

Dialect and Mathematics Performance in African American Children Who Use AAE: Insights from Explanatory IRT and Error Analysis

Katherine T. Rhodes, Julie A. Washington & Sibylla Leon Guerrero

To cite this article: Katherine T. Rhodes, Julie A. Washington & Sibylla Leon Guerrero (31 Jul 2024): Dialect and Mathematics Performance in African American Children Who Use AAE: Insights from Explanatory IRT and Error Analysis, Educational Assessment, DOI: [10.1080/10627197.2024.2370787](https://doi.org/10.1080/10627197.2024.2370787)

To link to this article: <https://doi.org/10.1080/10627197.2024.2370787>



© 2024 The Author(s). Published with license by Taylor & Francis Group, LLC.



[View supplementary material](#)



Published online: 31 Jul 2024.



[Submit your article to this journal](#)



Article views: 22



[View related articles](#)



[View Crossmark data](#)

Dialect and Mathematics Performance in African American Children Who Use AAE: Insights from Explanatory IRT and Error Analysis

Katherine T. Rhodes, Julie A. Washington, and Sibylla Leon Guerrero



University of California


ABSTRACT

Little is known about mismatches between the language of mathematics testing instruments and the rich linguistic repertoires that African American children develop at home and in the community. The current study aims to provide a proof of concept and novel explanatory item response design that uses error analysis to investigate the relationship between AAE child language and children's mathematics assessment outcomes. Here, we illustrate 2nd and 3rd grade children's qualitative patterns of performance on arithmetic tasks in relation to their AAE dialect use and elaborate a unified framework for examining child and item level linguistic characteristics. Results suggest that children draw upon their emerging (bi)dialectal repertoire with arithmetic problems when selecting appropriate problem-solving strategies on language-formatted problems. The mismatch of assessment language formatting with children's repertoires may disadvantage AAE speakers' strategy selections and result in a language-based performance disadvantage unrelated to mathematical ability.

It is often assumed that the language of math word problems provides clear linguistic cues to forming mental representations of the problem. Any difficulties children encounter during problem-solving are generally interpreted as an inability to correctly apply *math* concepts and procedures toward a solution. However, this interpretation is not necessarily warranted when children's linguistic repertoires diverge from the language used in math word problems. While there is some divergence between home language and the language of math assessment for all children, this divergence is magnified for those children who speak minoritized language varieties at home, such as a language other than English or a dialect other than general American English. In the case of bilingual children who are identified as English learners, linguistic barriers to accessing word problems have long been recognized as important to consider for testing accommodations (Abedi & Lord, 2001; Clinton et al., 2018). While a growing body of research has investigated this issue in children who speak a language other than English at home, the status of children who are bidialectal speakers of a variety of English such as African American English (AAE), often goes unrecognized and overlooked, both in research and in practice. The current study aims to address this gap in knowledge by exploring the role of AAE-speaking children's dialectal repertoires in mathematics assessment.

Similar to bilingual English speakers, bidialectal AAE speakers may experience language-related barriers accessing the problem-solving assessments they encounter at school, i.e. difficulties comprehending, encoding, and forming mental representations of what mathematics assessment problems expressed in a dialect that does not match the language of home and community are asking them to do (Terry, Hendrick, et al., 2010). Representational challenges may arise from linguistic complexity or linguistic difference or both. Item-level linguistic complexity is often introduced by the use of

CONTACT Katherine T. Rhodes  ktrhodes@uci.edu  School of Education, University of California, 3200 Education Building, Irvine, CA 92697, USA

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/10627197.2024.2370787>

© 2024 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

specialized or infrequent vocabulary words, complex syntax, and/or unfamiliar topics outside of children's experience or background knowledge (Banks et al., 2016). At the individual level, differences between AAE child language and the language of the assessment can introduce barriers to the encoding of oral language (Terry, Jackson, et al., 2010), reading comprehension (Craig & Washington, 2006; Gatlin & Wanzek, 2015; Washington et al., 2018), and mathematics problem solving (Terry et al., 2022; Terry, Hendrick, et al., 2010).

While child-level and item-level characteristics of mathematics assessment have been examined separately, to date there has been no research examining the interaction of child and item language within one, cohesive, analytic framework (Cruz Neri & Retelsdorf, 2022). Explanatory item response theory (EIRT), when used in conjunction with error analysis, can offer assessment users insights into the constellations of the item features and child characteristics that may contribute to language-based patterns of performance. In this paper, we explore the interaction of children's linguistic repertoires with the language of math assessment among African American children who use AAE, and illustrate how EIRT, combined with a careful error analysis, is distinctly qualified to reveal hidden sources of math performance variation among this understudied population of children.

Literature review

The language factor in mathematics assessment

The impact of language formatting on test performance is well-established for children who use a language other than English (i.e., bilingual English learners or ELLs) and children who have language disabilities (Abedi, 2004; Abedi et al., 1997; Martiniello, 2008; Rhodes et al., 2015). For bilingual children, item linguistic bias has been associated with semantic elements such as idioms and low frequency words that are not specific to math content as well as complex syntax such as subordinating clauses, conditional clauses, and the passive voice (Abedi, 2006). This impact of language formatting on mathematics assessments has been demonstrated from early elementary ages (e.g., Hopewell & Escamilla, 2014) to upper elementary and high school (Liu & Bradley, 2021; Martiniello, 2008; Shaftel et al., 2006), though the presence of assessment bias and whether particular linguistic features introduce bias for ELLs in standardized mathematics assessments varies by assessment, grade, and the tested population (Clinton et al., 2018; Cruz Neri & Retelsdorf, 2022). Among students identified as bilingual English learners, testing accommodations such as extra time, read-alouds, and access to dictionaries have thus become standard practice during mathematics instruction and assessment (Abedi, 2004; Kieffer et al., 2009).

Bidialectalism and African American English dialect

African Americans comprise approximately 14% of the U.S. population and are the second largest demographic minority in the U.S (Lee-James & Washington, 2018; U.S. Census Bureau, 2021). Many African American children grow up speaking a dialect, AAE, with distinctive linguistic features in all domains of language (i.e., syntax, morphology, phonology, semantics and pragmatics; Lee-James & Washington, 2018). Just as with bilinguals, AAE bidialectal speakers develop specialized language knowledge and processing strategies in response to their rich, linguistic environments (Beyer et al., 2015; Garcia et al., 2022). However, AAE is not represented in the classrooms of bidialectal African American children, or on academic assessments, including mathematics tests. Moreover, in educational practice and policy, African American children are generally identified as racial, rather than linguistic minorities. As a result, their (bi)dialectal experiences are not measured or considered as potentially meaningful facets of their cognitive profiles, and very little is known about how similarities and differences between AAE and the "general" American English (GAE) of the school curriculum may interact in language development at home and school.

Differences, or mismatches, between language forms in assessments and the language forms used in home and community appear to make the task of comprehending oral and written assessments more

demanding for AAE speakers, thus requiring more cognitive resources such as working memory (Brown et al., 2015; Jacobson et al., 2017; Terry, Hendrick, et al., 2010). From this perspective, when assessments contain more contrastive or mismatching features, children who use dense AAE dialect are thought to incur a greater cost in terms of cognitive load (as compared to less dense dialect speakers) because they must reconcile the differences among linguistic forms during processing. This theory has found support in several literacy studies, where greater dialect mismatch has been associated with lower levels of literacy performance (Gatlin & Wanzek, 2015) and increased working memory load (Jacobson et al., 2017).

Though language is integral in mathematics learning (Barwell, 2003; Moschkovich, 2010), most research on AAE dialect usage has focused on children's language and literacy development (e.g., Connor & Craig, 2006; Craig et al., 2004; Patton Terry, 2006, 2008; Puranik et al., 2020; Stockman, 2010; Washington et al., 2018) while very few studies have examined the relationship between AAE language usage and children's mathematical problem-solving. This gap in AAE literature is critical, because similar to literacy, mathematical problem-solving is often verbally-mediated, particularly when "real world" mathematics problems are represented with language formatting (Cruz Neri & Retelsdorf, 2022; Fedorenko et al., 2007).

To date, two studies (Terry et al., 2022; Terry, Hendrick, et al., 2010) have linked AAE dialect usage with performance on verbally-mediated mathematics problem-solving. Terry, Hendrick, et al. (2010) found that assessment items containing contrastive morphosyntactic features that are variably included in AAE (third person singular "s" morpheme (e.g., Jill eats), or third person singular conditionals with the "ed" past tense marker (if Jill walked to school . . .)) were particularly impactful on student's probabilities of correctly answering items. Terry et al. (2022) more specifically focused on the presence of 3rd person singular verbal "s" in math word problems and found that children performed better on AAE-consistent math word problems (i.e., with omitted verbal -s). Concomitantly, their online processing of GAE-consistent sentences containing verbal -s displayed EEG signatures of increased cognitive effort needed to integrate the -s feature into a mental representation of the sentence. Thus, at the item level, mismatches between morphological features that vary across AAE and GAE may bias mathematics assessment outcomes while at the child level, processing of these features may introduce cognitive load unrelated to the mathematical task.

A novel methodological approach for revealing issues of linguistic mismatch and match

While both item and child language characteristics can thus play a role in performance, the exact mechanisms of language mismatches in mathematics assessments remain unclear because, (1) when performance is measured by correct/incorrect response patterns on language formatted items, it is not possible to describe qualitative differences in performance that may be directly attributable to difficulties with mentally representing language-formatted problems (as opposed to mathematical difficulties that may also be apparent on problems formatted with only Arabic numerals), and (2) children's linguistic profiles must be cohesively related to item-level performance in order to estimate the full magnitude of language effects.

Error analysis and differential patterns of performance

We argue that the first issue can be addressed with an error analysis, which may be especially important for drawing conclusions about the nature of difficulties children may encounter as a result of linguistic mismatches with assessments. Beyond a dichotomous accuracy score, children's strategy use and types of errors contain a wealth of information about their mathematical cognition and sources of misunderstanding (Mazzocco et al., 2013; Ryan & Williams, 2007; Siegler & Shrager, 1984). These more qualitative metrics of children's mathematical problem-solving can yield important insights into underlying misconceptions, which often involve skills that are malleable for intervention.

In particular, error analysis is a powerful tool for gaining insight into cognitive mathematical processing (Ketterlin-Geller & Yovanoff, 2009; Ohlsson, 1996; Tatsuoka, 1983, 1985, 1990). While there are many forms of error analysis (e.g., Engelhardt, 1977; Greenstein & Strain, 1977; Roberts,

1968), for the purposes of understanding children's difficulties forming mental representations of problems as a result of language mismatch, it is useful to generate error patterns that are based on information processing (see Radatz, 1979), distinguishing between difficulties with problem approach and difficulties with solution execution. Difficulties with problem approach, including selecting a strategy that could feasibly result in a correct solution, would be indicative that a child has had difficulty forming a mental representation of what she is being asked to do (e.g., performing a subtraction operation on an addition problem). Difficulties with the execution of the solution strategy, including difficulties that occur during computation, would be indicative that a child has had difficulty with mathematical processing (e.g., counting one finger twice while using finger counting to solve an addition problem).

EIRT and cohesively examining both item- and child-level effects

While error analysis describes children's mathematical cognition, research has consistently demonstrated that children's performance on math test items depends on both children's abilities *and* the features of those items, including larger problem sizes, more difficult mathematical operations, or language formatting (as opposed to Arabic numeral formatting), all of which tend to be more difficult for all test-takers regardless of abilities (Campbell & Epp, 2005; Geary & Wiley, 1991; Imbo & Vandierendonck, 2007; LeFevre et al., 1996; Siegler, 1991; Siegler & Shrager, 1984; Siegler & Taraban, 1986). Explanatory item response theory is a collection of statistical modeling methods that allow both item- and child-level effects to be considered in one, cohesive modeling framework (De Boeck & Wilson, 2004; Van Den Noortgate et al., 2003). While EIRT is a broad family of models that includes many different model types, in the present study, we employ the linear logistic test model (LLTM; De Boeck & Wilson, 2004) because it is the most parsimonious model for relatively small datasets. The LLTM is an EIRT model that estimates the effects of item properties rather than estimating individual item parameters. Importantly for the current study, one would not expect to find item \times child interaction effects (i.e., formatting effects) unless a test is explicitly designed to elicit them. The unintentional interaction between African American children who use AAE and mathematics test items with linguistic formatting may not be apparent in correct responses – it may instead be evident in the nature of incorrect responses. Thus, observing the effect of interest would require both the examination of qualitatively different patterns of among incorrect answers (i.e., an error analysis), and the explicit examination of potential item \times child interactions using an EIRT modeling approach.

The current study

The current study extends the small body of literature examining AAE and mathematics word problem difficulties by offering proof of concept for a cohesive, explanatory item response theory (EIRT) paradigm examining item-formatting and person-language interaction effects through error analysis. In this exploratory study, we examined African American children's errors and strategies while solving math problems with various problem sizes, mathematical operations, and symbolic formatting. In particular, the study investigated the possibility that African American children may evidence patterns of AAE mismatch for forming mental representations with language-formatted arithmetic, and that this relationship may only be visible with explicit examination of children's qualitative patterns of performance (i.e., that they would be more likely to make strategic errors on language formatted items as a function of their AAE-consistent language productions). Therefore, we asked whether patterns of performance, in both error propagation and correct/incorrect responses, could be predicted by 1) item formatting (language vs. Arabic numeral); 2) children's AAE language usage; and 3) an interaction between item formatting and children's AAE language use.

In line with extensive research on word problem difficulty, we expected that language- as compared to numerical-, formatting would have a significant negative impact on children's math performance. We hypothesized that in our sample of highly dense AAE speakers, assessment mismatches with

children's AAE would present a challenge for children's ability to construct a mental representation of language-formatted items and derive a correct answer. We expected to see a negative relationship of AAE production with children's ability to select correct strategies to solve the problem (regardless of whether they perform the computation correctly), indicating that language formatting in the test creates difficulties for children's mental representations of the problems. We hypothesized that language mismatch would be more likely to affect language-formatted items, resulting in an interaction between language formatting and AAE production such that children with higher AAE scores would be more likely to select an incorrect strategy in language formatted items.

Method

Participants

Participants ($n = 42$, 25 male) were from a large metropolis in the Southeastern United States and drawn from the baseline (pre-intervention) time point of a larger study that aimed to evaluate language-based reading interventions for African American children. The students were 2nd and 3rd graders who, on average, were approximately one grade level behind in reading. Their oral reading fluency was slightly below average for similarly aged peers (Gray Oral Reading Test Oral Reading Index standard score $M = 79.71$, $SD = 10.02$), but they were not so far below average that they would necessarily have qualified for in-school services. This pattern of reading performance is often seen among African American children who are dense dialect speakers (Patton Terry, 2006; Terry, Jackson, et al., 2010; Washington et al., 2018). Students who did not have significant visual and/or hearing impairments and who had no reported intellectual disabilities were approached for participation in the study. Parental/guardian consent and child assent were obtained prior to testing. The sample had a mean age of 9 years, 2 months ($SD = 11.89$ months; range = 7 years, 6 months to 11 years, 2 months). Five of the participants had current IEPs for speech-language disorders, learning disability, and/or emotional behavioral disorders. The participants all attended Title I schools.

Measures

After children assented to participate, the measurement battery was administered by trained research assistants, one-on-one, in quiet areas of participating research sites. All data were double entered and compared for consistency.

Experimental math assessment

The math assessment contained eight items in a $2 \times 2 \times 2$ design, across symbolic format (word problems vs. Arabic numeral format), operation (addition vs. subtraction), and problem size (small, single digit operands vs. larger, mixed digit or double digit operands).¹ The experimental items are provided in Appendix A. Item 2, "8 + 8," for example, was designed as an Arabic numeral formatted addition problem with small operands, while item 4, "Melissa has four pieces of bacon. Her dog takes away three pieces. How many pieces of bacon does Melissa have left," was designed as a word, subtraction problem with small operands.

The experimental math assessment was administered with paper and pencil. Children were told to take as much time as they needed to complete each item and encouraged to show as much work as possible. Children were allowed to self-correct and told that they could change their answers if they desired, such that commission of errors and correct responses were not mutually exclusive. In order to avoid confounds for reading ability, word problems were read aloud by examiners and repeated as many times as children requested, which is a common accommodation for mathematics word

¹During the design phase, children were confused by written representations of larger operands (e.g., seventy-six). Thus, operands greater than ten were represented with Arabic numerals within the word problems. Similarly, Arabic numeral formatted items were represented in columns rather than as horizontal number sentences.

problem administration with students who are English language learners and one that does not appear to provide undue test performance advantages (Wolf et al., 2012). During administration, examiners took observational notes about children's problem-solving strategies for each item, noting verbalizations, gestures, and writing. Children's strategies, errors, and accuracies were coded using examiners' behavioral notes and children's written protocols into one of four categories: counting, fact retrieval, decomposition, and/or algorithm execution.

Outcome coding. After administration, responses were coded across three outcomes, (1) a strategic error, (2) a computational error, and (3) correct/incorrect answers. These outcomes were not mutually exclusive, such that it was possible for children to commit multiple types of errors on the same item and/or commit error(s) and then self-correct to achieve a correct solution. Outcome codes are described along with example problem-solving observations in Table 1.

Reliability. Two trained research assistants independently conducted coding of strategy usage, error commission, and accuracy for all responses. Reliability analyses indicated a 98% agreement in coding with a random sub-sample of 20% of participants. All score discrepancies were discussed until score agreement could be reached. Disagreements were resolved through discussion. If still unresolved, the Principle Investigator made the final scoring decisions.

Standardized assessments

Broad math ability. The Woodcock-Johnson III, Applied Problems subtest (McGrew & Woodcock, 2001) was used to assess broad mathematical ability. Children were presented a series of story problems visually and read aloud that required them to use a variety of math operations in order to

Table 1. Outcome response frequencies and example errors.

Possible Response Types & Example Responses	Frequency of Response
Incorrect, no errors (incorrect other)¹ <i>A child said, "I don't understand," and chose to skip the problem entirely.</i>	11
Incorrect, computation error <i>A child attempts to solve the problem "25-19" using the subtraction algorithm, has difficulty borrowing from the 10s column, and answers "10."</i>	73
Incorrect, strategy error <i>A child attempts to solve an addition problem using subtraction, or a child attempts to recall the answer to the problem "3 + 4," incorrectly recalls the math fact, does not select a back-up strategy, and reports the answer, "6."</i>	42
Incorrect, both errors <i>A child uses addition to solve the problem "25-19," has difficulty carrying from the 1s column, and reports the answer "54."</i>	10
Correct, no errors <i>A child provides a correct answer without making an error.</i>	199
Correct, strategy error self-corrected <i>A child attempts to recall the answer to the problem "3 + 4," incorrectly recalls the math fact as "6," then pauses and says, "No. That's seven."</i>	1
Correct, computation error self-corrected <i>A child uses finger counting to solve the problem "3 + 4," misses a finger, and answers, "6," then checks her work and corrects her answer to, "7."</i>	0
Correct, both errors self-corrected <i>A child uses addition to solve the problem "25-19," has difficulty carrying from the 1s column, and reports the answer "54." The child then decides to check his work and is able to use subtraction to answer correctly, "6."</i>	0

¹For instance, "incorrect, no errors" means that the student solved the problem incorrectly but did not use an inappropriate strategy or make a common computational error (e.g., a child responded with, "I don't understand," or chose to skip the problem entirely)..

Reliability analyses indicated a 97.74% agreement with a random sub-sample of 20% of participants. Total raw scores from the experimental math task correlated significantly with age-referenced standard scores on the Woodcock-Johnson III, Applied Problems subtest (McGrew & Woodcock, 2001), suggesting convergent validity with a standardized test of mathematical achievement, $r = .53, p < .001$.

solve questions of increasing difficulty. The published median reliability of this task is .92 in the 5- to 19-year-old age range. Standard scores were used for the current analysis.

AAE production. The Diagnostic Evaluation of Language Variation Screener (DELV-S, Seymour et al., 2003), Language Variation subtest was used to assess children's AAE dialect usage. The Language Variation subtest classifies children's phonological and morphosyntactic productions according to their consistency with and variation from general American English dialect (Seymour et al., 2003). Children were presented with a series of repetition and cloze items, and their responses were coded as (A) AAE-consistent, (B) general American English-consistent, (C) other, or (D) no response. Protocols were individually scored by two trained research assistants to 100% consensus using test manual procedures. All participating children were African American AAE speakers, and average dialect density was above 50% (the DELV-S LV subtest average degree of language variation = 1.76, SD = .58, representing strong variation from general American English dialect). Raw sums of AAE consistent responses were used for the analyses. Importantly, AAE-consistent responses are not measures of *proficiency*, as only simple phonological and morphological features with AAE contrasts are assessed. Rather, AAE scores in bidialectal assessments such as the DELV-S represent the child's *usage* in the assessment context of one dialect form.

Oral reading. The Gray Oral Reading Test, 5th Edition (GORT-5; Wiederholt & Bryant, 2012) Oral Reading Index was used to assess children's oral reading ability. The GORT-5 consists of sixteen short passages, which gradually increase in length and complexity. Examinees are instructed to read the passages aloud as quickly and accurately as possible and then answer a series of five comprehension questions regarding the passage they have just read. Fluency is calculated using age-normed benchmarks of accuracy and speed after each story. The Oral Reading Index serves as a combined standard score for oral reading fluency and comprehension. Reported average internal consistency for GORT-5 oral reading index exceeds $\alpha = .90$.

Data analyses

First, a univariate data matrix was created using $n = 42$ children's responses to $n = 8$ experimental math items for a total of $n = 336$ responses. Table 1 reports children's response frequencies across all possible response patterns. See Supplemental Table A for the frequencies of strategic errors, computational errors, and correct responses by item in the experimental assessment. Correlations among the mathematical outcomes, language predictors, and a standardized measure of broad mathematics are shown in Supplemental Table B.

Next, a series of linear logistic test models (LLTMs) were tested using the PROC GLIMMIX procedure in SAS version 9.4 (SAS Institute Inc., 2013). Model testing progressed in a buildup fashion for each of three outcomes, (1) likelihood of making a strategy error, (2) likelihood of making a computational error, and (3) likelihood of correctly answering a problem. Baseline models tested the hypothesis that students differed significantly from one another in their likelihoods of each outcome (i.e., a random intercept model with no predictors). A lack of significant variance in the random intercept of the baseline model would indicate that students did not significantly differ from each other in their likelihoods of an outcome, and that subsequent testing was not warranted. When random intercepts were found, a model (Model 2) with fixed effects for the three item-level predictors was specified. Model 3 built upon Model 2 by adding a fixed effect for students' AAE dialect production and a control for reading (GORT Oral Reading Index). Finally, models examining the possibility of significant item by person interactions were examined in Models 4 (AAE by language format) and 4a (Reading by language format). Model equations are reviewed in Appendix B.

We examined a series of analyses to verify the robustness of our results. Evaluating potential confounding effects was particularly important given that this sample of children had been identified as "struggling readers." Despite testing accommodations (i.e., the word problems were read aloud as

many times as children requested during assessment), and despite the fact that the patterns of reading performance on standardized assessments are a common concern for dense AAE dialect speakers (Patton Terry, 2006; Terry, Jackson, et al., 2010; Washington et al., 2018), we attempted to ensure that reading ability was not in fact driving likelihood of errors, particularly on language-formatted problems. We added EIRT models controlling for the effect of reading ability to each of our outcome analyses. Given that the children in the current study were typically developing and had not been identified for cognitive disabilities, we did not measure or control for cognitive variables as confounds in the current study.

Results

Descriptives

The average standard scores for WJ-III Applied Problems fell within one standard deviation of the population mean of 100 ($M = 89.5$, $SD = 11.47$). Total scores on the experimental math measure ranged from 1 to 8 ($M = 4.76$, $SD = 2.00$). 41% of children's responses on the experimental measure included errors; of these errors, 53% were purely computational while 31% of errors were purely strategic errors (see Supplemental Tables A). In particular, children had difficulty retrieving math facts for simple addition and subtraction problems, and instead tended to rely on a variety of counting strategies ($\chi^2(1) = 7.71$, $p = .01$).

EIRT results

Likelihood of selecting an inappropriate strategy

As displayed in Table 2a, in the best-fitting model estimation of strategic errors, both item features and child characteristics predicted the likelihood that children would struggle with selecting appropriate strategies to solve problems (e.g., using the addition algorithm or counting up to solve a subtraction problem, reporting a retrieved math fact when not confident in its accuracy). Children were significantly less likely to struggle with selecting an appropriate strategy on addition items ($B_{addition} = -.81$, $SE = .34$, $p = .02$), as compared to subtraction items. AAE dialect production significantly interacted with item-language formatting to predict the likelihood of selecting an inappropriate strategy for solving problems, such that children with higher AAE scores were more likely to struggle with selecting an appropriate strategy on word problems ($B_{language} = .24$, $SE = .11$, $p = .03$). Figure 1 provides an illustration of this interaction with simple addition and simple subtraction problems. See Supplemental Table C for the full model taxonomy.

Likelihood of making a computational error

Conversely, the likelihood that children would make computational errors was predicted only by item-level features (see Table 2b). Children's computational errors were the most likely on subtraction ($B_{addition} = -.66$, $SE = .30$, $p = .03$) problems and problems with larger operands ($B_{simple} = -2.25$, $SE = .36$, $p < .001$). Inconsistent with existing mathematics cognition literature, however, children were actually less likely to make computational errors on language formatted items ($B_{language} = -1.00$, $SE = .31$, $p = .001$). A post hoc analysis indicated that use of counting strategies predicted the commission of computational errors ($B_{count} = 1.19$, $SE = .36$, $p = .001$), and once the use of counting strategies was included in the model, the likelihood of computational errors was no longer significantly predicted by language formatting (or Arabic numeral formatting). The full model taxonomy is provided in Supplemental Table D.

Likelihood of answering correctly

Finally, accuracy (i.e., answering correctly) was predicted by item features (Table 2c) but not by item language formatting ($B = .12$, $SE = .28$, $p = .68$). Overall, on average, children were just as likely to

Table 2. LLTM final model results.

	a) Likelihood of Strategic Error (Final Model 4)	b) Likelihood of Computational Error (Final Model 2)	c) Likelihood of Correctly Answering (Final Model 3)
<i>Fixed Effects</i>			
Intercept	-3.50 (2.79)	.36 (.30)	-1.07 (2.84)
<i>Item Fixed Effects</i>			
Language format	-1.40 (1.03)	- 1.00 (.31)**	.12 (.28)
Addition operation	-.81 (.34)*	- .66 (.30)*	1.18 (.29)***
Simple problem size	-.27 (.33)	- 2.25 (.36)***	2.06 (.31)***
<i>Person Fixed Effects</i>			
AAE dialect	.005 (.10)		-.12 (.08)
Reading	.02 (.03)		.01 (.03)
Counting			
<i>Interaction Effects</i>			
AAE dialect * Language format	.24 (.11)*		
Reading * Language format			
<i>Error Variance</i>			
Child Intercept	.81 (.50)**	.71 (.42)**	1.30 (.55)***
<i>Model Fit</i>			
-2LL	263.37	306.78	354.83
AIC	279.37	316.78	368.83

answer correctly as they were to answer incorrectly, ($B_0 = -1.07$, $SE = 2.84$, $p = .71$). Children were more likely to correctly answer addition items than subtraction items ($B = 1.18$, $SE = .29$, $p < .001$) and problems with single digit operands ($B = 2.06$, $SE = .31$, $p < .001$). However, across item features (including problem formats, operations, and problem sizes), children's production of AAE dialect was not a significant predictor of accuracy ($B = -.12$, $SE = .08$, $p = .15$). The full model taxonomy is provided in Supplemental Table E.

Reading proficiency

Reading ability was not a significant predictor of selecting an inappropriate strategy ($B = .01$, $SE = .04$, $p = .79$). Similarly, children's reading did not predict their likelihood of making computational errors ($B = -.003$, $SE = .03$, $p = .91$) or their overall accuracy on the math test ($B = .01$, $SE = .03$, $p = .73$).

Discussion

The goal of this study was to illustrate how explanatory IRT, in conjunction with error analysis, can be utilized to investigate whether and how language formatting characteristics of math word problems and bidialectal language knowledge are reflected in AAE-speakers' performance on math assessments. Children's language production in AAE was measured with the DELV-S and their math problem-solving was captured using an experimental task with both number- and language-formatted problems in a Latin square design. We used EIRT to examine the effects of assessment item formatting and person-language interactions, in conjunction with error analysis to distinguish between strategic and computational errors. We tested three hypotheses about the role of child and item level language in math problem solving. First, we expected that language formatted math problems would be more challenging than numeric formatted ones for all children, and that this challenge would be reflected in difficulty in forming a mental representation of the problem and leading to lower accuracy overall. Second, we hypothesized that language formatted items would mismatch with children's AAE dialect production, as measured by the DELV-S, thereby decreasing their likelihoods of selecting appropriate strategies to solve mathematics problems. Third, as we theorize this mismatch to be specific to language and not mathematical processing more broadly, we expected that there would be an

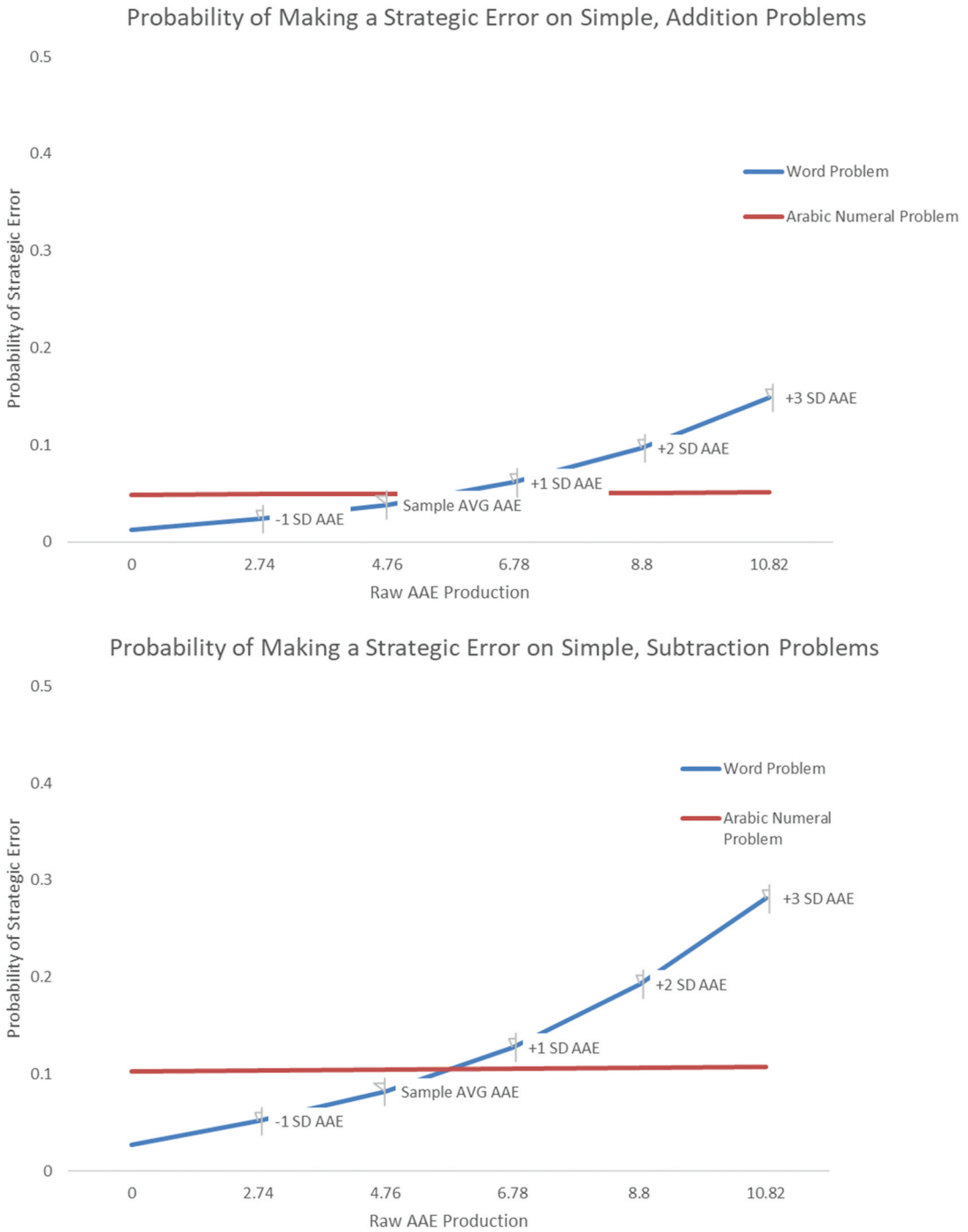


Figure 1. AAE production and item language formatting interact to predict the probability of making strategic errors (Illustrated here with simple addition and subtraction problems).

interaction between AAE dialect production and assessment language formatting, such that language formatted items (but not Arabic numeral formatted items) would differentially burden children's likelihoods of selecting appropriate strategies as a function of their AAE dialect production.

Math representations are more challenging with language formatting

The first aim of the current study was to investigate whether and how formatting was related to assessment outcomes in a sample of 2nd grade AAE-speaking bidialectal children with high proportions of AAE dialect density. In the study, there was no evidence of a relationship between language formatting and children's overall accuracy on the math assessment. However, when computational and strategic errors were disaggregated, two distinct relationships with language formatting and AAE production emerged. Overall, error responses represented only 41% of children's responses, and of these, strategic errors accounted for only 31% (i.e., strategic errors accounted for only about 13% of all responses). When children did make strategic errors, they were about 5.7 times more likely to make a strategic error on problems expressed in words rather than numbers, when holding other factors constant. Although it may appear that these strategic errors could be due to general language abilities, the AAE predictor in our analyses was not a measure of language skill but of AAE dialect production. Furthermore, the language-formatted mathematics items were designed to be relatively, linguistically simple (using words with an average age of acquisition of 4.54 years of age and a maximum of 7.42 years of age; Kuperman et al., 2012).

Understanding when children make strategic errors is important because these types of errors signal that the child has not been able to construct an accurate mental representation of the problem (Verschaffel et al., 2020), thereby misunderstanding what they are being requested to do. In the literature on word problems, these misunderstandings are often attributed to problems of reading comprehension (e.g., Fuchs et al., 2018; Vilenius-Tuohimaa et al., 2008). In the present study, children were provided with written problems as an anchor text but were read problems aloud in order to address potential problems of reading comprehension. Findings here highlight the need to further investigate the assumption, in the case of language formatted problems, that reading accommodations largely resolve language differences when measuring mathematical competence.

When children made errors, they were far more likely to make computational errors on subtraction problems and problems with larger operands. Children were only about one-third as likely to make a computational error on language as compared to numeric formatted items. While this finding is counterintuitive, it can be explained by the presence of more strategy errors on these items, that were followed by correct, albeit inappropriate, computations. Although we have no evidence as to *why* children selected particular strategies, one speculation might be that children are more likely to select strategies (albeit incorrect) with known or "easier" computations when they have not understood what the problem is asking.

Math representations are more challenging when items mismatch child language

The second and third aims of the study were to understand the relationship between children's AAE dialect use and their assessment outcomes, both directly and as interaction of language formatting on assessment items with children's AAE dialect production. Children's AAE dialect production was not a predictor of their likelihood of making computational errors, nor was it a predictor of their overall accuracy, supporting prior research suggesting that individual language characteristics are less likely to impact arithmetic problems expressed in purely numerical form (Vukovic & Lesaux, 2013). However, and aligned with prior research on AAE use and mathematics performance (Terry et al., 2022; Terry, Hendrick, et al., 2010), item formatting interacted with AAE dialect usage such that on average, children with denser dialect made more strategy errors on language-formatted items. Importantly, in the study, children's AAE dialect scores were an indicator of their AAE dialect production, and not general American English language proficiency. If these word problems were indeed language-neutral

and not assessments of language knowledge, AAE dialect usage should not have been a predictor of strategic error propagation. However, in this study, item language formatting *differentially* predicted children's likelihood of making strategic errors, depending on their levels of AAE dialect usage.

These findings, though exploratory, call for expanded investigation as mathematics assessments with language formatted items have the potential to differentially burden AAE dialect speakers, depending on their AAE language usage, perhaps through imposing a barrier to the mental representation of language formatted math problems. Using both an error analysis *and* EIRT modeling allowed for item \times person interactions to be revealed here. Without both, only item design effects (largely driven by traditional, computational errors) would have been apparent.

A unified model of test and child language

Results from the current study illustrate a novel paradigm with the potential to reveal how AAE-speaking children might draw on bidialectal resources in mathematics, using their AAE dialect to construct appropriate mental models of word problems and to select appropriate strategies for their solution. Assessment formats may present barriers to mental representation and appropriate strategy selection for reasons unrelated to mathematical knowledge or abilities, when item language does not match children's home and community language systems. A traditional analysis of correct/incorrect answers would not generally reveal these difficulties. Rather, an examination of the quality of children's errors (in this case, their difficulties selecting appropriate strategies to solve language-formatted items) was necessary to observe this effect.

Error analysis is typically used in mathematics education, on the one hand, in teacher training, where pre-service teachers are acquiring pedagogical knowledge and insights into learner misconceptions (e.g., Ryan & Williams, 2007). Additionally, error analysis is itself a pedagogical technique aimed at math learners, who benefit from both correct worked examples and examples demonstrating errors, which are then analyzed by the students themselves (e.g., Radatz, 1979; Rushton, 2018). Here, we illustrate that error analysis may have an important role beyond pedagogy, in adding nuance to analytic approaches to mathematics assessment, particularly for children from minoritized backgrounds. This diagnostic approach to analyzing children's errors has the potential to become a critical feature of psychometric evaluation that provides insight into children's strategy choices and allows us to more sensitively understand their mathematical competence by asking, what cognitive *mathematical* behaviors might link language and test performance? The study findings demonstrate that one way that assessment has the potential to mismatch with dialect variation and mismatch children's outcomes, is at a cognitive level through specific cognitive behaviors (e.g., strategy selection) but not others (computation).

Equity and mathematics assessment

Importantly, the effects of assessment language formatting demonstrated in this study are not observable unless two important conditions are met, (1) African American children are understood as linguistic minorities as opposed to simply being racial minorities, and (2) a unified model that acknowledges children's linguistic repertoires and disaggregated test item characteristics, is examined with a focus on not only outcomes, but also children's strategies used to arrive at those outcomes, whether correct or incorrect. Together, a linguistic focus on African American children combined with a diagnostic approach to assessment permits researchers to make attributions about performance that move away from a deficit perspective.

To address the first issue, including African American children in mathematical cognition research is important, and it should be done with sensitivity to cultural and linguistic identity. Assessment design features have the potential to interact with children's language systems to differentially predict performance. This is a subtle issue of assessment differences, and it likely colors our understanding of African American children's achievement disparities across the many language-formatted assessments

that are used to monitor national achievement, including mathematics, reading, writing, and cognition. Whether consistent achievement difficulties on these assessments are attributed to children's racial identities or to their environments, the conclusion has often been one of immovable, non-malleable deficit for African American children. To address this history of inequity in mathematics, and indeed in education more broadly, it is critical that children's AAE not be erased to match the assessment. Rather, potential mismatches between AAE-speaking children's rich language knowledge and the language of mathematics assessments must be better understood, in order to support children in understanding mathematical tasks, communicating their mathematical knowledge, and providing a true measure of their mathematical abilities.

Limitations and future directions

This study demonstrates the utility of using EIRT with error analysis in investigating the relationship of (bi)dialectal language to mathematics assessment outcomes and contributes to effect size and power estimation for future work. Nevertheless, caution should be used when interpreting the current model results, as increasing the sample size for both items and children is needed for more robust findings in future research. One limitation of the LLTMs employed in the current study is that they did not permit estimation of individual item parameters, but rather estimated item effects as a function of item properties. Future research should investigate EIRT models which have adequate sample sizes to permit the estimation of random effects for both items and persons. In addition, generalizations of AAE dialect users across various geographical, regional, and socioeconomic contexts should be made with care, as AAE features, prevalence, and patterns of use vary across communities and regions. Finally, AAE usage is also related to both SES and race, and SES is related to race and ethnicity. These relations are rooted in the historical and socio-political landscape of the United States, and not necessarily separable on the level of simply "control" variables. The children in the current study sample were from disadvantaged socioeconomic backgrounds. Future research should examine this phenomenon across African American children with a variety of socioeconomic resources and language usage patterns.

Conclusion

If researchers and educators are to advance theory and practice about children's mathematical learning, it is of paramount importance that we work to identify and remediate issues of equity and potential cultural/linguistic testing bias, and ultimately, design culturally and linguistically sustaining instructional approaches (Paris & Alim, 2014). Though the issue of linguistic bias is relevant for all children who are members of minoritized cultural and linguistic communities in their assessment contexts, it has often been overlooked for African American children in the United States who have bidialectal repertoires that include African American English dialect. The current study may help to inform the design and use of measurement batteries for researchers and ultimately, for clinicians by examining the contributions of seemingly subtle, but potentially great, differences in item modality to arithmetic performance. In effect, this research begins to lay the groundwork for a more nuanced model of identification of mathematical cognition among elementary school-aged children, in particular, minoritized African American children who have a rich cultural and linguistic heritage.

Acknowledgments

The authors would like to thank participating families and children, the support staff across participating research sites, and Dr. Lee Branum-Martin for feedback and initial reactions to preliminary drafts of this manuscript. [SLG] gratefully acknowledges support from a National Science Foundation grant, SMA-2204272, during the revision of this manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The Georgia State University Research on the Challenges of Acquiring Language and Literacy (ReCALL) seed grant “Contrasting two theory-based approaches to literacy instruction for poor readers who speak non-mainstream American English” helped to support this work. Also, this research was completed in part under funding from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) of the National Institutes of Health under award number HD090868. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. S.L.G. was supported by the National Science Foundation Social and Behavioral Directorate (NSF SMA-2204272) during the revision of this manuscript. Portions of these findings were presented as a poster at the 2019 Biennial Meeting of the Society for Research in Child Development, Baltimore, MD.

References

- Abedi, J. (2004). Will you explain the question? *Principal Leadership*, 4(7), 27–31.
- Abedi, J. (2006). Psychometric issues in the ELL assessment and special education eligibility. *Teachers College Record*, 108(11), 2282–2303. <https://doi.org/10.1111/j.1467-9620.2006.00782.x>
- Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, 14(3), 219–234. https://doi.org/10.1207/S15324818AME1403_2
- Abedi, J., Lord, C., & Plummer, J. R. (1997). *Final report of language background as a variable in NAEP mathematics performance (CSE Technical Report 429)*. National Center for Research on Evaluation, Standards, and Student Testing, University of California, <https://cresst.org/wp-content/uploads/TECH429.pdf>
- Banks, K., Jeddeeni, A., & Walker, C. M. (2016). Assessing the effect of language demand in bundles of math word problems. *International Journal of Testing*, 16(4), 269–287. <https://doi.org/10.1080/15305058.2015.1113972>
- Barwell, R. (2003). Linguistic discrimination: An issue for research in mathematics education. *For the Learning of Mathematics*, 23(2), 37–43.
- Beyer, T., Edwards, K. A., & Fuller, C. C. (2015). Misinterpretation of African American English bin by adult speakers of standard American English. *Language & Communication*, 45, 59–69. <https://doi.org/10.1016/j.langcom.2015.09.001>
- Brown, M. C., Sibley, D. E., Washington, J. A., Rogers, T. T., Edwards, J. R., MacDonald, M. C., & Seidenberg, M. S. (2015). Impact of dialect use on a basic component of learning to read. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.00196>
- Campbell, J. I. D., & Epp, L. J. (2005). Architectures for arithmetic. In J. I. D. Campbell (Ed.), *The handbook of mathematical cognition* (pp. 347–360). Psychology Press.
- Clinton, V., Basaraba, D. L., & Walkington, C. (2018). English learners and mathematical word problem solving: A systematic review. In D. L. Baker, D. L. Basaraba, & C. Richards-Tutor (Eds.), *Second language acquisition: Methods, perspectives and challenges* (pp. 171–208). Nova Science Publishers, Inc.
- Connor, C. M., & Craig, H. K. (2006). African American preschoolers’ language, emergent literacy skills, and use of African American English: A complex relation. *Journal of Speech, Language, and Hearing Research*, 49(4), 771–792. [https://doi.org/10.1044/1092-4388\(2006/055\)](https://doi.org/10.1044/1092-4388(2006/055))
- Craig, H. K., Thompson, C. A., Washington, J. A., & Potter, S. L. (2004). Performance of elementary-grade African American students on the Gray Oral Reading Tests. *Language, Speech, and Hearing Services in Schools*, 35(2), 141–154. [https://doi.org/10.1044/0161-1461\(2004/015\)](https://doi.org/10.1044/0161-1461(2004/015))
- 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.
- Cruz Neri, N., & Retelsdorf, J. (2022). The role of linguistic features in science and math comprehension and performance: A systematic review and desiderata for future research. *Educational Research Review*, 36, 100460. <https://doi.org/10.1016/j.edurev.2022.100460>
- De Boeck, P., & Wilson, M. (Eds.). (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. Springer New York. <https://doi.org/10.1007/978-1-4757-3990-9>
- Engelhardt, J. M. (1977). Analysis of children’s computational errors: A qualitative approach. *British Journal of Educational Psychology*, 47(2), 149–154. <https://doi.org/10.1111/j.2044-8279.1977.tb02340.x>
- Fedorenko, E., Gibson, E., & Rohde, D. (2007). The nature of working memory in linguistic, arithmetic and spatial integration processes. *Journal of Memory and Language*, 56(2), 246–269. <https://doi.org/10.1016/j.jml.2006.06.007>

- Fuchs, L. S., Gilbert, J. K., Fuchs, D., Seethaler, P. M., & Martin, B. N. (2018). Text comprehension and oral language as predictors of word-problem solving: Insights into word-problem solving as a form of text comprehension. *Scientific Studies of Reading*, 22(2), 152–166. <https://doi.org/10.1080/10888438.2017.1398259>
- Garcia, F. M., Shen, G., Avery, T., Green, H. L., Godoy, P., Khamis, R., & Froud, K. (2022). Bidialectal and monodialectal differences in morphosyntactic processing of AAE and MAE: Evidence from ERPs and acceptability judgments. *Journal of Communication Disorders*, 100, 106267. <https://doi.org/10.1016/j.jcomdis.2022.106267>
- Gatlin, B., & Wanzek, J. (2015). Relations among children's use of dialect and literacy skills: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 58(4), 1306–1318. https://doi.org/10.1044/2015_JSLHR-L-14-0311
- Geary, D. C., & Wiley, J. G. (1991). Cognitive addition: Strategy choice and speed-of-processing differences in young and elderly adults. *Psychology and Aging*, 6(3), 474–483. <https://doi.org/10.1037/0882-7974.6.3.474>
- Greenstein, J., & Strain, P. S. (1977). The utility of the Key Math Diagnostic Arithmetic Test for adolescent learning disabled students. *Psychology in the Schools*, 14(3), 275–282. [https://doi.org/10.1002/1520-6807\(197707\)14:3<275:AID-PITS2310140305>3.0.CO;2-T](https://doi.org/10.1002/1520-6807(197707)14:3<275:AID-PITS2310140305>3.0.CO;2-T)
- Hopewell, S., & Escamilla, K. (2014). Struggling reader or emerging biliterate student? Reevaluating the criteria for labeling emerging bilingual students as low achieving. *Journal of Literacy Research*, 46(1), 68–89. <https://doi.org/10.1177/1086296X13504869>
- Imbo, I., & Vandierendonck, A. (2007). The development of strategy use in elementary school children: Working memory and individual differences. *Journal of Experimental Child Psychology*, 96(4), 284–309. <https://doi.org/10.1016/j.jecp.2006.09.001>
- Jacobson, L. A., Koriakin, T., Lipkin, P., Boada, R., Frijters, J. C., Lovett, M. W., Hill, D., Willcutt, E., Gottwald, S., Wolf, M., Bosson-Heenan, J., Gruen, J. R., & Mahone, E. M. (2017). Executive functions contribute uniquely to reading competence in minority youth. *Journal of Learning Disabilities*, 50(4), 422–433. <https://doi.org/10.1177/0022219415618501>
- Ketterlin-Geller, L. R., & Yovanoff, P. (2009). Diagnostic assessments in mathematics to support instructional decision making. *Practical Assessment, Research & Evaluation*, 14(16), 1–11. <https://doi.org/10.7275/VXRK-3190>
- Kieffer, M. J., Lesaux, N. K., Rivera, M., & Francis, D. J. (2009). Accommodations for English language learners taking large-scale assessments: A meta-analysis on effectiveness and validity. *Review of Educational Research*, 79(3), 1168–1201. <https://doi.org/10.3102/0034654309332490>
- Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavioral Research Methods*, 44(4), 978–990. <https://doi.org/10.3758/s13428-012-0210-4>
- Lee-James, R., & Washington, J. A. (2018). Language skills of bidialectal and bilingual children. *Topics in Language Disorders*, 38(1), 5–26. <https://doi.org/10.1097/TLD.0000000000000142>
- LeFevre, J.-A., Sadesky, G. S., & Bisanz, J. (1996). Selection of procedures in mental addition: Reassessing the problem size effect in adults. *Journal of Experimental Psychology*, 22(1), 216–230. <https://doi.org/10.1037//0278-7393.22.1.216>
- Liu, R., & Bradley, K. D. (2021). Differential item functioning among English language learners on a large-scale mathematics assessment. *Frontiers in Psychology*, 12, 12. <https://doi.org/10.3389/fpsyg.2021.657335>
- Martiniello, M. (2008). Language and the performance of English-language learners in math word problems. *Harvard Educational Review*, 78(2), 333–368. <https://doi.org/10.17763/haer.78.2.70783570r1111t32>
- Mazzocco, M., Murphy, M., Brown, E., Rinne, L., & Herold, K. (2013). Persistent consequences of atypical early number concepts. *Frontiers in Psychology*, 4, 4. <https://doi.org/10.3389/fpsyg.2013.00486>
- McGrew, K. S., & Woodcock, R. W. (2001). *Woodcock-Johnson III Tests of Achievement*. Riverside Publishing Company.
- Moschkovich, J. N. (2010). *Language and mathematics education: Multiple perspectives and directions for research*. Information Age Publishing.
- Ohlsson, S. (1996). Learning from performance errors. *Psychological Review*, 103(32), 241–262. <https://doi.org/10.1037/0033-295X.103.2.241>
- Paris, D., & Alim, H. S. (2014). What are we seeking to sustain through culturally sustaining pedagogy? A loving critique forward. *Harvard Educational Review*, 84(1), 85–100. <https://doi.org/10.17763/haer.84.1.982l873k2ht16m77>
- Patton Terry, N. (2006). Relations between dialect variation, grammar, and early spelling skills. *Reading and Writing*, 19(9), 907–931. <https://doi.org/10.1007/s11145-006-9023-0>
- Patton Terry, N. (2008). Addressing African American English in early literacy assessment and instruction. *Perspectives on Communication Disorders & Sciences in Culturally & Linguistically Diverse (CLD) Populations*, 15(2), 54–61. <https://doi.org/10.1044/cds15.2.54>
- Puranik, C., Branum-Martin, L., & Washington, J. A. (2020). The relation between dialect density and the codevelopment of writing and reading in African American children. *Child Development*, 91(4). <https://doi.org/10.1111/cdev.13318>
- Radatz, H. (1979). Error analysis in mathematics education. *Journal for Research in Mathematics Education*, 10(3), 163–172. <https://doi.org/10.2307/748804>
- Rhodes, K. T., Branum-Martin, L., Morris, R. D., Rowski, M., & Sevcik, R. A. (2015). Testing math or testing language? The construct validity of the Key Math-revised for children with intellectual disability and language difficulties. *American Journal on Intellectual and Developmental Disabilities*, 120(6), 542–568. <https://doi.org/10.1352/1944-7558-120.6.542>
- Roberts, G. H. (1968). The failure strategies of third grade arithmetic pupils. *The Arithmetic Teacher*, 15(5), 442–446. <https://doi.org/10.5951/AT.15.5.0442>

- Rushton, S. J. (2018). Teaching and learning mathematics through error analysis. *Fields Mathematics Education Journal*, 3(1), 4. <https://doi.org/10.1186/s40928-018-0009-y>
- Ryan, J., & Williams, J. (2007). *Children's mathematics 4-15: Learning from errors and misconceptions*. McGraw-Hill Education (UK).
- Seymour, H. N., Roeper, T. W., & de Villiers, J. (2003). *Diagnostic Evaluation of Language Variation-Screening Test*. NCS Pearson, Inc.
- Shaftel, J., Belton-Kocher, E., Glasnapp, D., & Poggio, J. (2006). The impact of language characteristics in mathematics test items on the performance of English language learners and students with disabilities. *Educational Assessment*, 11(2), 105–126. https://doi.org/10.1207/s15326977ea1102_2
- Siegler, R. S. (1991). Strategy choice and strategy discovery. *Learning and Instruction*, 1(1), 89–102. [https://doi.org/10.1016/0959-4752\(91\)90020-9](https://doi.org/10.1016/0959-4752(91)90020-9)
- Siegler, R. S., & Shrager, J. (1984). Strategy choices in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *Origins of cognitive skills: The eighteenth annual Carnegie symposium on cognition* (pp. 229–293). Lawrence Erlbaum Associates Publishers. <https://cir.nii.ac.jp/crid/1571980075464441088>
- Siegler, R. S., & Taraban, R. (1986). Conditions of applicability of a strategy choice model. *Cognitive Development*, 1(1), 31–51. [https://doi.org/10.1016/S0885-2014\(86\)80022-3](https://doi.org/10.1016/S0885-2014(86)80022-3)
- Stockman, I. J. (2010). A review of developmental and applied language research on African American children: From a deficit to difference perspective on dialect differences. *Language, Speech, and Hearing Services in Schools*, 41(1), 23–38. [https://doi.org/10.1044/0161-1461\(2009/08-0086\)](https://doi.org/10.1044/0161-1461(2009/08-0086))
- Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, 20(4), 345–354. <https://doi.org/10.1111/j.1745-3984.1983.tb00212.x>
- Tatsuoka, K. K. (1985). A probabilistic model for diagnosing misconceptions by the pattern classification approach. *Journal of Educational Statistics*, 10(1), 55–73. <https://doi.org/10.3102/10769986010001055>
- Tatsuoka, K. K. (1990). Toward an integration of item-response theory and cognitive error diagnoses. In N. Fredericksen, R. Glaser, A. Lesgold, & M. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 453–488). Lawrence Erlbaum Associates Publishers.
- Terry, J. M., Hendrick, R., Evangelou, E., & Smith, R. L. (2010). Variable dialect switching among African American children: Inferences about working memory. *Lingua*, 120(10), 2463–2475. <https://doi.org/10.1016/j.lingua.2010.04.013>
- 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, J. M., Thomas, E. R., Jackson, S. C., Hirotani, M., & Finley, S. (2022). African American English speaking 2nd graders, verbal-s, and educational achievement: Event related potential and math study findings. *PLoS One*, 17(10), e0273926. <https://doi.org/10.1371/journal.pone.0273926>
- U.S. Census Bureau. (2021). *QuickFacts: United States*. <https://www.census.gov/quickfacts/fact/table/US/PST045221>
- Van Den Noortgate, W., De Boeck, P., & Meulders, M. (2003). Cross-classification multilevel logistic models in psychometrics. *Journal of Educational and Behavioral Statistics*, 28(4), 369–386. <https://doi.org/10.3102/10769986028004369>
- Verschaffel, L., Schukajlow, S., Star, J., & Van Dooren, W. (2020). Word problems in mathematics education: A survey. *International Journal on Mathematics Education*, 52(1), 1–16. <https://doi.org/10.1007/s11858-020-01130-4>
- Vilenius-Tuohimaa, P. M., Aunola, K., & Nurmi, J. (2008). The association between mathematical word problems and reading comprehension. *Educational Psychology*, 28(4), 409–426. <https://doi.org/10.1080/01443410701708228>
- Vukovic, R. K., & Lesaux, N. K. (2013). The language of mathematics: Investigating the ways language counts for children's mathematical development. *Journal of Experimental Child Psychology*, 115(2), 227–244. <https://doi.org/10.1016/j.jecp.2013.02.002>
- Washington, J. A., Branum-Martin, L., Sun, C., & Lee-James, R. (2018). The impact of dialect density on the growth of language and reading in African American children. *Language, Speech & Hearing Services in Schools*, 49(2), 232–247. PMID: PMC6105135. https://doi.org/10.1044/2018_LSHSS-17-0063
- Wiederholt, J., & Bryant, R. (2012). *Gray Oral Reading Tests* (5th ed.). Pro-Ed.
- Wolf, M. K., Kim, J., & Kao, J. (2012). The effects of glossary and read-aloud accommodations on English language learners' performance on a mathematics assessment. *Applied Measurement in Education*, 25(4), 347–374. <https://doi.org/10.1080/08957347.2012.714693>