The effects of race and maternal education level on children’s retells of the *Renfrew Bus Story* — *North American Edition*

Anne van Kleeck  
University of Texas at Dallas  
Alissa Lange  
State University of New York at Buffalo  
Amy Louise Schwarz  
University of Texas at Dallas
Abstract

Purpose: The Renfrew Bus Story North American edition (RBS-NA) is widely used clinically and in research for determining children’s language abilities, although possible influences of race and maternal education on RBS-NA performance are unknown. The current study compared RBS-NA retells of four groups of children: African American (AA) and European American (EA) whose mothers had an education level of high school or less (≤HS), or had an education level higher than high school (>HS).

Method: Statistical tests were conducted on 163 kindergartners’ story retells using raw scores for all four RBS-NA measures: Information, Sentence Length, Complexity, and Independence.

Results: A 2 x 2 ANOVA indicated main effects for maternal education and race for the Information score, with ≤HS and AA children scoring lower. For measures not meeting ANOVA assumptions, 2 x 2 ANOVAs using ranked data indicated significant differences across groups for maternal education on Sentence Length, Complexity, and Independence scores, with ≤HS scoring lower. There were no additional effects of race.

Conclusion: There are systematic effects of maternal education and race on children’s RBS-NA performance, which is important for both researchers and clinicians to take into account when using this instrument.
The effects of race and maternal education level on children’s retells of the Renfrew Bus Story –
North American edition

Producing narratives requires the integration of a variety of linguistic and social-cognitive skills. In the linguistic realm, producing a narrative requires coordinating semantic, syntactic, morphologic, and phonological elements both within and across sentences. Social-cognitive skills impact narrative production via the cultural conventions that shape the way narratives are organized, shared, and adapted to a listener’s level of background and shared information on the topic. Due to this complexity, assessing young children’s narrative skills is considered to be both “an ecologically valid and educationally relevant means of language assessment” (e.g., Pankratz, Plante, & Insalaco, 2007, p. 390).

Narrative is important in its own right as a “ubiquitous and universal construct used to inform, socialize, entertain, and teach” (McGregor, 2000, p. 55). A number of researchers have hypothesized that skill with narratives in preschool and the early elementary years would also be a helpful component in later ability to comprehend connected text when reading (Anderson, Anderson, Lynch, & Shapiro, 2003; Scarborough, 2001; Sénéchal & LeFevre, 2002; Storch & Whitehurst, 2002; Tabors, Snow, & Dickinson, 2001; van Kleeck, 2007, 2008). Indeed, several empirical studies have found that narrative abilities do predict later reading comprehension, reading fluency, or written narrative skills in both children who have learning disabilities and those who are typically developing (e.g., Feagans & Appelbaum, 1986; Griffin, Hemphill, Camp, & Wolf, 2004; Reese, Suggate, Long, & Schaugency, 2009; Snyder & Downey, 1991; Tabors, et al., 2001). In a large study of over 1,500 Spanish/English bilingual children from kindergarten to grade three, oral narratives (retells of a wordless picture book) were in fact the best predictor of third grade reading comprehension (Miller, et al., 2006). Furthermore, narrative skills
uniquely contribute to reading fluency even after controlling for receptive vocabulary and decoding skills (Reese, et al., 2009).

In the studies in which narrative ability has not predicted later reading ability (e.g., Menyuk, et al., 1991; O’Neill, Pearce, & Pick, 2004; Roth, Speece, & Cooper, 2002; Snow, Tabor, Nicholson, & Kurland, 1995; Stein & Glenn, 1982), researchers have measured children’s reading outcomes in the second grade or earlier when word recognition still dominates both the curriculum and tests of reading ability. Narrative ability would not be expected to predict performance until children are required to begin engaging in independent comprehension of longer and more complex connected text, which typically occurs in the third or fourth grade. And indeed, using structural equation modeling, Storch and Whitehurst (2002) found that the strong and direct link between oral language and literacy found during preschool disappeared during the first two years of formal reading instruction, but reappeared in the third and fourth grade. Others have also found that the links between oral language skills and reading comprehension become stronger as children get older (e.g., Vellutino, Tunmer, Jaccard, & Chen, 2007). So, the importance of narrative ability during the preschool years appears to accrue when the curriculum shifts from what is frequently referred to as “learning to read” in kindergarten through second grade to independently “reading to learn” beginning in third grade.

When evaluating children’s narrative productions, the complexity of the task is acknowledged in that two levels of measurement are typically conducted, a micro- and a more macro-level (Petersen, Gillam, & Gillam, 2008). The micro-level measures focus on such things as word diversity and frequency and sentence level complexity. These measures are the same as those typically used when analyzing spontaneous language samples of conversational dialogue. So, for example, syntactic complexity is often assessed via measures such as mean length of
utterance in words or morphemes (MLU) and percent of complex sentences produced. The lexical skills of a child are captured by determining the diversity of words used in the sample, often using a measure such as the number of different words (NDW). When using these measures, narratives are considered to be one type of language sample, one that can either be spontaneously generated by the child or consist of a retell of a story told to the child. The *Renfrew Bus Story North American edition (RBS-NA)* story retell used in the current study captures the micro-level of narrative analysis at the sentence level with two measures. One is called Sentence Length, and it indicates the length of the child’s five longest sentences, the other is called Complexity, and it indicates the amount of complex syntax produced. The RBS-NA does not include micro-level measures at the word level.

To consider how children are able to organize their narratives at a more global level that goes beyond the word and sentence level, macro-level measures are used. This level is typically tapped by assessing children’s ability to organize their narratives into the elements of story grammar (e.g., Serpell, Baker, & Sonnenschein, 2005), or by assessing the amount of information from the story that is included in story retell tasks. The RBS-NA contains one macro-level measure called Information.

To date, The *Bus Story Test* (Renfrew, 1969) is the only norm-referenced screening measure of young children’s narrative abilities that spans the preschool and kindergarten age range, a period of development during which it is important to be able to measure the language skills foundational to later reading, and work to strengthen those that are weak. This assessment tool was originally developed in England as the *Bus Story Test*, and now has a North American version, the *Renfrew Bus Story – North American edition* (RBS-NA: Glasgow & Cowley, 1994) that was used in the current study. This assessment includes a story and twelve accompanying
pictures, presented on four pages, about a bus that runs away from its driver. The story is read to the child who then retells the story to the examiner. In addition to the micro- and macro-level measures mentioned above, the child’s retell is also analyzed for the amount of prompting required for the child to discuss the story, a measure called Independence. Both versions of the *Bus Story* have been shown to be predictors of persistent language impairment (Bishop & Edmundson, 1987; Pankratz, et al., 2007; Stothard, Snowling, Bishop, Chipchase, & Kaplan, 1998), and the British version has also shown a strong relationship to later literacy (Stothard, et al., 1998).

The *Bus Story* is widely used in research in the United States and the United Kingdom that has focused on children who are developing their language in a typical fashion as well as those with language impairments (e.g., Anderson, Morse, Catroppa, Haritou, & Rosenfeld, 2004; Anderson, et al., 1997; Bishop & Edmundson, 1987; Botting, Conti-Ramsden, & Crutchley, 1997; Briscoe, Gathercole, & Marlow, 1998; Conti-Ramsden & Botting, 1999; Conti-Ramsden, Crutchley, & Botting, 1997; Dockrell & Lindsay, 2001; Gallagher, Frith, & Snowling, 2000; Girolmetto, Wiigs, Smyth, Weitzman, & Pearce, 2001; Gray, 2004; Hohen & Stevenson, 1999; Hughes, Dunn, & White, 1998; Kennedy, et al., 2006; Kovas, et al., 2005; Mackie & Dockrell, 2004; Nathan, Stackhouse, Goulandris, & Snowling, 2004; Paul, 1996; Paul, Hernandez, Taylor, & Johnson, 1996; Paul & Smith, 1993; Storch & Whitehurst, 2002; Stothard, et al., 1998). In the United Kingdom, it is believed to be widely used clinically (Norbury & Bishop, 2003), being “regularly employed in primary language units” (Botting, 2002, p. 2) that serve children with specific language impairments (Conti-Ramsden & Botting, 1999). According to the distributor of the North American version of the test, test sales are about 60% to Canada and 40% to the
U.S.A. Buyers include both speech-language pathologists and researchers in speech-language pathology and education (Lange, 2010).

Concerns have been raised that the North American version of this assessment tool may over-identify minority children as having poor narrative skills. In a review of this test, Haccoun (2001) noted that African American (AA) children in the normative data appeared to obtain lower scores, but unfortunately, information on SES is not provided in the norms. Data from 2008 shows that AA children are over three times more likely to be poor than their European American (EA) counterparts (Federal Interagency Forum on Child and Family Statistics, 2010). Furthermore, children living in households headed by single mothers were five times more likely to be living in poverty than those living in households headed by two parents. For African Americans, the poverty rate for households headed by single mothers was 50.2%, and the number of female headed households in which there are children under 18 is approximately 2 ½ times higher for African Americans than for the total US population. Given the overall confound of race and income, it is unclear if AA children’s lower scores are related to race or SES.

Two key variables often used in determining SES are income and parental education. As Dollaghan and her colleagues discuss (Dollaghan, Campbell, Paradise, Feldman, Janosky, Pitcair & Kurs-Lasky, 1999), education level is considered to be the more stable (Huston, McLoyd, & Garcia-Coll, 1994) and less controversial (Hauser, 1994) of the two. Furthermore, maternal education level is the best predictor of intellectual attainment when family income, maternal education level and paternal education level are compared (Mercy & Steelman, 1982). This may be because, except for cases of extreme poverty, maternal education level is the component of SES most strongly related to parenting measures (Bornstein, Hahn, Suwalsky, & Haynes, 2003; Hoff, Laursen, & Tardif, 2002).
More directly tied to language ability, in a study of 701 African American preschoolers, performance on the *Preschool Language Scale-3* (Zimmerman, Steiner, & Pond, 1992) was significantly different for different maternal education levels, but not for monthly income for children of the 590 mothers who provided this information (Qi, Kaiser, Milan, Yzquierdo, & Hancock, 2003). Furthermore, 81 percent of these mothers reported being single female heads of household. Research from these various perspectives supports our use of maternal education level as a marker for SES in the current study.

Previous research does not illuminate the potential impact of race on children’s performance on the *RBS-NA*. Although a study by Price and her colleagues (Price, Roberts, & Jackson, 2006) used the *Bus Story Test* to study the structural development of AA children’s narratives, their data are not helpful for determining whether EA and AA children perform the same or differently on the *Renfrew Bus Story – North American edition* because (a) no EA children were studied, (b) the British version of the test was used (the North American version was not yet available when their data were collected), and (c) these researchers developed their own method of coding the narratives using a story grammar paradigm rather than following the standardized analyses in the test manual.

Pankratz et al. (2007) obtained *Renfrew Bus Story – North American edition* retells from racially and ethnically heterogeneous groups of preschoolers with specific language impairment and those who were developing language in a typical fashion. When they grouped together minority children in their sample (Hispanic, African American, and other), they found that, compared to the EA preschoolers, the minority children scored significantly lower on the Sentence Length score (number of words in the five longest utterances produced in the retells), but not on the Information score (information provided by the child that corresponded to
information in the original story). This pattern also held when comparing just the Hispanic children to the EA children, but there were not enough children in the other minority groups to conduct similar analyses.

These studies just reviewed lead to potential concern regarding impact of race on children’s performance on the RBS-NA. We know of no research related to the impact of maternal education level on the RBS-NA. However, there are two relevant studies that have used other narrative measures.

Reese had her colleagues (Reese et al., 2009) explored the relationship between maternal education level and three aspects of story book retell narratives produced by 61 six-year-old (Study 1) and 39 seven-year-old (Study 2) children. In both studies, maternal education did not significantly correlate with any of the oral narrative measures. It should be noted, however, that there were no children of university graduates in this sample. This narrow range of maternal education level may have obscured the impact of maternal education level on the children’s retells.

Kulkofsky and Klemfuss (2008) examined narratives of 112 preschoolers that focused on a previously witnessed staged event. In Study 1 \((n = 46)\), 89\% of the mothers had at least a college degree; in Study 2 \((n = 66)\), 72\% of the mothers had a college degree. There were no main effects or interactions for maternal education, nor did maternal education correlate with narrative quality. However, here again the range of maternal education in this study may have been too narrow, although in this case containing too many mothers with higher education levels.

Given that previous research sheds little light on the topic, the goal of the current study was to determine the effects, if any, of race and maternal education on kindergarten children’s

**Method**

**Participants**

Data were gathered from a sample of the children participating in the large-scale research project, *Scaling-up TRIAD: Teaching Early Mathematics for Understanding with Trajectories and Technologies*, for which approval from the primary university’s Institutional Review Board was obtained following all local and federal guidelines for the protection of human subjects. This study was a U.S. Department of Education-funded randomized control trial involving two urban public school districts. District agreement to participate was obtained, and all eligible schools within each district were then randomly assigned to Control or Treatment groups, including all preschool classrooms in each school (the treatment took place the school year before the retell data for the current study was collected when the children were in kindergarten). Thus, all classrooms within each school were assigned to one condition. Children in the initial sample (n = 1305) were from 106 classrooms, in 42 schools in Buffalo, New York and Boston, Massachusetts. On average, 81.6% of children in these schools were receiving free or reduced school lunches, and the racial/ethnic distributions across the two school districts were the following: 43% African American, 35% European American, 19% Hispanic or Latino, and 4% Asian American).

**Maternal education and race.** The maternal education status was taken from a question in the *Parent Questionnaire of the TRIAD study*. The questionnaire included questions about parent education and employment, as well as questions about the child relevant to the math study, such as parent-perceived child counting ability. The question
about maternal education from this measure asked the respondent to report the highest level of education attained by the child's mother by indicating one of seven categories (see Table 1). The self-report Parent Questionnaire data were obtained by mail or by phone by a trained interviewer during the child's pre-kindergarten year.

Table 1 shows the number of African American (AA) and European American (EA) children according to the seven maternal education categories used in the TRIAD study. The race information was provided by the school districts involved. In the current study, we divided these seven education categories to create two levels of maternal education, high school and less (≤HS), and higher than high school (>HS). However, to assure equivalence in maternal education for the AA and EA children in the current study, we made sure that our sample had equal numbers of children in each of the original seven maternal education categories. To do this, we set the number of children in each of the seven maternal education categories as the smallest number of either AA or EA children in that group in the original sample.

For example, the top section of Table 1 indicates that our search of the TRIAD database yielded 26 AA and 13 EA children whose mothers had an eighth grade education. As shown in the middle section of Table 1, we included all 13 of the EA children in our sample for this study. We then randomly selected 13 out of the 26 AA children to participate in this study. We followed this same procedure for each maternal education category until we had an equal number of children by race in each maternal education category, as shown in the middle section of Table 1. We then collapsed the seven maternal education categories into two categories as shown in the bottom section of Table 1, which assigned children to either the ≤HS or the >HS maternal education level. Splitting the data into these two maternal education levels made sense conceptually and, given the sample we were using, this particular split for maternal education
level also allowed us to have roughly equivalent sample sizes at each maternal education level.

**Children excluded.** We excluded children who were enrolled in special education programs because they were not distributed proportionally in the original sample across race and the seven maternal education categories. We also excluded control group children because there were far fewer of them, and we did not want to have groups in which treatment and control children were mixed, since the narrative-retell data was collected at the end of the math intervention. This was the conservative approach to take given that in other research on a math intervention with preschoolers (reported in Preschool Curriculum Evaluation Research Consortium, 2008), very small effects (effect sizes ranged from $d = 0.08$ to 0.17) on language abilities were found (immediately post treatment or one year later) as measured by the *Peabody Picture Vocabulary Test, Third Edition* (Dunn & Dunn, 1997) or the *Test of Language Development Primary, Third Edition* (Newcomer & Hammill, 1997). Our final sample for this study included 172 children, with 86 AA and 86 EA children (see Table 2). A total of eighty-two children fell within the $<\text{HS}$ maternal education level, and 90 fell within the $>\text{HS}$ maternal education level.

**Measures: Renfrew Bus Story – North American edition**

The *RBS-NA* is a standardized, norm-referenced screening measure of oral language using a story retell narrative language sample. The assessment involves telling a child a story, and then asking the child to retell the story using the pictures in a wordless storybook as prompts. The stories that children retold while taking the *RBS-NA* were audio and video recorded, transcribed and scored on the four measures described in the test manual. For transcription, utterances were segments of language that consisted of main clauses and their attached subordinate clauses or sentence fragments. As noted earlier,
there are three primary measures rating the content (macro-level) and form (micro- or sentence-level) of the story retells, including Information, Complexity, and Sentence Length. One additional subscale, Independence, taps the child’s self-initiative in generating the story retell.

The Information subscale is a macro-level measure of how many of the 32 key pieces of information (including both words and phrases) from the original story the children used in their story retell. The emphasis is on how much of the information the child uses in the retell, so children can get credit for a response that matches the ‘key’ in meaning, even if not in identical word use. Still, many of the ‘key’ words or phrases are only given full credit if they do match exactly. The total possible raw score, 52, includes some items worth 2 points. This subscale represents the proficiency level of a set of integrated skills. In order to score well on this measure, children must remember the names of the key words or concepts (memory), know the meaning of the words well enough to use them appropriately in their retell (vocabulary), and have a sufficient understanding of story structure to use the words in the right sequence (book/story knowledge).

Sentence Length is a straightforward micro-level measure of morpho-syntactic complexity. For this purpose, the RBS-NA uses the average number of words in the five longest sentences a child generates in the retell. Complexity, also a micro-level narrative measure, looks at the number of utterances containing a subordinate clause that children use in their story retells. It is also an indicator of syntactic development.

The Independence score relates to neither macro- nor micro-level structure of the narrative itself, but is focused instead on the amount of prompting the child needs from the assessor in order to retell the story. The total prompt score is the summed prompt scores with a
maximum of 4 possible prompts available per picture, and prompts are reverse scored. For example, “no prompt” is given 4 points, while the 4th prompt, “The bus…,” is scored 0 points. Children who retell the story without any need for prompting from the assessor will get the highest Independence score, while children who need prompts will score lower on Independence. The Independence measure was scored as directed in the manual.

**Procedure**

Assessors were trained prior to administering the **RBS-NA**. Assessors included research assistants and retired school teachers. Training consisted of a lecture on the importance of strict adherence to protocol, viewing practice assessments, and real-time test administration practice with feedback.

Children were individually administered the **RBS-NA** in the fall semester of kindergarten; the math treatment had occurred during the prior school year in pre-kindergarten. They were assessed in an open space (e.g., library, hallway) in their school by a trained assessor. Assessments were videotaped and audio taped to facilitate transcription. Audiotapes served as a backup in case of videotape malfunction, and also aided in the transcription process when the videotapes were difficult to hear.

**Reliability.** The reliability of the transcription and scoring procedures was calculated for the full sample \((n = 1027)\). Reliability of the transcription procedure was determined by evaluating a one-way, single-measure interclass correlation (ICC) using a random 7% sample of stories that were transcribed by two trained research assistants. The two sets of transcriptions were then expert scored by the same researcher, and the ICCs between the pairs of scores on three of the measures, which represented the reliability of the transcription process from the tape, were 0.99 for Information, 0.97 for Sentence Length, and 0.94 for Complexity. Transcription
reliability for the Independence score was not determined, because this was an objective recording of the type of prompt that was provided and recorded by the assessor during test administration. Scorers/transcribers checked 100% of the transcriptions for the accuracy of the Independence scores given by the assessors. Errors were made in less 1% of the time, and these were corrected.

Inter-rater reliability for the scoring procedure was calculated using ICCs to determine scoring agreement between two different scorers using the same transcriptions (a random 10% of samples for Information, Sentence Length, and Complexity, and a random 8% for Independence). A one-way ICC with single measures were used to account for error due to having more than two raters, but not always the same two raters rating each target. Agreement was 0.97 for Information, 0.94 for Sentence Length, 0.81 for Complexity, and 0.94 for Independence.

Results

Box plots were used to identify outliers for each dependent measure. Specifically, data points were considered outliers if they were beyond the first and third quartiles by 1.5 times the interquartile range (Hinkle Wiersma, & Jurs, 2003). The box plots indicated that each dependent measure included two to three outlier scores. A total of nine children produced the outlier scores across all four dependent measures. These nine children were distributed relatively equally across the four groups. For ease of reporting, as shown in Table 2, we removed these same nine children from the analysis of each dependent measure; so our analysis is based on an \( n = 163 \) with 82 African American children and 81 European American children. Seventy-nine children fell within the \(<\)HS category while 84 fell within the \(>\)HS category.

We first wanted to know whether our results were due to differences in gender and site of
testing (Buffalo vs. Boston). Only the Information score was normally distributed so we conducted a two-tail independent t-test on the Information score variable and two-tail Mann Whitney Tests on the remaining three measures. Table 3 contains the results of the gender analysis and Table 4 contains the results of the site analysis. The sample contained 80 males and 83 females (eliminating any outliers as discussed above). There were no significant differences for gender for any of the variables. The sample contained 133 children from Buffalo and 30 children from Boston. There were no significant differences for site for any of the variables. As such, we proceeded in our planned analyses collapsing over both gender and site.

Two-factor, fixed-effect, analyses of variance (ANOVA) with two levels for the independent variable race (AA or EA) and two levels for the independent variable maternal education (<HS and >HS) were planned for four measures in the RBS-NA test manual (Information, Sentence Length, Complexity, and Independence). A power analysis indicated that an experimental effect could be detected only if the independent variables explained at least 3.86% of the variance of a given dependent measure (Abdi, Edelman, Valentine, & Dowling, 2009). For main effects, pooled variances to determine Cohen’s $d$ effect sizes were calculated according to Cortina and Nouri (2000) and confidence intervals were calculated according to Altman (2000). We followed Cortina and Nouri’s (2000) recommendation to only report Cohen’s $d$ effect sizes for main effects because there is disagreement on what procedure to use when calculating Cohen’s $d$ values for interactions. We also followed their procedure to calculate the pooled variances needed for the Cohen’s $d$ calculations. Since only two of the test manual measures had standard scores, we chose to report findings for all measures using raw scores for consistency and ease of reporting. Each dependent measure was tested for meeting the assumptions of ANOVA using Chi Square for independence, the Lilliefors test (Lilliefors, 1969)
for normality, and the O’Brien test of equal variance (O’Brien, 1979, 1981; Trujillo-Ortiz & Hernandez-Walls, 2003). The means and standard deviations for the measures from the RBS-NA are shown in Table 5. Only the Information score met the assumptions of ANOVA. The Sentence Length data and the Complexity data had a right skewed distribution and equal variance, while the Independence data had a left skewed distribution and unequal variance. Attempts to transform these data to meet the assumptions of ANOVA were unsuccessful. When no nonparametric test exists that fits the experimental design, Conover (1999) recommends conducting an ANOVA on the raw data and on the rank transformed data. “If the two procedures give nearly identical results, the assumptions underlying the usual analysis of variance are likely to be reasonable and the regular parametric analysis valid” (p. 419-420). Thompson (1991) states that ANOVAs on ranked data are valid only when there is one significant main effect. In this situation, the ranked data conform to a non-central $X^2$ distribution. When the ANOVA on ranked data indicates more than one significant main effect and a significant interaction, the approach of using ranked data in ANOVA is invalid (Leys & Schumann, 2010; Thompson, 1991).

When we conducted ANOVAs on both the raw and ranked data for the three variables not meeting the ANOVA assumptions (i.e., sentence length, complexity, independence) as Conover (1999) recommends, we found that there was only one main effect in each case. Given that our data conform to the situation in which ANOVA using ranked data is valid according to Thompson (1991), we reported the results for sentence length, complexity, and independence as ANOVAs based on ranked data. We set the alpha level at .05 and applied the Bonferonni correction, which resulted in a correct alpha of .0125 (.05/4, because we had one inferential test for each of the four dependent measures). The results of the inferential tests are shown in Table 6.
For the Information score, there was a main effect for both maternal education $F(1,159) = 10.75$, $MS = 450.96$, $p = 0.001$, Cohen’s $d = 0.49$, $95\%$ CI [0.47, 0.52] and race $F(1,159) = 17.94$, $MS = 752.36$, $p < 0.001$, Cohen’s $d = 0.65$, $95\%$ CI [0.63, 0.68], but no interaction $F(1,159) = 0.48$, $MS = 20$, $p = 0.491$. Maternal education explained 6.33% of the variance, with $>$HS children ($M = 22.45$) producing an Information raw score that was significantly higher than the score achieved by the $\leq$HS children ($M = 19.10$). Race explained 9.51% of the variance, with EA children ($M = 23.01$) producing an Information raw score that was significantly higher than the score achieved by the AA children ($M = 18.67$). Recall that our power analysis indicated that only 3.86% of the variance needed to be explained in order to detect an experimental effect. As such, the power of our sample was more than adequate.

For the remaining three dependent measures, which did not meet the assumptions of ANOVA, there were main effects for maternal education, no main effects for race, and no significant interactions. The following results are reported as $X^2$ values instead of as $F$ values because ANOVAs conducted on ranked data conform to a non-central $X^2$ distribution (Thompson, 1991). For Sentence Length, the main effect for maternal education $X^2(1,159) = 7.72$, $MS = 16,348.99$, $p = 0.006$, Cohen’s $d = 0.47$, $95\%$ CI [0.45, 0.50] explained 4.63% of the variance, with the five longest sentences of $>$HS children ($M = 9.28$) having greater average length than the $\leq$HS children ($M = 8.28$). For Complexity, the main effect for maternal education $X^2 (1,159) = 9.17$, $MS = 17,951.20$, $p = 0.029$, Cohen’s $d = 0.48$, $95\%$ CI [0.46, 0.51] explained 5.45% of the variance, with $>$HS children ($M = 2.73$) producing more complex sentences than the $\leq$HS children ($M = 2.10$). For the Independence score, the main effect for maternal $X^2 (1,159) = 13.02$, $MS = 25,194.50$, $p = 0.004$, Cohen’s $d = 0.66$, $95\%$ CI [0.64, 0.69] explained 7.31% of the variance, with $>$HS children ($M = 41.32$) requiring less prompting than the $\leq$HS children ($M
Discussion

Maternal Education Effects

All four scores of the *RBS-NA* showed significant differences between children whose mothers had a lower education level (≤HS) as compared to a higher education level (>HS), with the children in the higher maternal education group scoring higher in all instances. These findings are supported by other research using spontaneous conversational language samples and standardized, norm-referenced tests. Dollaghan and her colleagues (1999), for example, analyzed the language samples and *Peabody Picture Vocabulary Test–Revised (PPVT-R)* (Dunn, 1980) scores of 240 three-year old children. Comparing three levels of maternal education, they found statistically significant linear trends across educational levels for their three spontaneous language sample measures (mean length of utterance in morphemes, the number of different words, and the total number of words) and on the standardized test of receptive vocabulary (*PPVT-R*).

The mean length of utterance measure used by Dollaghan et al. has been found to be very highly correlated with the mean length of a child’s five longest utterances (e.g., Malakoff, Mayes, Schottenfeld, & Howell, 1999), which was the Sentence Length measure used in the *RBS-NA*. As such, these two findings are logically connected and support each other (although the children were different ages and the language samples were collected differently). On a more general level, the *RBS-NA* and the *PPVT-R* used by Dollaghan and her colleagues are both standardized, norm-referenced language tests. At this level, the two sets of results are similar in finding maternal education effects. Qi and her colleagues (Qi, Kaiser, Milan, & Hancock, 2006) also found that among the 482 low-income AA preschoolers in their sample, maternal education
level had a significant impact on the children’s *PPVT-III* performance, with children’s whose mothers had a high school degree or less scoring 11 points lower than children whose mothers had a Bachelor’s degree. However, the *PPVT* and the *RBS-NA* each focus on very different aspects of language competence (oral narrative skill versus single word receptive vocabulary), which weakens support that results on one test provide for the other.

A standardized, norm-referenced language test more closely aligned with the *RBS-NA* (which requires the integration of many language skills in producing a narrative) is the *Preschool Language Scale-3, or PLS-3* (an omnibus test of language abilities). Qi et al., 2003, found significant effects of maternal education level on preschoolers’ performance on the *PLS-3* within low-income groups of both AA and EA preschoolers. The findings with the *PLS-3* support those obtained for the *RBS-NA* in the current study.

For research focused specifically on narratives, the findings of the current study are at odds with previous research. Recall that Reese and her colleagues (Reese et al., 2009) found no relationship between maternal education and three aspects of the storybook retell narratives in six and seven year olds. However, also recall that this study contained no mothers who were university graduates. This compressed range may have obscured any impact of education level on the children’s narrative language abilities. In the current study, 48 mothers (28% of our sample) had Bachelor’s or Master’s degrees. The greater spread in maternal education levels in our study may account for the differences in findings between our study and that of Reese et al. (2009). Furthermore, different story retell tasks were used in each study, so it is also possible that the different stories that were retold in these two studies might explain differences in the results.

In their study of preschoolers, Kulkofsky and Klemfuss (2008) found no main effects or interactions for maternal education, nor did maternal education correlate with narrative quality.
However, this study may have had too large a proportion of college graduates in the sample. Recall that in Study 1 (n = 46), 89% of the mothers had a college degree, and in Study 2 (n = 66), 72% did. Furthermore, both the type of narrative elicited and measures used in this study were quite different than those used in the RBS-NA, so it is difficult to compare the results obtained from them. The preschoolers’ retells in the Lulkofsky and Kelmfyss study were of a previously witnessed staged event, which was measured for overall quality and for the following four components of narrative quality: volume (number of propositions), complexity (number of words per proposition), descriptive texture (number of descriptive words), and cohesion (marking temporal, causal, conditional, & optional states). These do not correspond well with any of the RBS-NA measures.

In the Price et al. (2006) study of 65 African American children retelling the Bus Story at age 4 and again at kindergarten entry, it was likewise found that their methods of scoring the Bus Story (using a story grammar analysis yielding a maximum of 39 points) did not correlate with maternal education, which ranged from 9 to 18 years (M = 13.0; SD = 2.02). This does not align with the current findings, because we found a main effect for race, and maternal education level on the RBS-NA, but there was no significant interaction. If there had been an impact for maternal education level only for the EA children in the current study, we would have found a significant interaction. However, the lack of similar findings in these two studies could possibly be due to differences between the Information score as determined in the RBS-NA manual and the story grammar analysis these authors conducted on the Bus Story narratives. Although both are macro-level measures, they may be dissimilar enough to yield different findings. Furthermore, the pictures are quite different for the British and North American editions, with the British edition
featuring a double-decker bus that most children in the U.S. would be unfamiliar with. It is possible that this had an influence on the different findings of these two studies.

In summary, our findings that maternal education level impacted all measures of the \textit{RBS-NA} correspond well with research on spontaneous language samples produced by three-year olds that were not narratives (Dollaghan et al., 1999) and performance on other standardized language tests (Dollaghan et al., 1999; Qi et al., 2003; Qi et al., 2006), but they stand in contrast to research that has specifically analyzed narratives produced by children ages 4 to 7 (Kulkofsky & Klemfuss, 2008; Price et al., 2006; Reese et al., 2009).

\textbf{Race Effects}

Although there was a main effect for race for the Information score only, with AA children performing more poorly than their EA peers, the Information score is a particularly important one. When researchers report only one score from either version (British or American) of the \textit{Bus Story}, the Information score is almost always chosen (e.g., Bishop & Edmundson, 1987; Botting, et al., 1997; Conti-Ramsden & Botting, 1999; Conti-Ramsden, et al., 1997; Kovas, et al., 2005; Mackie & Dockrell, 2004; Paul, 1996; Paul, et al., 1996; Paul & Smith, 1993; Storch & Whitehurst, 2002; Stothard, et al., 1998). When two scores are reported, they are the Information score and the Sentence Length score, so here again, the Information score is of considerable importance in how either version of the \textit{Renfrew Bus Story} tends to be used in research (e.g., Anderson, et al., 2004; Anderson, et al., 1997; Briscoe, et al., 1998; Dockrell & Lindsay, 2001; Gallagher, et al., 2000; Girolametto, et al., 2001; Gray, 2004; Hohen & Stevenson, 1999; Hughes, et al., 1998; Kennedy, et al., 2006; Nathan, et al., 2004). Indeed, few studies report more than one or two of these measures. The current findings regarding the relatively strong effect of children’s race on their \textit{RBS-NA} Information score is as such a very
important one, at least for use of this language assessment tool in the United States, from which the AA sample in the current study was drawn.

These findings are not consistent with other data on the RBS-NA reported by Pankratz et al. (2007), who found that the small subset of minority children (mostly Hispanic) in their sample scored significantly lower on Sentence Length, but not on the Information score. However, given that these children were mostly Hispanic, it is not possible to compare their findings to those of the current study.

The current findings are also not consistent with a small study conducted by Serpell and his colleagues (2005). These researchers read a published children’s story, *The Monkey and the Crocodile* (Galdone, 1969) to four groups of AA and EA children second graders, two from low-income ($n = 27$ AA and $14$ EA) and two from middle-income ($n = 26$ AA and $14$ EA) backgrounds. To assess their comprehension of the story, the children were asked a number of literal and inferential questions tapping seven key features of the story. Finally, they were asked to retell the story. None of the groups differed on two of the three measures used — answers to the questions about the story, and the length of their retells in words. However, on story telling quality, as measured by the number of story structure elements included in the retell, the low-income AA children scored significantly higher than the middle-income AA children. Furthermore, the low-income children’s story telling quality scores were on average higher than, but not significantly different from, both groups of EA children. Serpell et al. noted “that low-income African American children exhibited oral narrative skills on a par with the middle-income European American children is a striking finding” (p. 163) and that these findings were “strikingly different from those obtained for most of the other competency measures” (p. 163).

The findings on the effect of race on children’s Information score in the current study do
Race, maternal education and Bus Story retells

dovetail with some, but not all, more general level with findings regarding other standardized, norm-referenced language tests. The effect size in the current study was $d = 0.65$, which means that the average information score performance of the AA children was 0.65 standard deviations below that of the EA children. This number is similar to the discrepancy between the performance of AA and EA children on other standardized, norm-referenced language tests focusing on other aspects of language development besides narrative ability.

For example, Brooks-Gunn and her colleagues (2003) report findings from two large scale studies on children’s performance on the *Peabody Picture Vocabulary Test – Revised, PPVT-R* (Dunn, 1980), a test of single-word receptive vocabulary. Data from 315 five year-olds from the Infant Health and Development Program (IHDP) study showed a white-black gap of 1.63 standard deviations. When adjusted for family income, maternal education, and home environment, the gap was reduced to 0.86 standard deviations. They reported additional data from 1,354 five and six-year olds from the National Longitudinal Survey of Youth (NLSY), which showed a white-black gap of 1.15 standard deviations. When adjusted for family income, maternal education, and home environment, this gap was reduced to 0.73 standard deviations, which is very close to the .65 standard deviations found in the current study.

Numerous subsequent studies of a new edition of this test, the *PPVT –III* (Dunn & Dunn, 1997), which some scholars believe may be less biased than the *PPVT-R* (Washington & Craig, 1999), also show that low-income AA children score on average substantially below the expected mean based on national norms, ranging from -0.60 to -1.48 SD from the mean (e.g., Campbell, Bell, & Keith, 2001; Champion, Hyter, McCabe, & Bland-Stewart, 2003; Qi et al., 2006; Restrepo, et al., 2006; Washington & Craig, 1999). And, there is evidence here again that maternal education explains part, but not all, of this discrepancy in the scores of AA children.
Race, maternal education and Bus Story retells

(Washington & Craig, 1999).

Although race differences have been found on various versions of the PPVT in some studies, the study by Qi et al., 2006, found no significant race differences when low-income AA children (n = 482) were compared to low-income EA preschoolers (n = 42) on the PPVT-III, the Expressive Vocabulary Test (Williams, 1997), or the PLS-3. The authors note that caution should be used in interpreting their results, since the sample sizes of AA and EA children were substantially different. With this caution in mind, we do not find support for the race differences found on the RBS-NA in the current study on these tests of receptive and receptive vocabulary and an omnibus test of language abilities.

In sum, the race difference on the Information score of the RBS-NA dovetails with some, but not all, previous research regarding the performance of AA children on standardized, norm-referenced language tests.

Future Research

There is clearly a need to sort out possible reasons for the discrepancy in scores between children of different races on the RBS-NA (and on other standardized language tests). One reason often suggested for standardized test score discrepancies and/or achievement discrepancies (often measured by standardized test scores) is that these children are so much more likely to live in poverty (e.g., Duncan, Yeung, Brooks-Gunn, & Smith, 1998). We attempted to control for this possibility in the current study by carefully equalizing the number of children from the two different races that were in each of the seven specific categories of maternal education that comprised our two levels of maternal education. Our results did indicate that children’s performance on all measures used in the RBS-NA, at least when using maternal education as a measure of socio-economic status, was clearly influenced by socio-economic status. However,
there was also a performance gap on the Information score on the RBS-NA based on race alone, and the effect size for it ($d = 0.65$) was larger than the effect size for maternal education ($d = 0.49$).

This general phenomenon of a test and/or achievement discrepancy between EA and AA children has been referred to as the “black-white achievement gap” (e.g., Craig & Washington, 2006), and other possible explanations for it in addition to poverty remain to be explored. Scholars have suggested a variety of additional factors such as the potential cultural bias of tests (e.g., Chamberlain & Madeiros-Landurand, 1991; Flowers, 2007; Thomas-Tate, Washington, Craig, & Packard, 2006; Wyatt, 2002), discrepancies in teacher quality (e.g., Connor, Son, Hindman, & Morrison, 2005; Talbert-Johnson, 2004; Wayne & Youngs, 2003), differences in home literacy practices for EA and AA preschoolers (e.g., Anderson-Yockel & Haynes, 1994; Vernon-Feagans, Hammer, Miccio, & Manlove, 2001), and mismatches for AA children between the language of the classroom (and of standardized language tests) and the language of home (see van Kleeck & Schwarz, under review).

**Limitations of the Current Study**

Before we are assured of maternal and racial difference in children’s performance on the Information score of the RBS-NA, it is essential that the current study be replicated. First of all, the children in the current study all took place in a mathematics intervention during the previous school year. It is possible, although we consider it unlikely, that the mathematics treatment differentially impacted the narrative retell skills of these kindergartners as a function of either race or maternal education.

An additional weakness stems from the fact that all of the children in this study were from urban classrooms in two cities in the Northeastern United States, and most of them were
from just one of those cities. Greater geographic diversity would be preferable. Another limitation was that there could be other unmeasured child, teacher, family, school, or even neighborhood characteristics that impacted the results. Also, we had information only on mothers’ education levels as a measure of socio-economic status in the current study. Other unmeasured aspects of a child’s SES background were factors such as family income and parental occupation. Further work that could tease out the independent contribution of these factors would provide a more detailed explanation for our findings. And finally, the current study compared only EA and AA children. It is essential to additionally determine how children from other non-mainstream backgrounds perform on the RBS-NA.

**Clinical Implications**

The findings of the current study suggest that clinicians should use caution in using the RBS-NA to determine the presence or absence of language impairment in AA children (at least for the Information score) and children from low-SES backgrounds (for all RBS-NA scores). Similar discrepancies in scores for African American children on other standardized language tests have lead to suggestions that spontaneous language sampling, supplemented with other informal language assessments, provide more valid methods for assessing these children’s language skills in an unbiased fashion (e.g., Craig & Washington, 2006).

Spontaneous language samples can be conversational, or they can be a narrative generated by the child. As such, we might have expected that children’s story retells, such as those generated in the RBS-NA, would produce results similar to spontaneous language samples and not show differences based on race alone. For the RBS-NA measures focused at the sentence level, Sentence Length and Complexity, this was true, as there were no significant differences on these measures based on a child’s race.
The Information score, however, showed a medium large ($d = 0.65$) effect size for race, and the importance of this score is highlighted by the fact that it is often the only RBS-NA score used in many research studies, as noted earlier. The Information score requires more than being able to generate language at the sentence level. As also noted earlier, it requires the integration of other skills, such as memory (of key words and concepts), adequate knowledge of the specific vocabulary in the story to be able to use it in a retell, and sufficient knowledge of story structure to convey the information in the proper sequence. So, the language differences found on the RBS-NA that were impacted by a child’s race appear to reflect differences in language ability beyond the sentence level. But this may not mean that AA children have a deficit in producing language beyond the sentence level. It may instead reflect that the narrative a child is required to produce in the RBS-NA is one that was not chosen by the child, and hence has no social importance to him or her, unlike narratives the child may be quite adept at producing.

More specifically, it may be that the RBS-NA reflects what Bloome and his colleagues (Bloome, Katz, & Champion, 2003) have referred to as “narrative as text.” Narrative as text is emphasized in school, and it tends to separate the existence of the storyteller from the story. Narrative as performance, on the other hand, explicitly builds social relationships and identities, but is of diminished importance in school. AA children are often skilled at narrative as performance, but may lack experience with narrative as text that is important in school.

The instructions for The Renfrew Bus Story reflect a more “school-like” task. The examiner says, “I’m going to tell you a story about this bus. Listen very carefully because when I’m finished, you’re going to tell me the story on this tape recorder. Are you ready?” To the extent that The Renfrew Bus Story emphasizes narrative as text as opposed to
narrative as performance, AA children might be expected to perform more poorly on it than their EA peers. Nonetheless, narrative as text is an important academic skill, so the \textit{RBS-NA} may be predicting future school difficulties for AA children, even if it does not reflect a general problem with narrative generation. Future research will need to distinguish these different types of narratives before the source of AA children's difficulty with the Information score on the \textit{RBS-NA} can be discerned.

The clinical implications of the findings in the current study for children whose mothers had lower education levels – they scored significantly lower on all \textit{RBS-NA} measures – was underscored by Dollaghan et al. (1999). These scholars concluded that there is a “critical need to examine the influences of maternal education and other sociodemographic variables on all norm-referenced tests and measures of spontaneous language production and comprehension, as a prerequisite to using such measures to compare children or to identify children with language disorders” (p. 1441). Our results indicate that the \textit{RBS-NA} will likely over-identify children from low SES backgrounds as having language impairments.

\textbf{Acknowledgments}

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant No. R305K05157 to the University at Buffalo, State University of New York, D. H. Clements and J. Sarama, “Scaling Up TRIAD: Teaching Early Mathematics for Understanding with Trajectories and Technologies.” The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education.
References


Race, maternal education and Bus Story retells


Race, maternal education and Bus Story retells 32


Table 1

*Frequency Counts of Children Who Received the TRIAD Treatment by Race and Maternal Education Categories and Sample*

*Selection Procedure for the Current Study*

<table>
<thead>
<tr>
<th>TRIAD Maternal Education Categories</th>
<th>8th Grade</th>
<th>High School (GED)</th>
<th>Vocational / Technical</th>
<th>1 to 2 years College</th>
<th>Associate’s Degree</th>
<th>Bachelor’s Degree</th>
<th>Master’s Degree</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>African American (AA)</td>
<td>26</td>
<td>96</td>
<td>8</td>
<td>77</td>
<td>28</td>
<td>27</td>
<td>5</td>
<td>267</td>
</tr>
<tr>
<td>European American (EA)</td>
<td>13</td>
<td>28</td>
<td>1</td>
<td>14</td>
<td>6</td>
<td>19</td>
<td>15</td>
<td>96</td>
</tr>
<tr>
<td>Total</td>
<td>39</td>
<td>124</td>
<td>9</td>
<td>91</td>
<td>34</td>
<td>46</td>
<td>20</td>
<td>363</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quasi-Random Sample Selection for the Current Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>African American (AA)</td>
</tr>
<tr>
<td>European American (EA)</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Re-categorization of Maternal Education Categories for the Current Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>High School or Less (≤ HS)</td>
</tr>
<tr>
<td>African American (AA)</td>
</tr>
<tr>
<td>European American (EA)</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>
Table 2

**Sample characteristics by group**

<table>
<thead>
<tr>
<th>Classification</th>
<th>Initial N</th>
<th>N Less Outliers</th>
<th>Boys</th>
<th>Boston$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA ≤ HS</td>
<td>41</td>
<td>40</td>
<td>21</td>
<td>7</td>
</tr>
<tr>
<td>AA &gt; HS</td>
<td>45</td>
<td>42</td>
<td>21</td>
<td>9</td>
</tr>
<tr>
<td>EA ≤ HS</td>
<td>41</td>
<td>39</td>
<td>18</td>
<td>5</td>
</tr>
<tr>
<td>EA &gt; HS</td>
<td>45</td>
<td>42</td>
<td>20</td>
<td>9</td>
</tr>
<tr>
<td>TOTAL</td>
<td>172</td>
<td>163</td>
<td>80</td>
<td>30</td>
</tr>
</tbody>
</table>

$^a$ = Data meeting our race and maternal education search criteria were collected at a site in Buffalo, New York and in Boston, Massachusetts.
Table 3

Analysis of gender differences on each dependent measure

<table>
<thead>
<tr>
<th>Variable</th>
<th>Female (n = 83)</th>
<th>Male (N = 80)</th>
<th>Test Statistic</th>
<th>Cohen’s d</th>
<th>Confidence Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information</td>
<td>21.11</td>
<td>7.74</td>
<td>20.54</td>
<td>6.15</td>
<td>0.57</td>
</tr>
<tr>
<td>Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sentence Length</td>
<td>8.95</td>
<td>2.32</td>
<td>8.63</td>
<td>1.99</td>
<td>3,063.50</td>
</tr>
<tr>
<td>Complexity</td>
<td>2.46</td>
<td>1.35</td>
<td>2.39</td>
<td>1.31</td>
<td>3,225.00</td>
</tr>
<tr>
<td>Independence</td>
<td>39.48</td>
<td>4.98</td>
<td>39.86</td>
<td>5.85</td>
<td>3,579.00</td>
</tr>
</tbody>
</table>

*p < .0125, **p < .001
Table 4

*Analysis of site differences on each dependent measure*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Buffalo (n=133)</th>
<th>Boston (n=30)</th>
<th>t-test values</th>
<th>Cohen’s d</th>
<th>Confidence Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information</td>
<td>20.77 6.81</td>
<td>21.07 7.83</td>
<td>-0.21</td>
<td>0.04</td>
<td>-0.12, 0.20</td>
</tr>
<tr>
<td>Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complexity</td>
<td>2.45 1.33</td>
<td>2.30 1.29</td>
<td>-0.11</td>
<td>-0.23</td>
<td>0.00, 0.00</td>
</tr>
<tr>
<td>Independence</td>
<td>39.52 5.53</td>
<td>40.33 4.89</td>
<td>0.15</td>
<td>0.00, 0.30</td>
<td></td>
</tr>
</tbody>
</table>

Two-Tailed t-test Results (t statistic) df = 161

Sentence Length 8.83 2.19 8.64 2.05 2,043.50 -0.09 -0.21, 0.04

Two-Tail Mann Whitney Test Results (U statistic)

Complexity 2.45 1.33 2.30 1.29 2,123.50 -0.11 -0.23, 0.00

Independence 39.52 5.53 40.33 4.89 1,813.50 0.15 0.00, 0.30

*p < .0125, **p < .001
Table 5


<table>
<thead>
<tr>
<th>Race</th>
<th>Maternal Education</th>
<th>Information</th>
<th>Sentence</th>
<th>Independence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Raw Score</td>
<td>Length</td>
<td>Complexity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(SD)</td>
<td>(SD)</td>
<td>(SD)</td>
</tr>
<tr>
<td>AA</td>
<td>High school or less</td>
<td>17.33</td>
<td>7.00</td>
<td>8.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(6.10)</td>
<td>(2.14)</td>
<td>(1.05)</td>
</tr>
<tr>
<td></td>
<td>More than high school</td>
<td>19.95</td>
<td>10.00</td>
<td>9.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(5.97)</td>
<td>(2.61)</td>
<td>(1.50)</td>
</tr>
<tr>
<td>EA</td>
<td>High school or less</td>
<td>20.92</td>
<td>10.00</td>
<td>8.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(6.13)</td>
<td>(1.51)</td>
<td>(1.22)</td>
</tr>
<tr>
<td></td>
<td>More than high school</td>
<td>24.95</td>
<td>11.00</td>
<td>9.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(7.54)</td>
<td>(2.02)</td>
<td>(1.31)</td>
</tr>
</tbody>
</table>

*Note:* SD = standard deviations; Information = total score for amount of key information used in the story retell. Sentence length = Average length in words of 5 longest utterances; Complexity = number of utterances with one or more subordinate clause; Independence = inverse of number of prompts needed to complete retell. AA = African American; EA = European American
Table 6

*Two-way fixed effect analysis of variance results*

<table>
<thead>
<tr>
<th></th>
<th>ANOVA Results using Raw Data</th>
<th></th>
<th>ANOVA Results using Ranked Data</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MS</td>
<td>( F ) (1,159)</td>
<td>Cohen’s ( d )</td>
<td>Confidence Intervals</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Information Score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maternal Education</td>
<td>450.96</td>
<td>10.75*</td>
<td>0.49</td>
<td>0.47, 0.52</td>
</tr>
<tr>
<td>Race</td>
<td>752.36</td>
<td>17.94*</td>
<td>0.65</td>
<td>0.63, 0.68</td>
</tr>
<tr>
<td>Maternal Education x Race</td>
<td>20.00</td>
<td>0.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sentence Length</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maternal Education</td>
<td>16,348.99</td>
<td>7.72*</td>
<td>0.47</td>
<td>0.45, 0.50</td>
</tr>
<tr>
<td>Race</td>
<td>5,974.36</td>
<td>2.82</td>
<td>0.14</td>
<td>0.11, 0.16</td>
</tr>
<tr>
<td>Maternal Education x Race</td>
<td>1,388.69</td>
<td>0.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Complexity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maternal Education</td>
<td>17,951.20</td>
<td>9.17*</td>
<td>0.48</td>
<td>0.46, 0.51</td>
</tr>
<tr>
<td>Race</td>
<td>8,246.40</td>
<td>4.21</td>
<td>0.26</td>
<td>0.24, 0.29</td>
</tr>
<tr>
<td>Maternal Education x Race</td>
<td>1,229.10</td>
<td>0.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Independence Score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maternal Education</td>
<td>25,194.50</td>
<td>13.02*</td>
<td>0.66</td>
<td>0.64, 0.69</td>
</tr>
<tr>
<td>Race</td>
<td>579.30</td>
<td>0.30</td>
<td>0.03</td>
<td>0.00, 0.05</td>
</tr>
<tr>
<td>Maternal Education x Race</td>
<td>10,594.40</td>
<td>5.47</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note: * = significant at the corrected alpha (Bonferroni) of 0.0125, \( a \) = the results are reported as \( X^2 \) because ANOVAs conducted on ranked data conform to a non-central \( X^2 \) distribution (Thompson, 1991).*