Volume 16 Article 2, January 2023

EBP *briefs*

A scholarly forum for guiding evidence-based practices in speech-language pathology

Assessing for Developmental Language Disorder in the Context of African American English

Isabelle Francois, MA
Stefanie Lapka, MS
University of Houston

Nan Bernstein Ratner, EdD
University of Maryland

Monique T. Mills, PhD
University of Houston

Pearson

EBP Briefs

Cite this document as:
Francois, I., Lapka, S., Bernstein Ratner, N., & Mills, M. T. (2023). Assessing for developmental language disorder in the context of African American English. *EBP Briefs, 16*(2), 1–8. NCS Pearson.

For inquiries or reordering:
800.627.7271
www.PearsonAssessments.com

NCS Pearson, Inc. 5601 Green Valley Drive Bloomington, MN 55437

Produced in the United States of America.

v.C

Structured Abstract

Clinical Question: For young AAE speakers, how useful is the Developmental Sentence Scoring (DSS) compared with Index of Productive Syntax (IPSyn) in identifying developmental language disorder (DLD) in the presence of African American English (AAE)?

Method: Structured Review

Study Sources: PsycInfo®, Education Source, Education Resources Information Center (ERIC), Communication & Mass Media Complete (CMMC), PubMed, Scopus, ASHAWire

Search Terms: (1) "African American English" OR "African American Language" OR "African American Vernacular English" OR "Black English" OR "AAE" OR "AAVE" AND (2) "child" AND (3) "language assessment" OR "language testing" OR "speech evaluation"

Number of Included Studies: 3

Primary Results: DSS and IPSyn appear to be dialect-neutral measures of morphosyntax in young AAE speakers. DSS was better able to detect morphosyntactic differences between children with typical language development (TLD) and children with DLD. DSS and its variant, Black English Sentence Scoring (BESS), appear to be clinically useful language sampling analysis tools.

Conclusions: Available evidence suggests that DSS is a more useful clinical tool over IPSyn for evaluating DLD within the context of AAE because it provides the opportunity to evaluate mastery and accuracy of grammatical features and not only the presence of structures.

Assessing for Developmental Language Disorder in the Context of African American English

Isabelle Francois, MA
Stefanie Lapka, MS
University of Houston

Nan Bernstein Ratner, EdD
University of Maryland

Monique T. Mills, PhD
University of Houston

Clinical Scenario

Anna is a speech-language pathologist (SLP) at a local elementary school. Like 92% of SLPs in the United States (American Speech-Language-Hearing Association [ASHA], 2019), she is White/non-Hispanic. She is tasked with administering a full evaluation for Marcus, an African American male child age 4 years 9 months, who has been referred for concerns with morphosyntax. His teacher, Ms. Williams—who, like 79.3% of teachers in the United States, is White/non-Hispanic (U.S. Department of Education & National Center for Education Statistics [NCES], 2020)—stated that she sometimes has a hard time understanding what he is trying to say because he does not speak in full sentences. Therefore, Anna uses the RIOT (review, interview, observe, test) procedure to review Marcus's educational and medical history, interview his parents, observe him in class, and test him with informal and formal assessments (Langdon, 2002).

From her review of Marcus's case history, Anna learns that Marcus is attending preschool for the first time this year. He is the youngest of four children and he lives with his parents who are both high school graduates. Marcus's parents report they also have a hard time understanding Marcus when he speaks; when Marcus talks about events, they are not sure if he is referring to the past, present, or future. They are concerned because their other children were speaking in full sentences and able to clarify temporality at his age. Anna notes characteristics of adult African American English (AAE) in the speech patterns of Marcus's parents during their interview.

As she observes Marcus answering questions in class, Anna notices that Marcus (1) demonstrates variable use of copula and auxiliary form of the verb *to be,* (2) uses multiple negation within clauses, (3) does not mark plurals morphologically or use quantifiers, (4) does not mark past tense with inflection or context, and (5) does not denote noun or pronominal possessives. Therefore, she decides to administer the Diagnostic Evaluation of Language Variation™ (DELV™) – Screening Test (Seymour et al., 2003) to characterize Marcus's dialect variation and to determine his risk for developmental language disorder (DLD). The DELV – Screening Test is a criterion-referenced measure composed of two parts. By probing for targets that are contrastive across dialects of American English (e.g., third person singular), Part I of the DELV – Screening Test provides criterion scores to classify children, ages 4–12 years, as having no, some, or strong variation from Mainstream American English (MAE). By probing for targets that are noncontrastive across dialects of American English (e.g., nonword repetition), Part II provides diagnostic scores to categorize children, ages 4–9 years, with one of the following risk groups for DLD: lowest, low-to-medium, medium-to-high, or highest. Part I of the DELV – Screening Test classified Marcus as having strong variation from MAE, whereas Part II of the DELV – Screening Test placed Marcus in the highest risk for DLD.

Next, Anna administered the Structured Photographic Expressive Language Test (3rd ed.; SPELT-3; Dawson et al., 2003). The SPELT-3 assesses use of morphologic and syntactic structures in children ages 4:0–9:11. The SPELT-3 uses colorful photographs as its visual stimuli. The auditory stimuli consist of statements or questions that the examiner says when presenting the pictures (e.g., Tell me about this picture). A response is then given by the child.

Results from Marcus's evaluation indicated a SPELT-3 raw score of 16, and a standard score of 77. Marcus's

standard score fell approximately 1.5 *SD*s (standard deviations) below the average performance of children his age on the SPELT-3 (mean = 100, standard deviation = 15). His standard score was within a 95% confidence interval such that his true standard score was likely between 68 and 85. Moreover, Marcus's percentile rank was 8, indicating that only 7% of children in his age range would receive a score lower than his.

Based on his performance on formal assessments, reports of difficulties with expressive morphosyntax, and best assessment practices (ASHA, 2004; Council for Clinical Certification in Audiology and Speech-Language Pathology of the American Speech-Language-Hearing Association, 2013), Anna decides it wise to additionally elicit a language sample during free play as an informal assessment. Language sample analysis is widely used to examine morphosyntax in young AAE speakers (Stockman, 1996). After collecting the language sample, Anna decides to conduct a literature search for language sampling analysis procedures appropriate for assessing morphosyntax in AAE-speaking children with a goal to accurately assess clinical markers of DLD in AAE.

Background Information

Language variation is composed of differences in sound and grammar patterns across groups of speakers. Within a single language there are multiple language varieties conventionally known as dialects. Dialects develop because of geographic and social influences such as gender, age, and cultural background. For example, American English includes several dialects such as MAE, AAE, Appalachian English, and Chicano English. In the United States, the most widespread dialect is MAE (also referred to as General American English, Standard American English, and idealized English) and is traditionally the linguistic variety employed by government, mass media, business, and education in the United States (ASHA, 2003). Although MAE is the dialect of wider communication (Wolfram & Schilling, 2015), AAE is spoken by many African American children at the time of school entry (Craig & Washington, 2006) and is, therefore, one of the most researched American English dialects.

AAE is a dialect of American English with set rules governing its grammar, morphology, syntax, phonology, semantics, pragmatics, and paralinguistics (Green, 2002; Hamilton, 2020). AAE has been referred to as Negro Dialect, Nonstandard Negro English, Black English, Vernacular Black English, Afro-American English, Ebonics, and African American Vernacular English. Scholars in education have referred to AAE as African American Language (Smitherman, 2006), and, most recently, as Black Language (Baker-Bell, 2020).

Accurate assessment of AAE speakers has been a challenge because of limited resources and the deficit language used to describe it. The nomenclature used to describe the linguistic patterns of AAE resembles descriptions of DLD (e.g., zero plural marker). This makes it difficult for clinicians to discriminate between AAE and DLD (Hamilton, 2020; Hamilton et al., 2018). Shifting the language used when discussing assessment of AAE speakers and the linguistic system used is necessary for accurate identification of DLD. Hamilton (2020) suggested using positive, noncomparative description of features in AAE (e.g., plurals may be marked by a numeral or quantifier preceding the noun).

Relatedly, another prominent challenge that arises when faced with assessment of AAE speakers is the overlap of features of AAE with features that are clinical markers of DLD in MAE speakers (Lee & Oetting, 2014; Oetting & McDonald, 2001). Oetting and colleagues (2016) discuss the implications of identifying disorder within dialect. Although it is well known that a dialect is a rule-governed systematic way of speaking that should not be misinterpreted as a language disorder, disorder and dialect are not mutually exclusive entities. In the past, the discussion of "disorder versus dialect" has given the impression that an individual must lie in either one category or the other. The authors point out that the relationship between the terms is a more complicated concept in which everyone speaks some form of dialect; however, within that population are a smaller number of those who present with a language disorder within their respective dialect (Oetting et al., 2016).

These challenges contribute to the misidentification of DLD in AAE-speaking children (Hendricks & Adolf, 2017). Prioritizing accurate assessment starts with distinguishing between patterns of variation that represent features of the speaker's dialect and patterns that represent disorders in language. Although efforts are underway to identify dialect-fair measures in AAE speakers with typical language development (Mills & Fox, 2016; Mills et al., 2017; Mills et al., 2021), more work is needed to improve access to accurate language assessments for AAE speakers with DLD.

Common measures used in the process of language sampling are Developmental Sentence Scoring (DSS; Lee

& Canter, 1971) and Index of Productive Syntax (IPSyn; Scarborough, 1990). Both these measures are based on MAE syntax and the normative scores are derived from a sample that no longer represents the demographics of the average SLP's caseload. This presents the risk of underestimating language performance of AAE speakers with DLD (Craig & Washington, 1994); thus, there is a need to determine how well the two measures identify DLD in AAE speakers.

DSS is a clinical procedure developed to serve as a scale of syntax acquisition that is based on the general order in which children with typical language development (TLD) use a set of grammatical elements in their expressive language. The child's grammar is measured against adult MAE; a minimum of 50 utterances is necessary for analysis. Each utterance is rated for use of specific words from eight grammatical categories: indefinite pronouns, noun modifiers, personal pronouns, primary verbs, secondary verbs, negatives, conjunctions, interrogative reversals, and wh questions. Each grammatical form that the child uses receives a weighted score (1–8) for correct use, with higher scores given to more complex and later-developing forms than to simpler and earlier-acquired forms. At least 50 utterances containing a noun and a verb are required for the completion of the DSS analysis. Additional criteria specify that the elicited utterances be complete (i.e., have at least a noun and a verb) and unique (e.g., a child cannot receive points for either echolalic responses or overused phrases such as "I don't know"). Repeated phrases may only be counted once. Fifty consecutive utterances must be selected even if they are taken from a larger sample to avoid only selecting the highest-scoring utterances. All utterances must be intelligible. An additional sentence point is awarded to all utterances that are grammatical and error free. Relevant to our question, a clinician must be aware of acceptable dialect variation to avoid marking AAE structures as incorrect.

To address potential bias in language sample analysis, Black English Sentence Scoring (BESS; Hyter, 1984) was developed as a scoring adjustment to DSS. BESS gives credit for features of MAE as well as AAE that would not be counted as correct on DSS. These features include past tense marked by inflection and context, negation inversion, and multiple negation to intensify a negation. BESS has demonstrated an ability to differentiate children with TLD from those with DLD (Hyter, 1984). BESS shares a strong positive correlation with the Structured Photographic Expressive Language Test (SPELT; Werner & Kresheck, 1983).

In addition to the DSS and its BESS variation, IPSyn is another tool for analyzing syntactic structure. IPSyn has been used to assess morphology and syntax in AAE-speaking toddlers (Horton-Ikard et al., 2005) and preschoolers (Oetting & Pruitt, 2005; Oetting et al., 2010; Overton et al., 2021). IPSyn consists of 59 target structures that awards points for up to two instances observed in a child's language sample and was designed to be a "summary scale of grammatical complexity that would be appropriate for the study of individual differences in language acquisition" (Scarborough, 1990, p. 1). In contrast to DSS, which tallies grammar at the word level, IPSyn assesses the variety of phrase structures in the child's sample. Classically, IPSyn appraises a language sample for production of 56 syntactic features under four grammatical domains: noun phrases, verb phrases, questions/negations, and sentence structure. A recent revision (Yang et al., 2022) that has been added to the computerized language analysis (CLAN; MacWhinney, 2000), reduces the required sample length from 100 to 50 utterances and eliminates structures that were determined to be psychometrically unstable or redundant with other structural categories. Unlike DSS, IPSyn does not credit grammaticality, only the surface structure of the child's utterances. Finally, both DSS and IPSyn have utility beyond diagnostic purposes. In contrast to a measure such as mean length of utterance (MLU), DSS and IPSyn may provide clinical insights into structures that the child does not appear to be able to use productively for clinical planning purposes. MLU would only suggest that a child needs to "produce longer utterances"; both DSS and IPSyn will provide clues about structures that the clinician might want to instruct to achieve that aim (Overton et al., 2021; Yang, et al., 2022).

Clinical Question

Anna's chief concern in her assessment of Marcus is to find a measure of morphosyntax that has been used with AAE speakers and is effective for identifying DLD in this population of young children. She needs to determine which measure, DSS or IPSyn, would best help her distinguish TLD from DLD with AAE speakers. To structure her question, Anna considered the clinical problem through the framework of a PICO question (Straus et al., 2018). A PICO question traditionally considers four aspects of the clinical problem: (1) the population (P), (2) the intervention (I), (3) the comparison group (C), and (4) the outcome (O). Because Anna was interested in assessment

rather than intervention, she adopted a broad definition of the intervention component of her question to include the assessment she used with her population of interest. This definition of intervention for the purposes of a PICO question is consistent with PICO definitions that consider intervention as "defined very broadly including an exposure, a diagnostic test, a prognostic factor, a treatment, a patient perception …" (Straus et al., 2018, p. 21). Thus, Anna's PICO question was: For young AAE speakers (P), how useful is DSS (I) compared with IPSyn (C) in identifying DLD in the presence of AAE (O)?

Search for the Evidence

Anna began her search by establishing the following selection criteria for all studies returned:

Inclusion and Exclusion Criteria:

- Although AAE is spoken by some Black Canadians (Edwards, 2004), because of differences in primary education and expectations across countries, studies had to be completed within the United States
- Children had to be enrolled in either preschool (Pre-K) or early elementary school (Grades K–2)
- Children had to be native speakers of AAE
- If bilingual students were included in the study, separate analysis had to be reported for a monolingual, English-speaking sample
- Children could not have additional known developmental impairment or cognitive impairment
- Measures of child morphosyntax had to be either DSS or IPSyn, and elicitation tasks had to be either conversation or free play
- Grammatical utterance and other language skills, such as reading and writing, could not be the primary target of the assessment
- Studies must have been original research and must have been peer reviewed
- Research design and publication date did not limit study inclusion

Search Strategy

To capture most relevant papers, Anna developed a comprehensive search strategy with three primary search term categories: (1) "African American English" OR "African American Language" OR "African American Vernacular English" OR "Black English" OR "AAE" OR "AAVE" AND (2) "child" AND (3) "language assessment" OR "language testing" OR "speech evaluation." Anna decided to search six subject databases to capture as many papers as possible. These databases included PubMed, PsycInfo, Education Source, Education Resources Information Center (ERIC), Communication & Mass Media Complete (CMMC), and Scopus. She also conducted a search of the journals of the American Speech-Language-Hearing Association via the ASHAWire database.

After Anna completed her search across each of the seven online databases, she had identified 754 papers. The open-source reference management software Zotero (Corporation for Digital Scholarship, 2013) identified 410 duplicate papers, and Anna reviewed the title and abstract of the remaining 345 papers. From the review of the titles and abstracts, Anna identified 14 to read in full to determine eligibility. After scanning the full text of all 14 papers, she found that most (n = 11) did not meet her inclusionary criteria. Four papers were excluded after further review; there was too little information on the population, task, and measure to be gleaned from the abstract. Seven papers were removed because although they focused on the assessment of morphosyntax, they did not examine it through DSS or IPSyn. Anna did, however, find three papers that were suitable to include in her review (Hyter, 1984; Oetting et al., 2010; Overton et al., 2021). Although Hyter (1984) is not a peer-reviewed article, it is a master's thesis that proposed scoring adjustments to BESS that are specific to AAE. Given the dearth of relevant papers on the topic, Anna included Hyter (1984) in her search results. She then reviewed the reference lists of all three papers to check for any additional papers that may not have been captured in her database search, but this search did not yield any additional studies. Because the current body of work is sparse, the low number of research articles that met Anna's selection criteria did not surprise her.

Evaluating the Evidence

Anna reviewed the three papers found in her comprehensive search to ascertain which measure would best guide her analysis of Marcus's language sample. The three studies included a total of 116 children between the ages of 3:8 and 6:11. Because Anna found pertinent external evidence, she then determined the level of

evidence represented by each study, as recommended in the communication sciences and disorders literature (Dollaghan, 2004; Gillam & Gillam, 2006). Anna's question relates to assessment rather than treatment, so she made use of a critical appraisal tool (CAT) centered on quantitative studies (Long et al., 2002). The CAT allowed Anna to evaluate each article across the following six areas: study overview; student, setting, sample, and ethics; ethics; group comparability and outcome measurement; policy and practice implications; and other comments.

Based on her CAT evaluation, Anna determined that Hyter (1984)—a validation study of BESS—provided the weakest level of evidence; it presented a master's thesis study that did not undergo peer review and did not include a comparison group of AAE-speaking children with TLD. Oetting et al. (2010) and Overton et al. (2021)—validation studies of IPSyn and of IPSyn and DSS, respectively—compared AAE speakers with TLD to AAE speakers with DLD and matched groups on variables such as age, clinical status, and maternal education level. All three studies used current enrollment in speech-language pathology services as a criterion for DLD status. However, Oetting et al. (2010) validated DLD status against a battery of tests, ensuring that children scored 1 *SD* (standard deviation) below the mean on two norm-referenced language tests and 1 *SD* above the mean on a norm-referenced test of nonverbal intelligence. Because both Hyter (1984) and Oetting et al. (2010) sampled from a restricted population of children from Detroit and southeastern Louisiana, respectively, the results from these studies are less generalizable than those of Overton et al. (2021) which included a national sample. Anna found the Overton et al. (2021) study most useful to her clinical decision-making because it directly compared performance on IPSyn to DSS across a national sample of AAE speakers with TLD versus DLD. The results of each study that Anna reviewed are summarized here.

Oetting et al. (2010) examined whether all 56 IPSyn items are appropriate for children who speak AAE. This study examined several factors that affect the validity of IPSyn for use with children who speak AAE. The study addressed whether all IPSyn items are appropriate for children who speak AAE and whether IPSyn scores varied based on age, clinical status, maternal educational level, and use of nonmainstream dialect density. The study also examined whether IPSyn was sensitive to age-related changes in AAE speakers' grammar between ages 4 and 6 years and compared performance of children with TLD to children with DLD. Sixty-two African American children ages 4–6 years participated in a play session to obtain a language sample that was then transcribed and scored following IPSyn scoring procedures. An item analysis was completed to determine the percentage of AAE speakers who earned a score of 0 on 17 items that are expected to vary based on AAE production. Of those 17 potentially biased items, only one led to 55% of AAE speakers who earned 0 points (e.g., Y/N question with inverted modal, copula, auxiliary). Furthermore, children's IPSyn scores did not vary as a function of important indexes such as age, clinical status, socioeconomic status, and AAE density. Thus, IPSyn was dialect-neutral, but not clinically sensitive.

Hyter (1984) assessed the concurrent validity and interscorer reliability of BESS for 17 AAE speakers with DLD, ages 3–7 years. BESS is the adaptation of the DSS system for use with children who speak or are learning to speak AAE. Like DSS, BESS involves rating each utterance for use of specific words from eight grammatical categories: indefinite pronouns, noun modifiers, personal pronouns, primary verbs, secondary verbs, negatives, conjunctions, interrogative reversals, and wh questions. The scoring system of BESS differs from DSS in that it credits AAE features not awarded in DSS. For example, BESS credits variable and systemic use of copula and auxiliary forms of the verb *to be* and third person singular presentations that retain the same form in person and number. Correlation analyses of 17 children's BESS scores with scores on the Screening Kit of Language Development (SKOLD; Bliss & Allen, 1983) and the SPELT revealed a high positive statistically significant correlation between BESS and the SPELT as well as a statistically significant (r = .99) but less strong correlation between the BESS and the SKOLD (r = .68). In terms of diagnostic accuracy, the SPELT held an advantage over the SKOLD and BESS. Ninety-four percent of children with DLD fell below the cutoff score on the SPELT. On the SKOLD, 82% of children with DLD fell below the cutoff score. Finally, BESS scores placed 82% of children with DLD below mean performance. Thus, the SPELT, the SKOLD, and BESS offer converging support for DLD in AAE speakers.

In a retrospective validation study, Overton et al. (2021) examined how well DSS and IPSyn identified DLD in children who speak AAE using KidEval, a free, open-access computerized language analysis program (CLAN) within the TalkBank project (talkbank.org; MacWhinney, 2007). The sample included 37 children under the age of 6 years (15 with DLD). Summary scores for DSS and IPSyn

were calculated. A language sample of 50–100 utterances was elicited during a variety of tasks including conversation, prompts for personal narrative, picture sequence, and exposition. Whereas IPSyn correctly identified 33% of child AAE speakers with DLD, DSS correctly identified 73% of child AAE speakers with DLD. In addition, DSS correctly identified 100% of child AAE speakers with TLD. Further, IPSyn scores tended to level off after 4 years.

Taken together, the results of Anna's structured review indicate that both DSS and IPSyn are dialect-fair measures for evaluating language samples of young AAE-speaking children. However, DSS offers more diagnostic accuracy at Marcus's age than does IPSyn. Moreover, BESS—a DSS adjustment—can be applied to language samples to improve diagnostic accuracy.

The Evidence-Based Decision

After evaluating the two measures through a structured review, Anna decides to complete her language sample analysis using DSS instead of IPSyn. She is confident that she will arrive at a description of Marcus's morphosyntax performance during a free-play task to complement his performance on formal assessments and a diagnosis of DLD in the context of AAE. Anna will use DSS scores to identify treatment targets: Syntactic structures that are absent in Marcus's language sample will be used to structure potential intervention goals.

Because negative and interrogative reversals subscales of DSS were inherently unstable in a recent psychometric study and because play sampling may not provide an opportunity for complex syntax forms (Yang et al., 2022), Anna will probe these syntactic structures separately, within the context of a sentence. For example, to prompt for production of a negative, she might show Marcus a picture and ask, *Why is the boy pushing the toy away?* Marcus is then expected to reply with the following target form: He *doesn't* want to play. Once treatment begins, Anna will elicit another language sample within a free-play task and apply DSS to monitor Marcus's progress in treatment.

As a future direction, Anna will collaborate with a child language researcher at a nearby university to learn how to use CLAN's KidEval program to apply both DSS and BESS analyses to play samples. She is excited to compute DSS and BESS is a way that is automated, saving time spent in intensive hand coding. Anna will serve as a guest speaker in this professor's child language disorders course, presenting on the use of language sample analysis in school-based speech-language pathology. She will also lead a journal club with other SLPs in her district who are interested in sharpening their skills in language sampling analysis. Finally, Anna will share insights learned about syntax development during the preschool years with the parent–teacher association and teachers in her building, engaging them in conversation about how best to support syntax development in preschoolers from various cultural and linguistic backgrounds.

Authors' Note

Address correspondence concerning this article to:
Isabelle Francois
https://orcid.org/0009-0000-6477-9744

Stefanie Lapka
https://orcid.org/0000-0002-9358-6068

Nan Bernstein Ratner
https://orcid.org/0000-0002-9947-0656

Monique Mills
https://orcid.org/0000-0003-2357-8318

Conflict of Interest

There are no relevant conflicts of interest.

Funding

Three of the authors are supported by the National Institute on Deafness and Other Communication Disorders Grant R01DC017152 (Nan Bernstein Ratner, Principal Investigator). The CLASP grant consultants include Jan Edwards, Brian MacWhinney, Monique Mills, and Barbara Zurer Pearson.

References

American Speech-Language-Hearing Association. (2003). *American English dialects* [Technical report]. https://www.asha.org/policy/tr2003-00044/

American Speech-Language-Hearing Association. (2004). *Knowledge and skills needed by speech-language pathologists and audiologists to provide culturally and linguistically appropriate services* [Knowledge and skills]. https://www.asha.org/policy

American Speech-Language-Hearing Association. (2019). A demographic snapshot of SLPs: Data highlight some key characteristics of ASHA's SLP members. *The ASHA Leader, 24*(7), 32. https://doi.org/10.1044/leader.AAG.24072019.32

Baker-Bell, A. (2020). *Linguistic justice: Black language, literacy, identity, and pedagogy.* Routledge. https://doi.org/10.4324/9781315147383

Bliss, L. S., & Allen, D. V. (1983). *Screening kit of language development.* University Park Press.

Corporation for Digital Scholarship. (2013). *Zotero* [Computer software]. https://www.zotero.org

Council for Clinical Certification in Audiology and Speech-Language Pathology of the American Speech-Language-Hearing Association. (2013). *2014 standards and implementation procedures for the Certificate of Clinical Competence in Speech-Language Pathology.* https://www.tamuk.edu/artsci/_files_COAS/CSDO/AboutCSDOFiles/2014SLPCertificationStandards.pdf

Craig, H. K., & Washington, J. A. (1994). The complex syntax skills of poor, urban, African-American preschoolers at school entry. *Language, Speech, and Hearing Services in Schools, 25*(3), 181–190. https://doi.org/10.1044/0161-1461.2503.181

Craig, H. K., & Washington, J. A. (2006). *Malik goes to school: Examining the language skills of African American students from preschool–5th grade.* Psychology Press. https://doi.org/10.4324/9781410615602

Dawson, J., Stout, C., & Eyer, J. (2003). *Structured Photographic Expressive Language Test* (3rd ed.). PRO-ED.

Dollaghan, C. A. (2004). Evidence-based practice in communication disorders: What do we know, and when do we know it? *Journal of Communication Disorders, 37*(5), 391–400. https://doi.org/10.1016/j.jcomdis.2004.04.002

Edwards, W. F. (2004). African American Vernacular English: Phonology. In E. W. Schneider (Ed.), *Varieties of English: The Americas and the Caribbean* (Vol. 2, pp. 181–191). De Gruyter Mouton. https://doi.org/10.1515/9783110208405.1.181

Gillam, S. L., & Gillam, R. B. (2006). Making evidence-based decisions about child language intervention in schools. *Language, Speech, and Hearing Services in Schools, 37*(4), 304–315. https://doi.org/10.1044/0161-1461(2006/035)

Green, L. J. (2002). *African American English: A linguistic introduction.* Cambridge University Press. https://doi.org/10.1017/CBO9780511800306

Hamilton, M.-B. (2020). An informed lens on African American English. *The ASHA Leader, 25*(1), 46–53 . https://leader.pubs.asha.org/doi/full/10.1044/leader.FTR1.25012020.46

Hamilton, M.-B., Mont, E. V., & McLain, C. (2018). Deletion, omission, reduction: Redefining the language we use to talk about African American English. *Perspectives of the ASHA Special Interest Groups, 3*(1), 107–117. https://doi.org/10.1044/persp3.SIG1.107

Hendricks, A. E., & Adlof, S. M. (2017). Language assessment with children who speak nonmainstream dialects: Examining the effects of scoring modifications in norm-referenced assessment. *Language, Speech, and Hearing Services in Schools, 48*(3), 168–182. https://doi.org/10.1044/2017_lshss-16-0060

Horton-Ikard, R., Ellis Weismer, S., & Edwards, C. (2005). Examining the use of standard language production measures in the language samples of African-American toddlers. *Journal of Multilingual Communication Disorders, 3*(3), 169–182. https://doi.org/10.1080/14769670500170768

Hyter, Y. D. (1984). *Reliability and validity of the Black English Sentence Scoring System* [Unpublished master's thesis]. Western Michigan University. https://scholarworks.wmich.edu/masters_theses/1490

Langdon, H. W. (2002). *Interpreters and translators in communication disorders: A practitioner's handbook.* Thinking Publications.

Lee, L. L., & Canter, S. M. (1971). Developmental sentence scoring: A clinical procedure for estimating syntactic development in children's spontaneous speech. *Journal of Speech and Hearing Disorders, 36*(3), 315–340. https://doi.org/10.1044/jshd.3603.315

Lee, R., & Oetting, J. B. (2014). Zero marking of past tense in child African American English. *Perspectives on Language Learning and Education, 21*(4), 173–181. https://doi.org/10.1044/lle21.4.173

Long, A. F., Godfrey, M., Randall, T., Brettle, A., & Grant, M. J. (2002). *Developing evidence based social care policy and practice. Part 3: Feasibility of undertaking systematic reviews in social care.* University of Leeds & University of Salford. https://usir.salford.ac.uk/id/eprint/13071/1/Long_et_al_2002_Feasibility_Social_Care_Review_-_part_III.pdf

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Erlbaum.

MacWhinney, B. (2007). The talkbank project. In J. C. Beal, K. P. Corrigan, & H. L. Moisl (Eds.), *Creating and digitizing language corpora* (pp. 163–180). Palgrave Macmillan. https://doi.org/10.1057/9780230223936

Mills, M. T., & Fox, M. (2016). Language variation and theory of mind in typical development: An exploratory study of school-age African American narrators. *American Journal of Speech-Language Pathology, 25*(3), 426–440. https://doi.org/10.1044/2016_ajslp-15-0038

Mills, M. T., Mahurin-Smith, J., & Steele, S. C. (2017). Does rare vocabulary use distinguish giftedness from typical development? A study of school-age African American narrators. *American Journal of Speech-Language Pathology, 26*(2), 511–523. https://doi.org/10.1044/2016_ajslp-15-0157

Mills, M. T., Moore, L. C., Chang, R., Kim, S., & Frick, B. (2021). Perceptions of Black children's narrative language: A mixed-methods study. *Language, Speech, and Hearing Services in Schools, 52*(1), 84–99. https://doi.org/10.1044/2020_LSHSS-20-00014

Oetting, J. B., Gregory, K. D., & Rivière, A. M. (2016). Changing how speech-language pathologists think and talk about dialect variation. *Perspectives of the ASHA Special Interest Groups, 1*(16), 28–37. https://doi.org/10.1044/persp1.SIG16.28

Oetting, J. B., & McDonald, J. L. (2001). Nonmainstream dialect use and specific language impairment. *Journal of Speech, Language, and Hearing Research, 44*(1), 207–223. https://doi.org/10.1044/1092-4388(2001/018)

Oetting, J. B., Newkirk, B. L., Hartfield, L. R., Wynn, C. G., Pruitt, S. L., & Garrity, A. W. (2010). Index of Productive Syntax for children who speak African American English. *Language, Speech, and Hearing Services in Schools, 41*(3), 328–339. https://doi.org/10.1044/0161-1461(2009/08-0077)

Oetting, J. B., & Pruitt, S. (2005). Southern African-American English use across groups. *Journal of Multilingual Communication Disorders, 3*(2), 136–144. https://doi.org/10.1080/14769670400027324

Overton, C., Baron, T., Pearson, B. Z., & Ratner, N. B. (2021). Using free computer-assisted language sample analysis to evaluate and set treatment goals for children who speak African American English. *Language, Speech, and Hearing Services in Schools, 52*(1), 31–50. https://doi.org/10.1044/2020_LSHSS-19-00107

Scarborough, H. S. (1990). Index of Productive Syntax. *Applied Psycholinguistics, 11*(1), 1–22. https://doi.org/10.1017/s0142716400008262

Seymour, H. N., Roeper, T. W., de Villiers, J. (with de Villiers, P. A., & Pearson, B. Z.). (2003). *Diagnostic Evaluation of Language Variation—Screening Test.* Ventris Learning.

Smitherman, G. (2006). *Word from the mother: Language and African Americans.* Routledge. https://doi.org/10.4324/9780203006665

Stockman, I. J. (1996). The promises and pitfalls of language sample analysis as an assessment tool for linguistic minority children. *Language, Speech, and Hearing Services in Schools, 27*(4), 355–366. https://doi.org/10.1044/0161-1461.2704.355

Straus, S. E., Glasziou, P., Richardson, W. S., & Haynes, R. B. (2018). *Evidence-based medicine: How to practice and teach EBM* (5th ed.). Elsevier.

U.S. Department of Education & National Center for Education Statistics. (2020). *Data point: Race and ethnicity of public school teachers and their students.* https://nces.ed.gov/pubs2020/2020103/index.asp

Werner, E., & Kresheck, J. (1983). *Structured Photographic Expressive Language Test.* Janelle Publications.

Wolfram, W., & Schilling, N. (2015). *American English: Dialects and variation* (3rd ed.). Wiley.

Yang, J. S., MacWhinney, B., & Ratner, N. B. (2022). The Index of Productive Syntax: Psychometric properties and suggested modifications. *American Journal of Speech-Language Pathology, 31*(1), 239–256. https://doi.org/10.1044/2021_AJSLP-21-00084