EBP
briefs

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

Examining the Utility of Scoring Modifications as a Culturally Responsive Approach to Language Assessment in Speakers of African American English

Lakeisha Johnson, PhD, CCC-SLP
Shaylyn King-St. Remy, MEd, CCC-SLP
Jasmine Smith, MA, CCC-SLP
School of Communication Science and Disorders
Florida State University

Pearson

EBP Briefs

Cite this document as:
Johnson, L., King-St. Remy, S., & Smith, J. (2024). Examining the utility of scoring modifications as a culturally responsive approach to language assessment of speakers of African American English. *EBP Briefs, 16*(4), 1–9. NCS Pearson.

Volume 16, Article 4
September 2024

Examining the Utility of Scoring Modifications as a Culturally Responsive
Approach to Language Assessment in Speakers of African American English

Structured Abstract

Clinical Question: Among school-age children who speak AAE (P), does the use of dialect scoring modifications (I), when compared to traditional scoring methods (C), yield more reliable identification of true language disorders (O)?

Method: Structured Review

Study Sources: Google Scholar and ASHAWire

Search Terms: (1) "dialect" OR "AAE" OR "scoring modifications" AND "language" AND (2) "scoring modifications for AAE dialect users"

Number of Included Studies: 3

Primary Results:

- Dialect status should be confirmed before utilizing scoring modifications.

- Scoring modifications are recommended when standardized tests must be given to students who exhibit linguistic variations from General American English, but caution should be exercised. Students may obtain higher scores with modifications, but there are risks for false positives or false negatives, which impacts diagnostic accuracy.

- When comparing the performance of students with and without language disorders on a sentence recall task, dialect-strategic scoring modifications provide moderate to high levels of diagnostic accuracy.

Conclusions: Limited standardized, norm-referenced tests currently exist that offer detailed scoring modifications to accurately diagnose language impairment in children who use a variety of dialects. Additional research is needed to determine the validity of scoring modifications on a variety of language-based tasks beyond sentence recall and language sampling. SLPs should not rely solely on standardized measures when making clinical decisions for dialect speakers; they should consider a wide range of assessment practices that have been found to be less biased and culturally responsive (e.g., dynamic assessment, language sampling, and processing-based tasks).

Examining the Utility of Scoring Modifications as a Culturally Responsive Approach to Language Assessment in Speakers of African American English

Lakeisha Johnson, PhD, CCC-SLP
Shaylyn King-St. Remy, MEd, CCC-SLP
Jasmine Smith, MA, CCC-SLP
School of Communication Science and Disorders
Florida State University

Clinical Scenario

Ramon, a Cuban American speech-language pathologist (SLP), worked in an elementary school for a public district in the southeastern United States. He received a new student named Kamari during the middle of the academic year and must complete a reevaluation to determine if services are still needed. Upon reviewing Kamari's previous school records, Ramon discovered that Kamari was in the third grade and received specialized instruction for expressive language through an individualized education plan (IEP) aimed at addressing weaknesses in morphosyntax. The review of records reported that Kamari had no additional documented concerns related to academics or communication. Additionally, documents indicated that Kamari received his last language assessment in preschool. Considering this and other relevant factors at the appropriately convened IEP meeting, the team decided to reevaluate Kamari's language skills.

While working directly with Kamari, Ramon began to suspect that his expressive language was characteristic of African American English (AAE) dialect. Ramon met with Kamari's current classroom teacher to discuss his performance on oral and written language tasks in the academic setting. Kamari's teacher reported that he sometimes omitted plural *-s* forms, past tense *-ed* verb markers, and used multiple negatives in oral and written language tasks. Ramon observed similar patterns during therapy sessions. For example, when recalling a recent personal event Kamari stated, "Me and my sister was going to the movies over the weekend, but we changed our minds."

For Kamari's reevaluation, Ramon initially administered the Diagnostic Evaluation of Language Variation (DELV™) Screening Test (Seymour et al., 2003) to confirm whether Kamari's language varied from Generalized American English (GAE). Based on his performance, Kamari exhibited strong variation from GAE. Ramon reviewed research articles that discussed the characteristics of child AAE (Seymour et al., 1998; Thompson et al., 2004) and observed that many of these features were present in Kamari's speech. Ramon aimed to use a variety of alternative informal assessment procedures to gather more information about Kamari's expressive language skills, such as obtaining a language sample and conducting a contrastive feature analysis. Ramon knew that many standardized language tests he had access to do not account for linguistic variation, but his district still requires a standard score. However, he wondered how to best use standardized language assessments and informal measures to gather diagnostic information. He recalled learning that scoring modifications were one way to be culturally responsive to linguistic variation. Therefore, Ramon conducted a literature search for evidence regarding scoring modifications on formal and informal language assessments. He sought to determine if scoring modification practices would aid in more reliable identification of a true language disorder compared to scoring Kamari's responses based on traditional scoring methods.

Background Information

African American English (AAE) is a systematic, rule-governed dialect of English that many African American students enter school speaking various amounts of (Thompson et al., 2004). It is differentiated in all language domains but namely in language form that includes phonology, morphology, and syntax (Seymour et al., 1998). Its use varies across a continuum from low to high, which is considered dialect density: the rate at which speakers produce dialect features. Dialect density varies based on several factors including geographical region, gender, age,

Volume 16, Article 4
September 2024

Examining the Utility of Scoring Modifications as a Culturally Responsive
Approach to Language Assessment in Speakers of African American English

and socioeconomic status (SES; Craig & Washington, 2004; Terry, Connor, et al., 2010; Washington & Craig, 1998). Characteristics of AAE overlaps with many other dialects of English, including Southern White English, Gullah-Geechee, and Hispanic-Influenced English (Berry & Oetting, 2017; Gutiérrez-Clellen et al., 2008; Oetting & McDonald, 2001). See a full description of AAE features and characteristics in Thompson et al. (2004).

Linguistic distance is a concept often used to describe how languages or dialects differ from one another. In the United States, the magnitude of this distance is often how SLPs attempt to measure whether a child has a language impairment (Lee-James & Johnson, 2022). Thinking of language in this way, as most SLPs are trained, can be problematic because it reinforces standardized language ideology—the belief that the standard variety of a language is superior and is the only correct way it can be used across academic and professional settings (Lippi-Green, 2004). Because AAE has often been stigmatized as a broken and illiterate form of English, it is important for SLPs, educators, and other practitioners to have a deeper understanding and value for it as a legitimate form of English. It is important to note that GAE, the form of English most often used in academic settings, is also considered a dialect. Oetting (2020) discusses how the profession of speech-language pathology can benefit from reframing perspectives around GAE, "a dialect of English that is never named but always implied," when discussing treatment goals.

AAE has been studied extensively in the context of language disorders and literacy achievement, largely focusing on children from low-income backgrounds. SES has been a consistent predictor of AAE dialect density, but this may be problematic because we know many of the language weaknesses associated with disadvantage can present similarly to linguistic rules of AAE. Lee-James and Johnson (2022) describe this overlap as the intersection of dialect, disadvantage, and disorder. It is especially important for SLPs to have a good understanding of the impacts all three may have on language abilities in the domains of phonology, morphology, and syntax as they make evidence-based decisions on eligibility for services.

The American Speech-Language-Hearing Association (ASHA, n.d.) defines cultural responsiveness as "understanding and appropriately including and responding to the combination of cultural variables and the full range of dimensions of diversity that an individual brings to interactions." One method of culturally responsive service

delivery is utilizing alternative forms of assessment and scoring practices, which may provide a more accurate description of the language abilities of children who use dialects others than GAE. Johnson and Gatlin-Nash (2020) describe dynamic assessment, language sampling, and processing-based tasks as alternate measurements that have been proven to be less-biased toward African American students and others from culturally and linguistically diverse (CLD) backgrounds. In addition to these culturally responsive alternate measurements, one other recommendation is to use standardized or norm-referenced assessments that accommodate language variation in their scoring. Scoring modifications include giving students credit for responses that would be considered appropriate based on the linguistic rules of their dialect or language system (Johnson & Gatlin-Nash, 2020).

Clinical Question

Ramon is concerned with how he can meet his district's requirements to provide a standardized score as a part of the reevaluation report, while also ensuring he utilizes culturally and linguistically responsive approaches after determining that Kamari is an AAE speaker. He needs to decide which assessments allow for scoring modification approaches and whether these approaches will help him better distinguish if Kamari truly needs continued special education services.

Ramon believes using the PICO framework (Straus et al., 2018) he learned about in graduate school could help guide his search for evidence-based information. The PICO framework is used to solve clinical problems and is defined as: P – the population, I – the intervention, C – the comparison group, and O – the outcome. Based on this, Ramon formulated the following question: Among school-age children who speak AAE (P), does the use of dialect scoring modifications (I), when compared to traditional scoring methods (C), yield more reliable identification of true language disorders (O)?

Search for the Evidence
Inclusion/Exclusion Criteria

Ramon created clear search terms and inclusion guidelines to find information that specifically answered his clinical question. Inclusion criteria for the articles included: (a) published in peer-reviewed journals, (b) participants were in

Volume 16, Article 4
September 2024

Examining the Utility of Scoring Modifications as a Culturally Responsive
Approach to Language Assessment in Speakers of African American English

elementary school in the United States (i.e., grades kindergarten through fifth), (c) evaluation of AAE dialect speakers, and (d) norm-referenced measures of expressive language. Ramon excluded articles examining scoring modifications for bilingual assessment and dialects other than AAE.

Search Strategy

Ramon began by using combinations of the following search terms to systematically locate relevant information via Google Scholar: "dialect" OR "AAE" OR "scoring modifications" AND "language." Ramon found over 6,020 results, but he quickly realized that he could not access some of the articles based on paywalls. Additionally, this search included journals with a focus on medical research, speech technology, education and information technology, and applied psycholinguistics. Upon review of the titles, Ramon found that many were not directly related to his question. He believed his next best option was to search journals focused specifically on issues related to communication disorders to find the most up-to-date and evidenced-based articles on his clinical question.

Ramon conducted a second search for articles using the terms, "scoring modifications for AAE dialect users" in the journals sponsored by ASHA using the ASHAWire database. As an ASHA member, Ramon was able to gain full access to the published articles using his membership login information. His search yielded 119 results, 83 of which were research articles. The other results were tutorials, review articles, and clinical forums. The majority of the results (*n* = 44) were found in the journal *Language, Speech, and Hearing Services in the Schools*. Ramon was pleased to see the search included a variety of topics that addressed cultural and linguistic diversity in speech-language pathology, as well as language disorders, development, and school-based settings. He began to eliminate articles that did not meet his inclusion criteria by reviewing the titles and abstracts. Eliminated articles included those that focused on the perception of SLPs and teachers, preschool or adult subjects, bilingualism, and dialects outside of AAE. Ultimately, Ramon found three articles that met his inclusion criteria and addressed his clinical question. These articles, along with a description of their purpose, methodology, and findings, can be found in Table 1.

Evaluating the Evidence

First, Ramon noted that there were not many articles to answer his PICO question directly. Several focused on standardized assessments, whereas others measured criterion-referenced or informal measures. Nonetheless, he reviewed three articles that met his inclusion criteria to learn more about utilizing scoring modifications for dialect speakers. For each article, he reviewed both the study quality and level of evidence. Ramon used the Wichita State University Evidence-Based Practice Reference Analysis Worksheet (Apel & Self, 2003) to assess study quality by determining the extent to which a study was designed and implemented appropriately. The Scottish Intercollegiate Guidelines Network (SIGN) Levels of Evidence system was used to examine study quality based on the design's ability to protect against bias (Scottish Intercollegiate Guidelines Network, 2019). He recalled that randomized control trials and meta-analyses had the strongest level of evidence, whereas expert opinion had the weakest evidence to protect against biases.

Ramon began by first reviewing an article by Hendricks and Adlof (2017) based on its target population of elementary students and the focus on evaluating language assessment with children who speak nonmainstream dialects. The article explored the impact of scoring modifications on norm-referenced measures to gain a more accurate depiction of these children's language skills. Ramon was intrigued by the authors' evaluation of the impact of modified scoring procedures and how they would impact diagnosis on the Clinical Evaluation of Language Fundamentals™ (4th ed.; CELF®-4; Semel et al., 2003). The target population included a diverse group of second grade students in South Carolina. Ramon felt these were important factors to consider because the demographic information was similar to Kamari's. He also noticed that the study used the DELV-ST, and he previously confirmed that Kamari demonstrated strong variation from GAE when he was administered the screener. The study concluded that there was a small to moderate correlation between language ability measures and nonmainstream dialect use. Ramon noted that when scoring modifications were used, AAE speakers obtained higher scores on the CELF-4. However, these modifications affected the diagnostic accuracy of the CELF-4 because there was a lower false positive rate (i.e., classifying children who speak AAE as having developmental language disorder in error) and a higher false negative rate (i.e., classifying children who speak AAE as having typical performance,

Volume 16, Article 4
September 2024

Examining the Utility of Scoring Modifications as a Culturally Responsive
Approach to Language Assessment in Speakers of African American English

when they actually have developmental language disorder). The authors cautioned that clinicians must account for the possibility of false positives or false negatives when administering scoring modifications to ensure students are not being over- or under-identified. As a clinician, Ramon wants to help Kamari succeed in school but he realizes that he also must consider the impact of a misdiagnosis or lack of diagnosis.

Ramon noted several citations in the Hendricks and Adlof (2017) study that he initially thought would help to inform his decisions, but ultimately decided to forgo reviewing the entire article after reading each abstract. For example, an article from Rhyner and colleagues (1999) was removed from his reading list because it focused on using two language screeners with low-income African American kindergartners. Kamari's eligibility for language therapy had previously been identified, so he knew he would not administer another screener. Additionally, he did not consider Terry, Jackson, and colleagues' (2010) paper on the effects of AAE on the Recalling Sentences subtest of the Clinical Evaluation of Language Fundamentals (3rd ed.; CELF-3; Semel et al.,1995) because the measure is currently included in the fifth edition and he worried about how reflective the norming sample was for students like Kamari.

The next article Ramon reviewed was from Oetting and colleagues (2016) which explored performance on a sentence recall task in students who spoke AAE and Southern White English (SWE). Although studies have shown that the ability to recall sentences can be a clinical marker for language impairment, more information is needed to see if these results are accurate for dialect speakers. Ramon recalled that there is overlap between certain grammatical features not being used or used variably in GAE that would cause a red flag for language impairment but are perfectly acceptable in dialects such as AAE (e.g., past tense *-ed*, plural *-s*, copula). He hoped this article would provide some clarity on whether including a sentence recall task would be appropriate for Kamari. Oetting and colleagues used a scoring modification procedure called dialect-strategic scoring, which allowed students to receive credit for dialect-appropriate productions of the following features that were frequently used in both dialects but have not been shown to be a clinical marker of language

impairment: *is* for third person plural present progressive verbs, *was* for third person plural past progressive verbs, and zero marking of verbal *-s*. The authors found that students who were already identified as having specific language impairment[1] (SLI) achieved lower scores on the measure than the comparison group, even when accounting for dialect usage. Their dialect-strategic scoring system provided moderate to high levels of diagnostic accuracy using the sentence recall task within both AAE and SWE dialect groups to identify students with SLI (Oetting et al., 2016). Ramon was glad to see that a sentence recall task would be a culturally responsive way to assess students, like Kamari, who use linguistic variations.

The final article Ramon elected to review by Oetting and colleagues (2021) further explored the use of the dialect-strategic scoring approach in marking tense and agreement in language samples gathered from AAE- and SWE-speaking students with and without SLI. Focusing specifically on the overlapping tense and agreement grammatical features of past tense, verbal *-s*, auxiliary *BE* present, and auxiliary *BE* past, three scoring approaches were used to analyze the play-based language samples. The unmodified approach involved using a traditional scoring method where only GAE forms were considered correct. The modified approach allowed for forms that would be considered correct in GAE or dialect (e.g., ate, ated, had ate, had eat), with any other responses counted as incorrect. Finally, the strategic approach considered both GAE and dialect overt (grammatical forms that are marked, such as the double negative in "He ain't have no school") and zero forms (grammatical forms that are not used or used variably, such as the copula in "She __ going home") as correct but divided them using a ratio of the sum of both overt forms by the sum of the overt and zero forms. The results of this study showed that the strategic scoring approach exhibited the greater differences between the SLI and typically developing groups, accounting for effects of dialect usage. Additionally, using dialect specific cut scores on the measures of past tense marking led to higher diagnostic accuracy. Ramon realized he could use this article to inform his analyses and decision-making after eliciting Kamari's language sample.

---

[1] Although the consensus is to use the term developmental language disorder (DLD; Bishop et al., 2017) to describe language disorders not associated with a known biomedical etiology, the term specific language impairment (SLI) is used as described in the original articles cited.

Volume 16, Article 4
September 2024

Examining the Utility of Scoring Modifications as a Culturally Responsive
Approach to Language Assessment in Speakers of African American English

The Evidence-Based Decision

Each of the articles Ramon reviewed reinforced the idea that multiple factors need to be considered when evaluating the language skills of children from CLD backgrounds. He initially thought he would be able to find sufficient evidence for using scoring modifications on standardized measures, but quickly realized that more research is needed in this area. The studies Ramon reviewed confirmed that he was using culturally responsive approaches by first administering the DELV-ST to determine linguistic variation status and his plan to follow up with alternate measures, such as language sampling, that are considered less biased towards CLD populations. Language samples can be used in more naturalistic settings, are elicited in a variety of ways, are not influenced by prior experiences and background knowledge, and provide in-depth information about all aspects of the child's language skills (Lee-James & Johnson, 2022). Based on the criteria outlined by Oetting and colleagues (2021), Ramon has learned a new approach to coding and analyzing language samples and looks forward to putting this new knowledge into application.

Currently, there are limited standardized tests that offer detailed scoring modifications that are appropriate for accurately diagnosing language impairments in children who use a variety of dialects. As more of these measures begin to be validated in a way that authentically and reliably represent the true abilities of students like Kamari, who use various linguistic forms, Ramon and other SLPs will be able to make more informed, evidence-based decisions. In the meantime, Ramon wants to move forward with administering the CELF-5 (Wiig et al., 2013) to Kamari because it does provide the standard score his district requires for eligibility, but also includes an appendix with additional information on scoring modifications.

Because responses on the Word Structure, Formulated Sentences, and Recalling Sentences subtests may exhibit linguistic variations, Appendix I on Language Differences and Cultural Sensitivity in the *CELF-5 Manual* provides detailed instructions on acceptable alternate responses. Common phonological and morphsyntactical contrasts between GAE and AAE, Spanish-influenced English, Chinese-influenced English, and Southern English are provided. After confirming the student is a dialect speaker and familiarizing oneself with the features of the dialect, the CELF-5 recommends giving full credit for responses that are appropriate based on the student's linguistic background. In this case, Kamari's responses would be written down verbatim. For example, on the Word Structure subtest, one of the items is "Tell me, who is sleepy?" If Kamari responded, "They is," he would receive full credit because Ramon knows variable use of subject-verb agreement is characteristic of AAE, and "they is" is listed as an alternate response.

It will be extremely important to consider all the factors around Kamari's present levels of performance, including talking with his family and teachers, completing classroom observations, and reviewing academic performance. Results from Ramon's clinical question confirmed his original thought that he should use both standardized and informal measures to make the most informed decision. Considering all these factors together as a whole will help Ramon make the best decision about whether continued services are needed after completing the reevaluation.

Authors' Note

Correspondence concerning this article should be addressed to Lakeisha Johnson, School of Communication Science and Disorders, Florida State University, 201 W Bloxham St, Tallahassee, FL 32301. Email: lrjohnson@fsu.edu

References

American Speech-Language-Hearing Association. (n.d.). *Cultural responsiveness* [Practice portal]. https://www.asha.org/Practice-Portal/Professional-Issues/Cultural-Responsiveness/

Apel, K., & Self, T. (2003). Evidence-based practice: The marriage of research and clinical services. *The ASHA Leader, 8*(16), 6–7. https://doi.org/10.1044/leader.FTR1.08162003.6

Volume 16, Article 4
September 2024

Examining the Utility of Scoring Modifications as a Culturally Responsive
Approach to Language Assessment in Speakers of African American English

Berry, J. R., & Oetting, J. B. (2017). Dialect variation of copula and auxiliary verb BE: African American English–speaking children with and without Gullah/Geechee heritage. *Journal of Speech, Language, and Hearing Research, 60*(9), 2557–2568. https://doi.org/10.1044/2017_JSLHR-L-16-0120

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE-2 Consortium. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *The Journal of Child Psychology and Psychiatry, 58*(10), 1068–1080. https://doi.org/10.1111/jcpp.12721

Craig, H. K., & Washington, J. A. (2004). Grade-related changes in the production of African American English. *Journal of Speech, Language, and Hearing Research, 47*(2), 450–463. https://doi.org/10.1044/1092-4388(2004/036)

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test* (4th ed.). Wascana Limited Partnership.

Ehrler, D. J., & McGhee, R. L. (2008). *Primary Test of Nonverbal Intelligence*. PRO-ED.

Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation* (2nd ed.). NCS Pearson.

Gutiérrez-Clellen, V. F., Simon-Cereijido, G., & Wagner, C. (2008). Bilingual children with language impairment: A comparison with monolinguals and second language learners. *Applied Psycholinguistics, 29*(1), 3–19. https://doi.org/10.1017/S0142716408080016

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

Johnson, L., & Gatlin-Nash, B. (2020, Spring). Evidence-based practices in the assessment and intervention of language-based reading difficulties among African American learners. *Perspectives on Language and Literacy, 46*(2), 19–23.

Lee-James, R., & Johnson, L. (2022). Impact of nonmainstream dialect use in language assessment with young African American English-speaking children. In E. Saiegh-Haddad, L. Laks, & C. McBride (Eds.), *Handbook of literacy in diglossia and dialectal contexts: Psycholinguistic, neurolinguistic, and educational perspectives* (pp. 425–449). Springer. https://doi.org/10.1007/978-3-030-80072-7_18

Lippi-Green, R. (2004). Language ideology and language prejudice. In E. Finegan & J. R. Rickford (Eds.), *Language in the USA: Themes for the twenty-first century* (pp. 289–304). Cambridge University Press. https://doi.org/10.1017/CBO9780511809880.017

Newcomer, P. L., & Hammill, D. D. (2008). *Test of Language Development–Primary* (4th ed.). PRO-ED.

Oetting, J. B. (2020, November 16). *From my perspective/opinion: General American English as a dialect: A call for change.* The ASHA LeaderLive. https://leader.pubs.asha.org/do/10.1044/leader.FMP.25112020.12

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., McDonald, J. L., Seidel, C. M., & Hegarty, M. (2016). Sentence recall by children with SLI across two nonmainstream dialects of English. *Journal of Speech, Language, and Hearing Research, 59*(1), 183–194. https://doi.org/10.1044/2015_JSLHR-L-15-0036

Oetting, J. B., Rivière, A. M., Berry, J. R., Gregory, K. D., Villa, T. M., & McDonald, J. (2021). Marking of tense and agreement in language samples by children with and without specific language impairment in African American English and Southern White English: Evaluation of scoring approaches and cut scores across structures. *Journal of Speech, Language, and Hearing Research, 64*(2), 491–509. https://doi.org/10.1044/2020_JSLHR-20-00243

Rhyner, P. M., Kelly, D. J., Brantley, A. L., & Krueger, D. M. (1999). Screening low-income African American children using the BLT-2S and the SPELT-P. *American Journal of Speech-Language Pathology, 8*(1), 44–52. https://doi.org/10.1044/1058-0360.0801.44

Volume 16, Article 4
September 2024

Examining the Utility of Scoring Modifications as a Culturally Responsive
Approach to Language Assessment in Speakers of African American English

Scottish Intercollegiate Guidelines Network. (2019). *SIGN 50: A guideline developer's handbook* (Rev. ed.). https://www.sign.ac.uk/media/2038/sign50_2019.pdf

Semel, E., Wiig, E. H., & Secord, W. A. (1995). *Clinical Evaluation of Language Fundamentals* (3rd ed.). NCS Pearson.

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals* (4th ed.). NCS Pearson.

Seymour, H. N., Bland-Stewart, L., & Green, L. J. (1998). Difference versus deficit in child African American English. *Language, Speech, and Hearing Services in Schools, 29*(2), 96–108. https://doi.org/10.1044/0161-1461.2902.96

Seymour, H. N., Roeper, T. W., & de Villiers, J. (2003). *Diagnostic Evaluation of Language Variation – Screening Test*. Ventris Learning.

Seymour, H. N., Roeper, T. W., & de Villiers, J. (2005). *Diagnostic Evaluation of Language Variation – Norm Referenced*. Ventris Learning.

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

Terry, J. M., Jackson, S. C., Evangelou, E., & Smith, R. L. (2010). Expressive and receptive language effects of African American English on a sentence imitation task. *Topics in Language Disorders, 30*(2), 119–134. https://doi.org/10.1097/TLD.0b013e3181e04148

Terry, N. P., Connor, C. M., Thomas-Tate, S., & Love, M. (2010). Examining relationships among dialect variation, literacy skills, and school context in first grade. *Journal of Speech, Language, and Hearing Research, 53*(1), 126–145. https://doi.org/10.1044/1092-4388(2009/08-0058)

Thompson, C. A., Craig, H. K., & Washington, J. A. (2004). Variable production of African American English across oracy and literacy contexts. *Language, Speech, and Hearing Services in Schools, 3*(3), 269–282. https://doi.org/10.1044/0161-1461(2004/025)

Washington, J. A., & Craig, H. K. (1998). Socioeconomic status and gender influences on children's dialectal variations. *Journal of Speech, Language, and Hearing Research, 41*(3), 618–626. https://doi.org/10.1044/jslhr.4103.618

Wiig, E. H., Semel, E., & Secord, W. A. (2013). *Clinical Evaluation of Language Fundamentals* (5th ed.). NCS Pearson.

Volume 16, Article 4
September 2024

Examining the Utility of Scoring Modifications as a Culturally Responsive
Approach to Language Assessment in Speakers of African American English

Table 1. Summary of Articles That Met Ramon's Inclusion Criteria

Article & level of evidence	Participants	Purpose	Type of language testing	Findings	Clinical implications	Limitations
Hendricks & Adlof (2017) Level of evidence – 2	299 second-grade students (167 White, 106 African American, and 26 other)	To determine how modified scoring procedures affected the diagnostic accuracy of the CELF-4	DELV-ST and the CELF-4 The CELF-4 was scored with and without the recommended scoring modifications for children who spoke AAE.	Partial correlations controlling for SES revealed small to moderate correlations between measures of language ability and the use of NMAE features. Modified scoring yielded higher scores for children who spoke AAE and a reduced association between the use of NMAE features and CELF-4 scores was found. Modified scoring also affected the diagnostic accuracy of the CELF-4, resulting in a lower positive likelihood ratio and a higher negative likelihood ratio.	The risks of false positives (erroneously classifying children who speak NMAE dialects as having LI) and false negatives (classifying children who speak NMAE dialects as having typical language when they do have LI) must be considered to ensure over- or under-identification does not occur. Decisions regarding eligibility should not be limited to scores on standardized assessments only.	Scoring modification can be challenging for some SLPs to grasp. A small number of AAE speakers were included in the sample, especially those at highest risk for LI.
Oetting et al. (2016) Level of evidence – 2	106 children (70 AAE speakers and 36 SWE speakers)	To determine whether the ability to recall sentences serves as a clinical marker for SLI for children who speak AAE and SWE	DELV-ST to verify dialect status DELV-NR, PTONI, GFTA™-2, PPVT™-4, and TOLD:P-4 to determine clinical status Researcher-created sentence recall task	For both dialects, the SLI group earned lower sentence recall scores than the TD group. Children with SLI, as compared with TD controls, exhibited lower levels of verbatim recall; more ungrammatical recalls when the recall was not exact; and higher levels of error on targeted functional categories, especially those marking tense.	Dialect-strategic scoring offers an alternative to traditional scoring systems by ensuring children are not penalized for producing dialect-appropriate features that overlap with clinical markers of SLI.	Current findings need to be replicated with larger groups of children. AAE varies regionally, so differences may not be generalizable.

Volume 16, Article 4
September 2024

Examining the Utility of Scoring Modifications as a Culturally Responsive
Approach to Language Assessment in Speakers of African American English

Table 1. Summary of Articles That Met Ramon's Inclusion Criteria (continued)

Article & level of evidence	Participants	Purpose	Type of language testing	Findings	Clinical implications	Limitations
Oetting et al. (2021) Level of evidence – 2	106 kindergartners (70 AAE speakers and 36 SWE speakers)	Following up the Oetting et al. (2016) study, authors examined the marking of tense and agreement (T/A) in language samples by children with and without SLI in AAE- and SWE-speakers. Additionally, the clinical utility of different scoring approaches and cut scores to account for dialect were explored.	DELV-ST to verify dialect status DELV-NR, PTONI, GFTA-2, PPVT-4, and TOLD:P-4 to determine clinical status Play-based language sample	A large, diverse inventory of mainstream and nonmainstream T/A forms were produced within the samples. Of the three scoring approaches, strategic scoring led to the greatest differences between the clinical groups while reducing effects of the children's dialects. Dialect-specific cut scores resulted in better clinical classification accuracies with measures of past tense leading to the highest levels of classification accuracy.	T/A structures should be assessed in AAE- and SWE-speaking children with SLI using dialect-appropriate materials, coding, and scoring. Possible goals for AAE- and SWE-speaking children with SLI include targeting T/A forms and increasing children's use of various dialect-appropriate T/A forms to match the frequency distributions of their dialect community.	Exercise caution when applying the cut scores identified with other groups of children because more research is needed. SLPs may need targeted training to understand the three approaches and how to best apply them to language sample analysis.