrasting subscore.

Temporal expressions. Temporal expressions were coded online according to their syntactic and developmental sophistication (Berman, 1988; Berman & Slobin, 1994). Examiners were to note the use of conjunctive sequencers to link events, such as *then*, *and then*, or *next*, and the production of any adverbial clauses of time, such as "*when the girl came back*, she looked in the cabinet" (with a tensed verb in the *when*-clause). Stories with no time references received 0 points; with sequencers only, they received 1 point; and with a full adverbial clause, they received 2 points. For each narrative, the children were given a score equivalent to the highest level of temporal expression they used, so total temporal expression subscores for the two narratives varied from 0 to 4 for each child. Counting measures from the two stories, the maximum cohesion score, with reference contrasting and temporal expressions combined, was 6.

Probe-question coding. For this study, examiners neither coded nor scored the responses to the probe questions. They merely recorded the responses by hand on the record form. The authors then coded and scored the mental state descriptions and false belief explanations from the examiners' written responses. The probe-question coding was based on Wellman and Bartsch's (1994) finding that children understood and talked about people's desires and intentions before they were able to understand people's cognitive states. So for the responses to Probe Question 1, children received 0 points if they just described an action (e.g., *The girl is coming in the room*), 1 point if they specified the character's desire or intention (e.g., *The boy is coming to get his train*), and 2 points if they referred explicitly to the character's cognitive state (e.g., *The girl is thinking about the cake or he's wondering about the train*).

Similarly, for their response to the second probe question, the explanation of why the character was looking in the wrong place for the object (in the last panel of the story), children received 1 point for stating the character's motivation for looking (e.g., *So she could eat her cake or she wants her cake*), but 2 points for giving a ToM explanation. Full-credit ToM responses included direct references to the character's ignorance or false belief (e.g., *She thinks it's there* or *He doesn't know it was moved*), as well as responses without an explicit mental verb as long as the explanation was consistent with understanding the false belief, for example, *He put it there* or *He didn't see it moved* (Bartsch & Wellman, 1989). The total score for Probe Question 1 varied from 0 to 4 for responses in the two story sequences, and a separate score of 0 to 4 for responses to Probe Question 2. Measures for the two stories were summed to give an evaluation score (ToM and false belief understanding, maximum of 8 points). The cohesion and evaluation subscores ther made up the composite index of narrative cohesion and evaluation, with a maximum of 14 points.

Reliability and Validity

Reliability and validity of the Short Narrative subtest scoring was performed for the 4-, 5-, and 6-year-olds as part of the DSLT field testing; stories from children ages 7 to 9 years were checked for reliability and validity in connection with the dissertation research of the first author. Two stories each from 64 children were audio-recorded—55 participants in the DSLT field test, and 9 participants in the dissertation research—yielding 128 recordings. After the inclusion criteria for this study were applied, approximately half of the 4- to 6-year-olds and none of the older children in the reliability study were included in the current analyses, which is discussed further under Limitations of the Study. The 128 recorded stories were transcribed for reliability and validity analyses at the following five levels: (1) reliability of the transcriptions, (2) accuracy of the examiners' online notations of reference contrasting and temporal expressions, (3) reliability of the examiners' hand recording of responses to the probe questions, (4) a validity check of the presence/absence scoring of cohesion indices relative to the actual content of the stories, and (5) validity of the narrative indices relative to holistic judgments of story quality.

Transcription reliability. The 128 narratives were each independently transcribed by two research assistants in communication disorders who were familiar with AAE speech. Discrepancies between the transcriptions of the 4- to 6-year-old children's narratives were resolved by consensus with a third transcriber, the first author (an AAE speaker, then a doctoral student in communication disorders). By the end, there was 100% agreement on the transcriptions. For the 7- to 9-year-olds, the current authors retranscribed the portions of the narratives that were relevant to the scoring. There was only one disagreement in the transcriptions, that is, there was agreement on 71 of 72 data points (8 passages each for 9 stories), or 98.6% agreement.

Online coding reliability. Reference contrasting and temporal marking in the transcriptions were scored by two of the present authors, with 95% agreement for the younger children and >98% agreement for the older children. Comparing the authors' coding from the transcribed narratives with the codes assigned by the clinicians and researchers during the child's telling of the stories, author and clinician agreement was 95% for reference contrasting and 90% for temporal expressions. One coding error that accounted for most of the differences in reference contrasting coding occurred when a clinician awarded credit for a specific reference, but it was only to *one* character, for example, *the little girl* for the little girl, but only *she* for the older person. Typical errors found in temporal expression coding were cases where adverbial clauses that should have received two points were marked as one-point sequencers, when, for example, a clause with a full verb, *after he finished lunch* (2 points), was not distinguished from the phrase without a tensed verb, *after lunch* (1 point).

Probe-question coding reliability. Because coding of mental state and false belief descriptions was done from the SLPs' written responses and not "online," there was no SLP coding to confirm. Reliability procedures for these items checked the current authors' interrater consistency in assigning credit. Agreement between the third author's coding and the codes assigned by the first two authors was 98% for the younger children. For the older children, agreement between the first and third authors' independent coding was 100%.

Also, the examiners' hand-recording of the responses to the probe questions was checked against the reliability transcriptions. Only 10 discrepancies for the two questions per story (i.e., of 256 instances, or 3.9% disagreement) were noted in the wording on the response form compared to the transcriptions done from the tapes. These were for the most part abridgements after the child had said enough to be able to code the response as *action*, *intention*, or *ToM*. In three instances, the examiner regularized the child's words, such as *he wants his train*, where the transcribers heard "*he want his train*." For the most part, though, dialectal forms and even grammatical errors were faithfully rendered, and no departures from the transcript changed the number of points awarded.

Validity of coding schema. A check was also made to see if credit was given under our coding system of *presence* versus *absence* for reference contrasting devices or temporal expressions that had been used inappropriately by the child. For example, did the child say *the little boy moved the train after the big boy ate lunch*, using *after* instead of *when* or *while*. Only two such instances were noted, both involving misuses of *after* or *before*. This is not to say that all of the reference contrasting, for example, was adequately executed throughout the child's story. For most of the younger children's stories, it was not. But, the scores credited the child's recognition that some effort had to be made to designate each character clearly for the listener, not necessarily the ability to carry it out throughout the story.

Overall coding validity. Finally, to confirm the validity of our decision to award credit based on only a judgment of presence or absence of a feature, we looked among the transcripts for stories that appeared holistically to have been misjudged. There were no high-scoring stories that had many inappropriate usages, nor were there low-scoring stories that had high expressiveness (Ukrainetz et al., 2005) or exceptional organizational structure (Berman & Slobin, 1994; Pearson, 2002) that one might want to reward. Three transcriptions of typical stories and probe-question responses, along with the coding decisions and scores assigned, are provided in Appendix B, one each from a 4-, 6- and 8-year-old.

Statistical Analyses

The cohesion scores (for reference contrasting and temporal expressions), the evaluation score (for mental states and false beliefs), and the composite index were the dependent variables in various ANOVAs. Control variables—gender,

AQ23

region, and PED level—were explored first. Then the independent variables of interest—age, clinical status, and dialect—were analyzed.

RESULTS

To answer our research questions, the descriptive results provide information about the presence and the level of complexity of the two selected indices of cohesion and evaluation in the stories of children who differed in age, clinical status, and dialect. The ANOVAs tell whether any differences in the scores on the indices observed were significant between children of different ages (for Research Question 1), different clinical status (for Research Question 2), or different dialect groups (for Research Question 3). Indeed, for the features tested here, age and clinical status were significant effects, but dialect was not. Testing each of the narrative elements in separate subscales and also together in the composite score showed which measures had more or less diagnostic value and at which ages.

Descriptive Results

Extrapolating from the 128 stories, two each from 64 children, that were audio-recorded and transcribed for reliability purposes, we can characterize the short narratives as follows: The stories from the 4- to 6-year-olds ranged from 35 to 135 words, with an average of 75 words per story and an average of 20 words in the answers to the two probe questions. The shortest story contained three sentences, the longest story contained 15 sentences, and the average mean length of utterance in words (MLU_w) was 8.1. Average type-token ratio was 0.52. Descriptive statistics for the stories from the 7- to 9-year-olds were similar to those for the younger children. Older children's stories ranged from 48 to 163 words, with an average of 77 words per story and an average of 25 words in the answers to the two probe questions. The shortest story contained five sentences, the longest story contained 15 sentences, and the average MLU_w was 9.5. Average type-token ratio was 0.50. As illustrated in Appendix B, the highest scoring story was not necessarily the longest. Story C was a full-credit story and had 82 words, which was very close to the average. Mean values by subgroup for the dependent variables are reported individually in the following paragraphs.

Control variables. Before turning to the principal analyses, we tested the difference in PED level between the dialect groups noted in the Method section. A univariate ANOVA showed that the difference was significant, $F(1, 527) = 15.8$, $p < .0005$, $\eta^2 = .03$, but it was a small effect, so no special accommodation was made for the difference. We also compared the narrative indices of children whose transcripts were included in the reliability study to the indices of children

whose transcripts were not included. No significant difference was found: $F(1, 527) = 2.48$, $p = .116$, $\eta^2 = .005$.

To explore the potential effects of gender, region, and PED level on the narrative indices, we conducted a multivariate ANOVA (MANOVA) with cohesion score, evaluation score, and composite score as dependent variables and gender, region, and PED levels as independent variables. The MANOVA revealed one significant interaction, between PED level and region for the composite index, $F(6, 505) = 2.8$, $p = .01$, $\eta^2 = .03$, but no main effect for either factor for that score. There was a main effect of region for the cohesion subscore, $F(3, 505) = 6.2$, $p < .0005$, $\eta^2 = .04$, but no other significant effects for region or interactions involving region or PED. Estimates of the mean scores by region and PED level showed that scores for the West were higher in general than for the other regions, but the pattern was not consistent across the PED levels, as indicated by the significant interaction. Because there were very few children involved, and the effect size was small, no further accommodation was made for region.

Gender was a main effect for both the total score and the evaluation subscore, $F(1, 505) = 13.6$, $p < .0005$ and $\eta^2 = .03$, and $F(1, 505) = 14.4$, $p < .0005$, $\eta^2 = .03$, respectively. The lower means for boys, $M = 6.8$, relative to the girls' mean of 8.6 in the total score and 3.7 and 4.9 in the evaluation subscore were consistent with the well-documented preponderance of boys with a diagnosis of LI (Leonard, 1988). Therefore, the MANOVA was repeated with clinical status as a factor. With the addition of this factor, the gender differences were no longer statistically significant, gender for composite score, $F(1, 525) = 3.1$, $p = .08$, $\eta^2 = .006$, and for the evaluation subscore, $F(1, 525) = 3.28$, $p = .07$, $\eta^2 = .006$. Instead, clinical status was significant for both variables, and there were no interactions: clinical status, $F(1, 525) = 82.2$, $p = .0005$, $\eta^2 = .14$, and $F(1, 525) = 54.6$, $p < .0005$, $\eta^2 = .09$, for the total score and ToM subscore, respectively. Because clinical status was the variable of interest, gender was not investigated further.

Development, Discrimination, and Dialect Effects for Narrative Features

Turning to the variables of interest, we conducted a $6 \times 2 \times 2$ MANOVA for the cohesion and evaluation subscores, with age (4 to 9 years),⁴ clinical status (LI vs. TD), FN4

⁴One must use caution when interpreting the results relative to the 7-year-olds. As shown in Table 1, there were very few participants at age 7, and there was an imbalance between the LI and TD groups for the GAE 7-year-olds. The pattern of means for that age, therefore, was often inconsistent with the general trends observed for the other ages. The following analyses were conducted a second time, collapsing age into three groups (4–5, 6–7, and 8–9). Effect sizes were stronger, but there were no differences in the pattern of the effects. Because some distinctions in the post hoc analyses were lost with that grouping, the decision was made to report the figures for all six ages.

TBL2
TBL3
TBL4

and dialect (AAE vs GAE) as independent variables. Because the composite score was not independent of its subscores, the composite index was tested in a separate $6 \times 2 \times 2$ ANOVA. Tables 2 to 4 show the mean and standard error for the participants by age, clinical status, and dialect for the two subindices and the composite narrative index.

Cohesion index: Reference contrasting and temporal expressions. As shown in Table 2, the mean cohesion scores showed substantial differences for the TD and LI groups from 4 to 6 years of age, but for children 7 years and older, the TD and LI children's scores were closer together. This pattern is reflected in the significant interaction between age and clinical status for that dependent variable: $F(5, 505) = 4.03, p = .001, \eta^2 = .04$. Age and clinical status were also significant main effects (Clinical status, $F(1, 505) = 31.3, p < .0005, \eta^2 = .06$; age, $F(5, 505) = 21.81, p < .0005, \eta^2 = .18$). By contrast, dialect was not a significant effect, $F(1, 505) < 1, n.s., \eta^2 < .001$. Post hoc Scheffe comparisons showed that age 4 was different from all other groups: p level was 0.01 for the difference between age 4 and age 5, and $p < .0005$ for all other ages. The 5-year-olds were significantly different from all other ages except age 6, but age 6 was different only from age 4. Ages 6 to 9 did not differ significantly on this score: The developmental trend was greater at the younger ages.

AQ25 Evaluation subscores. The maximum score for the descriptions of the mental states in Probe Question 1 and the explanation of action based on a false belief in Probe Question 2 was 8 points, counting both stories. Within our protocol, all children commented in some way on the unseen displacement of an object in the story: 30% referred just to the action depicted, and 70% explained the character's action with reference to a mental state. Ninety percent of the 9-year-olds articulated the desire, intention, or thought of the characters, and even among the 4-year-olds, 45% of the

Table 2. Mean cohesion score (reference contrasting and temporal expressions) by age and dialect and age and clinical status (standard error in parentheses), maximum score = 6.

AQ26

	Age					
	4	5	6	7	8	9
Dialect						
AAE	2.2 (.20)	3.2 (.17)	3.5 (.15)	3.9 (.35)	4.5 (.29)	4.4 (.34)
GAE	2.3 (.21)	3.1 (.18)	3.5 (.18)	5.2 (.45)	4.1 (.29)	3.7 (.32)
Clinical status						
TD	3.3 (.16)	3.8 (.15)	4.1 (.14)	4.2 (.32)	4.7 (.30)	4.5 (.29)
LI	1.3 (.24)	2.5 (.19)	2.9 (.19)	4.9 (.47)	4.0 (.28)	3.7 (.36)

Table 3. Mean evaluation score (mental states and false belief understanding) by age and dialect and age and clinical status (standard error in parentheses), maximum score = 8.

AQ27

	Age					
	4	5	6	7	8	9
Dialect						
AAE	2.3 (.28)	3.5 (.23)	3.9 (.20)	4.5 (.48)	5.6 (.40)	6.3 (.46)
GAE	2.7 (.29)	3.7 (.25)	4.9 (.25)	5.3 (.63)	5.4 (.40)	5.9 (.44)
Clinical status						
TD	3.3 (.22)	4.6 (.21)	5.5 (.19)	5.8 (.45)	6.1 (.41)	6.6 (.40)
LI	1.7 (.33)	2.5 (.27)	3.4 (.26)	4.0 (.65)	4.9 (.39)	5.6 (.50)

participants did so. Means for these variables are found in Table 3.

The MANOVA with dialect, clinical status, and age as independent factors indicated that for this variable, there was no interaction of age and clinical status, $F(5, 505) < 1, n.s., \eta^2 < .01$. Clinical status and age were significant main effects, but dialect was not: clinical status, $F(1, 505) = 55.47, p < .0005, \eta^2 = 1.0$; age, $F(5, 505) = 28.95, p < .0005, \eta^2 = .22$; dialect, $F(1, 505) = 1.92, p = .17, \eta^2 = .004$. In post hoc Scheffe tests, ages 4, 5, and 6 were all significantly different from each other at the .006 level or lower. The developmental trend was less evident from ages 6 to 8, where differences between children in the different age groups were not statistically significant. However, age 9 was significantly different from age 6 ($p = 0.002$), so unlike the cohesion subscore, there was continued growth in this subscore after age 6. There was no interaction of dialect and clinical

AQ28

Table 4. Mean composite narrative index by age and dialect and age and clinical status (standard error in parentheses), maximum score = 14.

	Age					
	4	5	6	7	8	9
Dialect						
AAE	4.6 (.37)	6.7 (.31)	7.5 (.27)	8.5 (.63)	10.2 (.53)	10.7 (.61)
GAE	5.0 (.38)	6.7 (.32)	8.5 (.33)	10.5 (.82)	9.4 (.53)	9.6 (.58)
Clinical status						
TD	6.6 (.29)	8.4 (.28)	9.6 (.25)	10.0 (.59)	10.8 (.54)	11.0 (.53)
LI	3.0 (.44)	5.1 (.35)	6.4 (.34)	9.1 (.86)	8.8 (.52)	9.2 (.66)

status, indicating that the discrimination between TD and LI was similar in both dialect groups.

Composite index of narrative cohesion and evaluation.

The means and standard errors for the clinical status and dialect groups for the 14-point index are shown in Table 4. Clinical status and age were significant main effects with medium and large effect sizes, respectively: clinical status, $F(1, 505) = 75.8, p < .0005, \eta^2 = .13$; age, $F(5, 505) = 42.7, p < .0005, \eta^2 = .30$. However, there was no significant effect for dialect, $F(1, 505) = 1.32, p = .25, \eta^2 = .003$. There were no significant interactions, including no interaction of clinical status with dialect, indicating that the measure distinguished disorder equivalently in both dialect groups. The developmental pattern seen in the Scheffe post hoc tests was similar to the one observed for the evaluation subscore, except that p values were even lower.

Relationship Between Subscores

Correlations between the composite index and the cohesion and evaluation subscores were $r(529) = 0.76$ and 0.89 , respectively, which were both significant at the $p = .001$ level, indicating a high degree of association between them. Thus, although the evaluation score was more highly correlated, both scores appear to have contributed significantly to the composite score. The correlation between the subscores themselves was $r(529) = .38$, which is also a significant coefficient, here at the .01 level. The lower correlation coefficient indicates that the two subscores were contributing somewhat different information to the composite.

Thus, results showed that these measures, especially the composite index, captured significant differences in the scores of stories throughout this age range. There was strong developmental growth from ages 4 to 9 years and a significant impairment of knowledge of these four elements in the stories of LI children of both dialect groups. However, the AAE- and GAE-speaking children showed no reliable difference in performance on these measures.

contrasting expressions) and conjunctions (i.e., temporal expressions) have been the most reliable in discriminating between TD and LI in other studies such as Greenhalgh and Strong (2001). In the present study, specific aspects of both of these measures also differentiated between TD children and children receiving intervention for LI, although the clinical status effect for the cohesion measure was stronger for the younger children than for children >6 years. In the current study, we also sought to examine the understanding of behavior based on false belief, or ToM.

Unlike in previous studies of literate language, such as Greenhalgh and Strong (2001) and Curenton and Justice (2004), a difference in ToM by clinical status contributed strongly to the total index and was a robust finding in this study. In the previous studies (Curenton & Justice, 2004; Greenhalgh & Strong, 2001), there was less than one mental state reference per ~10 clauses. By contrast, this protocol foregrounded the characters' state of knowledge, and with targeted probe questions about ToM elements that were explicitly portrayed, it elicited many more specific references to that knowledge. Therefore, the effectiveness of the elicitation for the ToM aspect of the coding was greatly enhanced, and so children were more likely to demonstrate their understanding of cognitions than in other studies. With higher levels of occurrence, ToM indices revealed a strong developmental trajectory and the ability to discriminate between children by clinical status.

Indeed, in order to be effective discriminators throughout the age range, this study found both subscores, with all four components—reference contrasting, temporal expressions, ToM, and false belief—to be necessary. In recent work developing the NAP, Justice et al. (2010) eliminated the elements of evaluation from the measures that Curenton and Justice had used in their 2004 study and worked more specifically with just the linguistic aspects of cohesion. They found that the new measure of cohesion—without evaluation—showed promise for discriminating LI in their preschoolers. However, as we observed in the clinical status and age interactions for the indices of cohesion in the current study, cohesive elements alone did not discriminate impairment for the older groups. These elements appear to be easier, and by age 7, even the children with LI were near ceiling. On the other hand, the more difficult evaluation measures—ToM and false belief understanding—appeared to be essential for the older ages, where the absence of an interaction indicated that they continued to discriminate between TD and LI.

The current study emphasizes that, in the context of this narrative elicitation, all four of the narrative indices proved to be dialect neutral across the age range. AAE- and GAE-speaking children produced closely comparable performances on these indices, including evaluation—ToM and false belief understanding. This last finding contradicts previous work by Holmes, Black, and Miller (1996), who suggested that mastery of ToM reasoning in AA children was delayed relative to EA peers as a result of differential

DISCUSSION

The current findings contribute to the study of narrative generally, but they are especially useful in advancing our ability to assess stories from AAE speakers in a dialect-neutral way. As in previous studies (Greenhalgh & Strong, 2001; Justice et al., 2010), the indices of linguistic cohesion examined here were significant discriminators, but principally at the younger ages. The features examined had been shown by prior research to be central components of children's narrative development across many languages between 4 and 9 years of age (Berman & Slobin, 1994). Furthermore, out of the set of linguistic features that authors have proposed as aspects of literate language (Pelligrini, 1984; Westby, 1991), elaborated noun phrases (i.e., reference

AQ30

exposure to mental state language in their homes. Counter-evidence to that claim has previously been provided by Curenton and Justice (2004), who also found that a composite measure that included mental and linguistic verbs was dialect neutral between AA and EA samples. Nonetheless, their sample included preschoolers only, and as mentioned above, the mental and linguistic verbs occurred very rarely and thus their study was not a strong test of narrative evaluation in Labov's (1972) sense of the word.

AQ31
 An important feature of the present study, like that of Curenton and Justice (2004), was that the AAE and GAE children were approximately matched for PED level. Possible effects of PED differences were neutralized by using a very homogenous sample but were nonetheless investigated by entering PED level in the control variable analyses. The effect of PED level on these measures was relatively weak, with no significant effect on the subindices of cohesion and evaluation or on the composite index. For the ToM measure, even this truncated range of PED level was a significant factor but accounted for very little of the variance.

Finally, the stories elicited by the protocol used in this study fulfill the goal set out by Justice et al. (2010) to be "scalable" for use with large numbers of children. The single-page picture sequences in this study were easy to administer and code. The short stories they elicited provided useful information for SLPs about the children's language and narrative skills yet required only a short time and little background knowledge of narrative development from the clinician. Clinicians were able to reliably code these features and obtain important information about children's narrative abilities online as the children told the stories or from brief responses to probe questions immediately following the child's story.

Thus, these indices make an important contribution to the field of assessment of language and potentially to the field of assessment of school readiness by demonstrating an efficient, dialect-neutral method for investigating and evaluating children's narrative abilities, including ToM understanding. Although currently there are several measures of narrative development available that provide valuable information to SLPs (Gillam & Pearson, 2004; Justice et al., 2006; Schneider et al., 2005, among others), these measures have not established strong dialect neutrality, and so one cannot have full confidence in their appropriateness for CLD children. Furthermore, these measures require the ability and time to record and later listen to or transcribe the children's productions. By making several modifications to the protocols of prior studies to increase the narrator's need for reference contrasting and complex temporal expressions, and to highlight a classic false belief scenario, the elicitation of these features was more effective. When this protocol was used, children of all ages produced the relevant constructions for each measure, even those relating to ToM. This brief protocol proved to be sensitive to age differences, effective at finding narrative language difficulty, insensitive to dialect differences, and reasonably valid and reliable.

Study Limitations

Although the current study was extensive and innovative, it had several limitations. First, despite the large number of participants overall, there were some cells in the factorial analysis that were very small (e.g., 7-year-old TD GAE speakers). The smaller number of participants at the later ages meant that age differences, especially for the 7- to 9-year-olds, could not be as precisely delineated as those at the younger ages. Thus, there may be differences between 7-, 8-, or 9-year-olds that were not detected in this sample.

Second, the AA children's stories may have been somewhat constrained by having mostly EA examiners. There is some evidence from the DSLT field testing that AA children interacting with AA examiners used more of their home language features as compared to GAE children interacting with EA examiners (Pearson, Velleman, Bryant, & Charko, 2009). However, because the present study involved only *noncontrastive* elements, and the level of use of contrastive features is irrelevant to the scoring, the effect of this variable on outcomes should be relatively small. Furthermore, the predominantly EA demographic of our examiners represents the predominant ethnicity of the field (ASHA, 2005) and so is more representative of the actual assessment situations that AA children are likely to encounter.

A third important limitation to the evidence presented here was the fact that reliability procedures for the field testing were performed on a group of stories that only partially overlapped with the stories in the current study. All of the children in the reliability analysis for the field test took the tests under the same conditions and were scored according to the same guidelines by the same research group. Nevertheless, even though the procedure and scoring were confirmed in the process, strictly speaking, reliability was done on only 7% of the current participants, not at least 10% as had been planned and as would be more customary. Another potential problem with the reliability procedures was that there were no audiotapes of the older children in the current study. However, with a comparable sample recruited and tested under the same conditions as the children in this study, we were able to establish that the longer, more complex narratives from older children were not harder to score with the simplified coding presented here. That is, whether the story was three or 30 sentences long, the examiner listened online for the same three short pieces of information: (a) whether both characters were identified or not; (b) whether time relations were expressed or not; and, if so, (c) whether they were expressed with only simple adverbs (like *then*, *next*, or *later*) or with full clauses, such as *when the boy came back into the room*. Similarly, whether the story took 30 s or 5 min to relate, the remaining points were evaluated promptly from short written answers to two specific questions.

Although the stories of the older children were more competent overall, they were not necessarily longer. In fact, judging from the responses to the probe questions recorded

verbatim by the examiners, the older children's answers were more succinct and to the point than those of the younger children. Whereas several younger children were observed to recap the whole story or to launch into an imagined episode that was irrelevant from the point of view of the scoring system, the 7- to 9-year-olds answered just the question that was asked, similar to the older children in the Bamberg and Damrad-Frye (1991) study. (See Pearson, 2002, for a similar length reversal with frog stories from older vs. younger children.)

The brevity of the protocol and the simplicity of the scoring work in favor of the reliability of our measures. A similar observation was made by Ukrainetz et al. (2005). In their study, as in this one, stories were told from relatively constrained five- or six-picture sequences, as opposed to the 25 pages in the typical wordless storybook (such as by Mayer, 1967, 1969) or the more open-ended single-picture fantasy narrative of the Test of Narrative Language (Gillam & Pearson, 2004). Ukrainetz et al. noted that the short sequences with relatively uniform referential content provided a good platform for children to generate the targeted behaviors on their own, not following a model, as in story retelling paradigms.

The pointed probe questions in this protocol performed an important function. We noted in our transcripts that not all of the children articulated the details about mental states in their spontaneous stories, even those who later demonstrated advanced understanding of ToM. By asking the probe questions after the child's own story, we maximized the probability of eliciting some reference to the characters' mental states, still in the child's own words. Writing down the short responses as they were spoken was relatively easy for the examiners, and according to our transcriptions, very accurate. From a methodological point of view, then, the probe questions represented a middle ground between spontaneous speech and story retellings.

Clinical Implications

This study furthers our knowledge of the development of key features of narrative language in the spoken narratives of both AAE- and GAE-speaking children. Using the elicitation techniques provided by the narrative protocol (P. de Villiers, 2004), SLPs performed an informative assessment of young children's narrative development that was quickly and reliably carried out.

These indices are far from a full analysis of children's spoken narratives, like those carried out on transcribed narratives elicited by a wider range of different materials and communicative contexts (see Gillam & Pearson, 2004; Miller, Gillam, & Peña, 2003; Schneider et al., 2005; Strong, 1998). However, the measures of reference contrasting, temporal expression, and ToM reasoning are arguably among the most fundamental features of narratives (Berman & Slobin, 1994; Bruner, 1986). The diagnostic power of each

subscore by itself was relatively small, as shown by the small effect sizes for the clinical status variable ($\eta^2 = 0.04$), but when both measures were combined in the composite index, they explained a significant portion of the variance in children's scores ($\eta^2 = 0.13$, for clinical status).

The elements in the brief protocol presented here are central features of the decontextualized literate language that has been demonstrated to be prerequisite for fluent reading and writing skills in the early grades (Pelligrini, 1984; Snow et al., 1995; Westby, 1991, 1999), so they may be helpful in a classroom or other educational setting in identifying the presence or absence of these skills. Clinicians, however, will likely want a deeper assessment that will provide more information on children's specific strengths and weaknesses. In that case, this protocol could be a useful narrative screener that could be administered first to locate children for whom a fuller assessment with another instrument might be useful. The tasks can also be presented in a battery with other noncontrastive assessments to achieve a fuller picture of a child's level of language functioning. Indeed, one of the stories from this study was incorporated into the DELV-NR (Seymour et al., 2005), and informal age-graded benchmarks are available for these features (P. de Villiers, 2004; Pearson & Ciolli, 2004).

This study further demonstrated that literacy features in spoken narratives are reliable discriminators of LI. This study adds convincing evidence of their dialect neutrality as well, making them applicable to a larger, more diverse population of children with LI, including those who speak AAE.

ACKNOWLEDGMENTS

The research in this paper was funded in part by NIH Contract #N01-DC8-2104 (Harry Seymour, PI) to the University of Massachusetts in Amherst, Smith College, and The Psychological Corporation.

REFERENCES

- Allen, B. A., de Villiers, J. G., & François, S. (2001). Deficits versus differences: African American children's linguistic paths towards a theory of mind. In M. Almgren, A. Barrena, M.-J. Ezeizabarrena, I. Idiazabal, & B. MacWhinney (Eds.), *Research on child language acquisition: Proceedings of the 8th Conference of the International Association for the Study of Child Language* (pp. 1006–1014). Somerville, MA: Cascadilla Press.
- American Speech-Language-Hearing Association. (2005). *Annual counts of the ASHA membership and affiliation, year-end 2005*. Retrieved from <http://www.asha.org/uploadedfiles/research/05wwAethnicitytrace>.
- Astington, J. (1993). *The child's discovery of the mind*. Cambridge, MA: Harvard University Press.
- Bamberg, M. (1986). A functional approach to the acquisition of anaphoric relationships. *Linguistics*, 24, 227–284.

- Bamberg, M., & Damrad-Frye, R.** (1991). On the ability to provide evaluative comments: Further explorations of children's narrative competencies. *Journal of Child Language*, *18*, 689–710.
- Baron-Cohen, S., Leslie, A. M., & Frith, U.** (1985). Does the autistic child have a “theory of mind”? *Cognition*, *21*, 37–46.
- Bartsch, K., & Wellman, H. M.** (1989). Young children's attribution of action to beliefs and desires. *Child Development*, *60*, 946–964.
- Bartsch, K., & Wellman, H. M.** (1995). *Children talk about the mind*. New York, NY: Oxford University Press.
- Berman, R. A.** (1988). On the ability to relate events in narrative. *Discourse Processes*, *11*, 469–497.
- Berman, R. A., & Slobin, D. I.** (1994). *Relating events in narrative: A cross-linguistic developmental study*. Hillsdale, NJ: Erlbaum.
- Bliss, L. S., Covington, Z., & McCabe, A.** (1999). Assessing the narratives of African American children. *Contemporary Issues in Communication Science and Disorders*, *26*, 160–167.
- Bruner, J.** (1986). *Actual minds, possible worlds*. Cambridge, MA: Harvard University Press.
- Bruner, J., & Feldman, C.** (1993). Theories of mind and the problem of autism. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding other minds: Perspectives from autism* (pp. 267–291). Oxford, England: Oxford University Press.
- Burns, F. A.** (2004). *Elicited versus open-ended narratives in African American children*. Unpublished doctoral dissertation, University of Massachusetts, Amherst.
- Campbell, J.** (1949). *The hero with a thousand faces*. New York, NY: Pantheon Books.
- Champion, T.** (1997). “Tell me somethin’ good”: A description of narrative structures among African American children. *Linguistics and Education*, *9*, 251–286.
- Champion, T.** (2003). *Understanding storytelling among African American children: A journey from Africa to America*. Mahwah, NJ: Erlbaum.
- Champion, T., Seymour, H., & Camarata, S.** (1995). Narrative discourse of African American children. *Journal of Narrative and Life History*, *5*, 333–352.
- Charity, A.** (2007). Regional differences in low-SES African American children's speech in the school setting. *Language Variation and Change*, *19*, 281–293.
- Crago, M. B., Eriks-Brophy, A., Pesco, D., & McAlpine, L.** (1997). Culturally based miscommunication in classroom interaction. *Language, Speech, and Hearing Services in Schools*, *28*, 245–254.
- Curenton, S. M., & Justice, L. M.** (2004). African American and Caucasian preschoolers' use of decontextualized language: Literate language features in oral narratives. *Language, Speech, and Hearing Services in Schools*, *35*, 240–253.
- Dasinger, L., & Toupin, C.** (1994). The development of relative clause functions in narrative. In R. A. Berman & D. I. Slobin (Eds.), *Relating events in narrative: A cross-linguistic developmental study* (pp. 457–514). Hillsdale, NJ: Erlbaum.
- de Villiers, J., & Pyers, J.** (1997). Complementing cognition: The relationship between language and theory of mind. In E. Hughes, M. Hughes, & A. Greenhill (Eds.), *Proceedings of the 21st Annual Boston University Conference on Language Development* (pp. 136–147). Somerville, MA: Cascadilla Press.
- de Villiers, J. G., & Pyers, J. E.** (2002). Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief-understanding. *Cognitive Development*, *17*, 1037–1060.
- de Villiers, J. G., Roeper, T., Bland-Stewart, L., & Pearson, B. Z.** (2008). Answering hard questions: Wh- movement across dialects and disorder. *Applied Psycholinguistics*, *29*, 67–103.
- de Villiers, P. A.** (2004). Assessing pragmatic skills in elicited production. *Seminars in Speech and Language*, *25*(1), 57–72.
- de Villiers, P., de Villiers, J., Picone, C., Wilkins, A., Dinkins, E., & Burns, F.** (2010). Dialect and narrative skills in African American preschoolers. In K. Franich, K. M. Iserman, & L. L. Keil (Eds.), *Proceedings of the 34th Annual Boston University Conference on Language Development* (pp. 114–125). Somerville, MA: Cascadilla Press.
- Fiestas, C. E., & Peña, E. D.** (2004). Narrative discourse in bilingual children: Language and task effects. *Language, Speech, and Hearing Services in Schools*, *35*, 155–168.
- Gee, J. P.** (1985). The narrativization of experience in the oral style. *Journal of Education*, *167*, 9–35.
- Gee, J. P.** (1986). Units in the production of narrative discourse. *Discourse Processes*, *9*, 391–422.
- Gillam, R., & Pearson, N.** (2004). *Test of Narrative Language*. Austin, TX: Pro-Ed.
- Gorman, B. K., Fiestas, C. E., Pena, E. D., & Clark, M. R.** (2011). Creative and stylistic devices employed by children during a storybook narrative task: A cross-cultural study. *Language, Speech, and Hearing Services in Schools*, *42*, 167–181.
- Green, L. J.** (2008). *Literacy and developing African American English-speaking children*. Invited symposium at the annual meeting of the American Speech-Language-Hearing Association, Chicago, IL. **AQ32**
- Greenhalgh, K. S., & Strong, C. J.** (2001). Literate language features in spoken narratives of children with typical language and children with language impairments. *Language, Speech, and Hearing Services in Schools*, *32*, 114–125.
- Gutierrez-Clellen, V. F., & Quinn, R.** (1993). Assessing narratives of children from diverse cultural/linguistic groups. *Language, Speech, and Hearing Services in Schools*, *24*, 2–9.
- Hall, W. S.** (1984). African American English corpus. In B. MacWhinney (2000) *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Erlbaum. **AQ33**
- Halliday, M. A. K., & Hasan, R.** (1976). *Cohesion in English*. London, UK: Longman.
- Harris, P.** (1996). Desires, beliefs, and language. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind* (pp. 200–222). Cambridge: Cambridge University Press. **AQ34**
- Heath, S. B.** (1983). *Ways with words: Language, life, and work in communities and classrooms*. Cambridge, MA: Cambridge University Press.

- Hester, E. J.** (1996). Narratives of young African American children. In A. Kamhi, K. Pollock, & J. Harris (Eds.), *Communication development and disorders in African American children* (pp. 227–245). Baltimore, MD: Brooks.
- AQ35 Hickmann, M.** (2003). *Children's discourse: Person, space, and time across languages*. Cambridge, UK: Cambridge University Press.
- AQ36 Holmes, H. A., Black, C., & Miller, S. A.** (1996). A cross-task comparison of false belief understanding in a Head Start population. *Journal of Experimental Child Psychology*, 63, 263–285.
- Horton-Ikard, R.** (2009). Cohesive adequacy in the narratives samples of school-age children who use African American English. *Language, Speech, and Hearing Services in Schools*, 40, 393–402.
- Hyon, S., & Sulzby, E.** (1994). African American kindergartners' spoken narratives: Topic associating and topic centered styles. *Linguistics and Education*, 6, 121–152.
- Justice, L. M., Bowles, R., Kaderavek, J. K., Ukrainetz, T., Eisenberg, S., & Gillam, R.** (2006). The Index of Narrative Micro-Structure (INMIS): A clinical tool for analyzing school-aged children's narrative performance. *American Journal of Speech-Language Pathology*, 15, 1–15.
- Justice, L. M., Bowles, R., Pence, K., & Gosse, C.** (2010). A scalable tool for assessing children's language abilities within a narrative context: The NAP. *Early Childhood Research Quarterly*, 25, 218–234.
- Labov, W.** (1972). *Language in the inner city: Studies of Black English vernacular*. Philadelphia, PA: University of Pennsylvania Press.
- Leonard, L. B.** (1998). *Children with specific language impairment*. Cambridge, MA: MIT Press.
- Liles, B. Z.** (1985). Cohesion in the narratives of normal and language disordered children. *Journal of Speech and Hearing Research*, 28, 123–133.
- Liles, B. Z.** (1987). Episode organization and cohesive conjunctives in narratives of children with and without language disorder. *Journal of Speech and Hearing Research*, 30, 185–196.
- Mandler, J. M.** (1982). Some uses and abuses of a story grammar. *Discourse Processes*, 5, 305–318.
- Matsuyama, U. K.** (1983). Can story grammar speak Japanese? *The Reading Teacher*, 36, 666–669.
- Mayer, M.** (1967). *A boy, a dog, and a frog*. New York, NY: Dial Books.
- Mayer, M.** (1969). *Frog, where are you?* New York, NY: Dial Books.
- Mayer, M.** (1974). *Frog goes to dinner*. New York, NY: Dial Books.
- Mayer, M., & Mayer, M.** (1975). *One frog too many*. New York, NY: Dial Books.
- Melzi, G.** (2000). Cultural variations in the construction of personal narratives: Central American and European American mothers' elicitation styles. *Discourse Processes*, 30, 153–177.
- Michaels, S.** (1981). Sharing time: Children's narrative style and differential access to literacy. *Language and Society*, 10, 423–442.
- Michaels, S.** (1986). Narrative presentations: An oral preparation for literacy with first graders. In J. Cook-Gumperz (Ed.), *The social construction of literacy*. Cambridge, UK: Cambridge University Press.
- Miller, L., Gillam, R., & Peña, E.** (2003). *Dynamic assessment and intervention: Improving children's narrative abilities*. Circle Pines, MN: AGS.
- Pankratz, M. E., Plante, E., Vance, R., & Insalaco, D. M.** (2007). The diagnostic and predictive validity of the Renfrew Bus Story. *Language, Speech, and Hearing Services in Schools*, 38, 390–399.
- Pearson, B. Z.** (2001). Logic and mind in Spanish–English children's narratives. In L. Verhoeven & S. Stromqvist (Eds.), *Narrative development in a multilingual context* (pp. 373–398). Amsterdam, The Netherlands: John Benjamins.
- Pearson, B. Z.** (2002). Narrative competence among monolingual and bilingual school children in Miami. In D. K. Oller & R. E. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 135–174). Clevedon, UK: Multilingual Matters.
- Pearson, B. Z., & Ciolli, L.** (2004). Distinguishing dialect and development from disorder: Case studies. *Seminars in Speech and Language*, 25(1), 101–112.
- 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, 229–244.
- Pelligrini, A. D.** (1984). The effect of dramatic play on children's generation of cohesive text. *Discourse Processes*, 7, 57–67.
- Perera, K.** (1986). Language acquisition and writing. In P. Fletcher & M. Garman (Eds.), *Language acquisition* (pp. 494–518). Cambridge, UK: Cambridge University Press.
- Perner, J., Leekam, S., & Wimmer, H.** (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, 5, 125–137.
- Peterson, C., & McCabe, A.** (1983). *Developmental psycholinguistics: Three ways of looking at a child's narrative*. New York, NY: Plenum.
- Peterson, C. C., & Siegal, M.** (2000). Insights into theory of mind from deafness and autism. *Mind & Language*, 15, 123–145.
- Schneider, P., Dube, R. V., & Hayward, D.** (2005). *The Edmonton Narrative Norms Instrument*. Retrieved from <http://www.rehabmed.ualberta.ca/spa/enni>.
- Schraeder, T., Quinn, M., Stockman, I. J., & Miller, J.** (1999). Authentic assessment as an approach to preschool speech-language screening. *American Journal of Speech-Language Pathology*, 6, 195–200.
- Seymour, H. N., Roeper, T., & de Villiers, J. G.** (2000). *Dialect-Sensitive Language Test* (Unpublished pilot edition). San Antonio, TX: The Psychological Corporation.
- 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*. San Antonio, TX: The Psychological Corporation.
- Snow, C., Tabors, P., Nicholson, P., & Kurland, B.** (1995). SHELL: Oral language and early literacy skills in kindergarten

- and first grade. *Journal of Research in Childhood Education*, 10, 37–48.
- Stein, N. L., & Glenn, C. G.** (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), *New directions in discourse processing* (pp. 53–120). Norwood, NJ: Ablex.
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., . . . Levinson, S. C.** (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, 106(26), 10587–10592.
- Strong, C. J.** (1998). *The Strong Narrative Assessment Procedure*. Eau Claire, WI: Thinking Publications.
- Tager-Flusberg, H., & Sullivan, K.** (1995). Attributing mental states to story characters: A comparison of narratives produced by autistic and mentally retarded individuals. *Applied Psycholinguistics*, 16, 241–256.
- Ukrainetz, T. A., Justice, L. M., Kaderavek, J. N., Eisenberg, S. L., Gillam, R. B., & Harm, H. M.** (2005). The development of expressive elaboration in fictional narratives. *Journal of Speech, Language, and Hearing Research*, 48, 1363–1377.
- United States Bureau of the Census.** (2000). *Current population survey: School enrollment supplemental file* [CD ROM]. Washington DC: Author.
- Washington, J. A., & Craig, H. K.** (1998). Socioeconomic status and gender influences on children's dialectal variations. *Journal of Speech and Hearing Research*, 41, 618–626.
- Wellman, H. M., & Bartsch, K.** (1994). Before belief: Children's early psychological theory. In C. Lewis & P. Mitchell (Eds.), *Children's early understanding of mind: Origins and development*. Hillsdale, NJ: Erlbaum.
- Westby, C. E.** (1991). Learning to talk, talking to learn: Oral–literate language differences. In C. Simon (Ed.), *Communication skills and classroom success: Assessment and therapy methodologies for language and learning disabled students* (pp. 181–218). San Diego, CA: College Hill Press.
- Westby, C. E.** (1999). Assessing and facilitating text comprehension problems. In H. W. Catts & A. G. Kamhi (Eds.), *Language and reading disabilities* (pp. 154–221). Boston, MA: Allyn & Bacon.
- Wolfram, W.** (1991). *Dialects and American English*. Englewood Cliffs, NJ: Prentice Hall.

**APPENDIX A. PREVIOUS STUDIES SHOWING DEVELOPMENT, DISCRIMINATION,
AND DIALECT NEUTRALITY OF COHESION IN NARRATIVES**

AQ39

	<i>Cohesion</i>	<i>Evaluation</i>
Development	Westby, 1991 Dasinger & Toupin, 1994 Berman & Slobin, 1994 Berman, 1988 Hickmann, 2003	Westby, 1991 Bruner, 1986 Tager-Flusberg & Sullivan, 1995 Perner et al., 1987 Wellman & Bartsch, 1994 Bartsch & Wellman, 1995 Berman & Slobin, 1994 De Villiers et al., 2008 Pearson & Ciolli, 2004
Discrimination	Liles, 1985, 1987 Greenhalgh & Strong, 2001	Baron-Cohen et al., 1985 Bruner & Feldman, 1993 Tager-Flusberg & Sullivan, 1995 Bamberg & Damrad-Frye, 1991 Berman & Slobin, 1994 Petersen & Siegal, 2000 (Greenhalgh & Strong, 2001)
Dialect neutrality	Curenton & Justice, 2004 Horton-Ikard, 2009 de Villiers et al., 2010	(Holmes et al., 1996) Allen et al., 2001 Burns, 2004

AQ40

Note. References in parentheses are potential counter-examples in that category.

APPENDIX B (P. 1 OF 2). SCORING ELEMENTS AND SAMPLE STORIES (BASED ON FIGURE 1)

Story A. 1 point (of 7)/ 38 words, different words 19 (4-year-old, AA Male, PED level 2)

EXA: Look carefully at each picture to see what happened. I can't see the picture so tell me the whole story.

Elements of Narrative Index	Child's story	Points awarded
<p>1. REFERENCE CONTRAST Does the child identify each character with an adjective, proper noun, prepositional phrase, or other? No: ("girl" and "she" do not distinguish between the two girls)</p>	<p>She goin she g- she eating cake. The girl eating cake</p>	0
<p>2. Explicit TEMPORAL EXPRESSION Does the child use a sequencer (1) or an adverbial clause (2)? Yes: Sequencer—"then"</p>	<p>Then the girl goin outside. The girl th- the mouse is in the freezer and the girl went back in And the girl went went in the the cabinet But it wasn't there</p>	1
Probe question responses		
<p>3. THEORY OF MIND In picture 5, does the child refer to a cognition (2), desire or intention (1) or action only (0) No: Action only—0 points</p>	<p><i>EXA: Tell me again what's happening in this picture.</i> CHI: She she she she comin in.</p>	0
<p>4. Explanation of FALSE BELIEF Does the child explain why character is looking in a specific place (2), why the child is looking (desire) (1), or irrelevant (0) No: Shows no understanding of false belief – 0 points</p>	<p><i>EXA: The girl is looking for her cake in the cabinet. Why is she looking <u>there</u>?</i> CHI: Ca- Cause it was in the frigerator.</p>	0

Story B. 5 points (of 7)/ 79 words, different words 47 (6 yrs, AA, female, PED level 4)

Elements of Narrative Index	Child's story	Points awarded
<p>1. REFERENCE CONTRAST Does the child identify each character with an adjective, proper noun, prepositional phrase, or other? No: ("girl" does not make contrast between the two girls)</p>	<p>Once upon a time there was this girl she had some cake. She said You could play with my baseball if you give me cake.</p>	0
<p>2. Explicit TEMPORAL EXPRESSION Does the child use a sequencer or an adverbial clause? Yes: Adverbial clause—"when the girl came back"</p>	<p>Gave 'em cake– They put it in the cabinet and they shut it. and then they op- opened the refrigerator, she opened the refrigerator to put it in there. And when the girl came back in, she was thinking about it. And ah (looked under?) the cabinet and it it wasn't in there. That's the end.</p>	2

APPENDIX B (P. 2 OF 2). SCORING ELEMENTS AND SAMPLE STORIES (BASED ON FIGURE 1)

Story B. 5 points (of 7)/ 79 words, different words 47 (6 yrs, AA, female, PED level 4)

Elements of Narrative Index	Child's story	Points awarded
Probe question responses		
3. THEORY OF MIND In picture 5, does the child refer to a cognition (2), desire or intention (1) or action only (0) Yes: Mental state verb—thinking	<i>EXA: Tell me again what's happening in this picture.</i> CHI: She was thinking about the cake.	2
4. Explanation of FALSE BELIEF Does the child explain why the character is looking in a specific place (2), why the child is looking (desire) (1), or irrelevant (0) Yes: Shows desire, shows no understanding of false belief	<i>EXA: The girl is looking for her cake in the cabinet. Why is she looking <u>there</u>?</i> CHI: Cause, she wants her cake.	1

Story C. 7 points (of 7)/ 82 words, different words 36 (8 years, AA, Female, PED level 3)

Elements of Narrative Index	Child's story	Points awarded
1. REFERENCE CONTRAST Does the child identify each character with an adjective, proper noun, prepositional phrase, or other? Yes: Contrasts the big brother and the little boy.	The big brother has the toy in his hand, And he's reaching his arm up. And the little boy is trying to jump up for it.	1
2. Explicit TEMPORAL EXPRESSION Does the child use a sequencer or an adverbial clause? Yes: Uses both. Credit the adverbial clause—"when the big brother is eating"	Then the big brother hides the toy under the bed. And the little boy is crying. When the big brother is eating, he takes the toy under the bed and put it in his (toys box) toys box. And then when the big brother comes, he's thinking of the toy. And then he's reaching for the toy under his bed.	2
Probe question responses		
3. THEORY OF MIND In picture 5, does the child refer to a cognition (2), desire or intention (1) or action only (0) Yes: Mental state verb—thinking	<i>EXA: Tell me again what's happening in this picture.</i> CHI: He's thinking of the toy.	2
4. Explanation of FALSE BELIEF Does the child explain why the character is looking in a specific place (2), why the child is looking (desire) (1), or irrelevant (0) Yes: Full explanation of character's false belief	<i>EXA: The boy is looking for the toy under his bed. Why is he looking <u>there</u>?</i> CHI: Because that's where he put it.	2

Note. Exa = examiner; Chi = child. Word count does not include mazes and repetitions.