Metatheoretical Differences Between Running Records and Miscue Analysis: Implications for Analysis of Oral Reading Behaviors

Sinéad Harmey and Bobbie Kabuto

Sinéad Harmey, Ph. D, is a Lecturer in Literacy Education at the International Literacy Centre, University College London Institute of Education, and her research interests include young children's literacy development, early literacy assessment, and teacher professional development. Bobbie Kabuto, Ph.D., is an Associate Professor at Queens College, City University of New York and her research interests include bi/literacy and socially constructed identities.

Correspondence concerning this article should be addressed to Sinéad Harmey, International Literacy Centre, UCL Institute of Education. University College London, 20 Bedford Way, London WC1H 0AL, United Kingdom, Email: s.harmey@ucl.ac.uk

Abstract

The purpose of this article is to examine the metatheoretical differences that impact how running records and miscue analysis differ in (a) the quantification of readers' produced responses to text and (b) the analysis of oral reading behaviors. After providing historical and metatheoretical overviews of both procedures, we present the data source, which include 74 records of oral readings from an extant dataset collected from an informal reading inventory (IRI). Each record was coded using running record and miscue analysis procedures. We used inferential statistics to examine relationships across conceptually similar items of analysis (for example, the number of errors or miscues). Findings from the inferential statistics show that there were significant, positive correlations between three of the five conceptually similar items, and a lack of statistically significant correlations between the use of meaning and grammar between running records and miscue analysis. Based on the findings, we argue that both procedures, which are often confused and conflated, possess metatheoretical differences that influence how oral reading behaviors are interpreted. These differences, in turn, impact how reading ability is framed and socially constructed. We conclude with the significance of this research for educational professionals.

Metatheoretical Differences Between Running Records and Miscue Analysis: Implications

for Analysis of Oral Reading Behaviors

As children read in classrooms, it is common practice for teachers to observe and record their oral reading behaviors through two main modes of analysis: running records (Clay, 2000) and miscue analysis (Goodman, Watson, & Burke, 2005). The popularity and dominance of these procedures have resulted in them being incorporated into a variety of commercial assessment tools. Some tools draw upon running records, like the Fountas and Pinnell Benchmark Assessment System (Fountas & Pinnell, 2017) and the Teachers College Reading and Writing [TCRWP] Project General Running Records Assessments (TCRWP, 2014), while others draw on miscue analysis. Other assessments, like the Basic Reading Inventory (Johns, Elish-Piper, & Johns, 2017) and the Qualitative Reading Inventory-6 (Leslie & Caldwell, 2017) integrate a hybrid form of both procedures when evaluating oral reading behaviors.

Running record and miscue analysis assessment procedures are similar in three ways. First, oral reading behaviors are recorded and coded while readers read continuous text. Second, using standard conventions, teachers note substitutions, omissions, or insertions that readers produce. Third, teachers note oral reading behaviors like rereading or self-correcting.

Based on these similarities, some researchers (e.g. Goetze & Burkett, 2010) may assume that running records and miscue analysis can be used interchangeably because they attempt to measure and quantify the same construct: oral reading behaviors. In addition, there are researchers who suggest that the running record procedure is merely a simplification of miscue analysis (cf. Blaiklock, 2004). Although we do not disagree with the idea that both running record and miscue analysis procedures can be used effectively to code and analyze oral reading behaviors, we argue that there are fundamental differences in how oral reading behaviors are

quantified and how readers' produced responses, or errors (in the case of running records which focus on accuracy) or miscues (in the case of miscue analysis which focuses on acceptability) are analyzed. In fact, we contend that these differences exist at theoretical and conceptual levels and suggest that significant metatheoretical differences exist between the two procedures that are reflected both in quantification and analysis.

In studying how different tests (i.e. the Gray Oral Reading Test, the Qualitative Reading Inventory-3, the Woodcock-Johnson Passage Comprehension subtest, and the Peabody Individual Achievement Test Reading Comprehension test) measured and assessed comprehension, Keenan, Betjemann, and Olson (2008) found that the four tests did not measure the same skill and were not interchangeable in evaluating readers' reading comprehension abilities. Furthermore, Keenan et al. found greater variability among reading comprehension results when assessing children who were younger, novice readers. There is limited research (cf. Wilson, Martens, & Arya, 2005), however, along similar lines that investigates the analytic differences between oral reading measures and how different measures, such as running records and miscue analysis, compare in evaluating oral reading behaviors. Other studies in the area of comprehension (see Ukrainetz, 2017; Wixson, 2017) suggest caution when using one test or set of procedures to measure a construct that involves complex linguistic, cognitive, psychological, and social processes, such as oral reading behaviors.

As researchers argue, understanding how reading abilities are measured and how tests or measures are comparable with each other are critical for assessment practices, instruction, and research (Keenan et al., 2008; Pearson, Valencia, & Wixson, 2014). Based on the number and nature of errors or miscues, for instance, educators make important decisions about the books chosen for children to read, the reading groups children might be placed in, and the instructional

foci of reading lessons. Furthermore, educators or researchers may draw conclusions about children's reading abilities, positioning ability as unidirectional and singular, rather than multidimensional and situated depending on the texts being read and the social and cultural context that interprets children's oral reading behaviors (Pearson et al., 2014; Wixson, 2017; Ukrainetz, 2017). As Catts and Kamhi (2017) argued, while educators do not disagree that reading ability is dynamic, common educational practices, such as focusing on one type of oral reading evaluation tool or leveling books and readers, may define *reading ability* as a *single ability*. Therefore, analyzing how running record and miscue analysis procedures describe oral reading behaviors has important immediate and long-term consequences about how children's reading abilities are framed.

In this paper, we explore the metatheoretical and conceptual differences between running record and miscue analysis procedures by reanalyzing 74 records of oral readings from an extant dataset collected from the informal reading inventory, the Qualitative Reading Inventory-5 (QRI-5; Leslie & Caldwell, 2011). Metatheories are the underlying beliefs, assumptions, and ideologies that develop a particular approach to a field of study and involve the systematic investigation of the underlying structure of a theory or approach (Figueroa, 1994). Our use of the term *conceptual differences* indicates how the two approaches frame and define concepts, such as errors or miscues, related to oral reading. The purposes of this paper are to investigate how comparable running record and miscue analysis procedures are in evaluating oral reading behaviors and how both procedures interpret the process of oral reading. In the next section of this paper, we will provide a synopsis of the theory that informs each procedure, and we will explicate the metatheoretical differences between both procedures.

Running records

Marie Clay (2001), an influential scholar who coined the term *emergent literacy*, developed running records, which was informed by her literacy processing theory (see Doyle, 2013 for a full description). Clay's use of literacy processing theory incorporated a cognitive processing perspective of beginning reading that was built upon two theories: Rumelhart's (2013) interactive reading model, which posits that students *use* certain sources of knowledge or information (i.e. story knowledge or letter-sound information), and Holmes and Singer's (1961) notion that children employ problem-solving working systems as they read.

In 1968, as part of her doctoral dissertation that described the literacy development of 100 children during their first year of formal schooling, Clay developed and validated several measures (including running records) that could be used to document oral reading behaviors (see Ballantyne, 2009 for a full description). Clay (2001) characterized the running record as an 'unusual lens' that could be used to document changes in the literacy behaviors of young children (p. 42). It should be noted that running records were validated for younger children (see Clay, 2013). Indeed, Clay suggested that older children's reading behaviors were "too fast and too sophisticated for teachers to observe in real time" (Clay, p. 75).

Using a proforma and standard conventions, a teacher listens to and records a child's accurate reading and notes any errors, self-corrections, and observable reading behaviors like rereading or pausing. While commonly used to gauge a child's oral reading accuracy on a levelled or benchmarked text, Clay (2000) stated that "any texts can be used for running records – books, stories, information texts, children's published writing" (p. 8). When the oral reading is complete, the teacher counts the number of errors and calculates the percentage of text read accurately. This percentage is used to determine text difficulty or whether the book is too hard

(less than 89% accuracy), instructional (90 – 94% accuracy), or easy (95% accuracy or above). Self-corrections are also marked because Clay (2001) described self-corrections as an early reading behavior that decreases as readers become more proficient and signals a child is self-monitoring the reading process. There is no comprehension element in the original form of running records, as Clay argued that comprehension is "very dependent on the difficulty level of the text" (2000, p. 14).

The next level of analysis involves coding the sources of information readers used when they produced a word substitution as an error. Clay (2013) suggested that coders must note all sources of information used, which include (1) meaning or what sounds meaningful, (2) structure or what sounds grammatically correct, and (3) visual information which include letter, cluster, or word features. If readers self-corrected, coders then decide what additional sources of information led readers to self-correct the errors. Rodgers et al. (2016) described how, over time, emergent readers begin to control the integration and use of these sources of information and the problem-solving actions that they take.

Miscue Analysis

Similar to Marie Clay, Kenneth Goodman revolutionized the study of oral reading behaviors with his publication *Reading: A Psycholinguistic Guessing Game*. Originally published in 1967, *Reading: A Psycholinguistic Guessing Game* (Goodman, 2003) introduced the field of reading to the theoretical foundation that undergirds miscue analysis. Rather than use the term *error*, K. Goodman argued that readers produce *miscues*, or observed responses produced by readers that differ from the expected written text (K. Goodman, 1996). Observing readers read continuous, cohesive texts and studying readers' miscues while reading, K. Goodman argued that the reading process should also be viewed as a language process. The

reading process, thus, draws from linguistic, psycholinguistic, and sociolinguistic perspectives, and became known as a socio-psycholinguistic perspective to reading.

K. Goodman (1996) further explicated the reading process as composed of cycles and strategies. Psycholinguistic strategies involve the interaction between thought and language and include initiation strategies, sampling strategies, prediction strategies, confirmation strategies, and correction strategies (K. Goodman, 1996). Reading includes four cycles—the syntactic (grammatical), the semantic (meaning), visual (graphophonic), and perceptual—that allow readers to draw from a minimal amount of textual information as they are in the process of constructing meaning with written texts. These four cycles are also referred to as the linguistic cuing systems, and a socio-psycholinguistic perspective to reading suggest that readers sample, predict, and confirm their produced responses as they draw upon and integrate these linguistic cues.

Y. Goodman et al. (2005) presented a set of procedures for documenting, analyzing, and evaluating readers' miscues—a general procedure to be known miscue analysis. Y. Goodman et al. (2005) presented miscue analysis in the form of a Reading Miscue Inventory (RMI), which is composed of two main types of procedures for miscue analysis. The first is the classroom procedure and informal procedure. The second procedure is the in-depth procedure. The key difference between the two procedures is that the in-depth procedure allows coders to give partial acceptability for how miscues impact the syntactic and semantic structures of the sentences and the entire text. For this study, we used the classroom procedure, which is the most commonly used procedure and analyzes readers' miscue for syntactic and semantic acceptabilities at the sentence level without considering partial acceptability (further information will be provided in the methodology section).

Miscue analysis employs similar coding procedures as running records but also requires the elicitation of retellings after the oral readings. After the miscues are coded, coders calculate percentages for syntactic acceptability (or the percentage of sentences that are grammatically acceptable), semantic acceptability (or the percentage of sentences that make sense), and meaning change (or the percentage of meaningful sentences that affect the meaning of the sentence or entire text). Miscue analysis is based on readers reading a complete text with natural language that is, not only, unfamiliar to readers, but also challenging (for more information on text selection see Y. Goodman et al., 2005). Miscue analysis allows educators and researchers to listen to and observe as readers read, providing a window into the reading process (Y. Goodman et al., 2005).

Unlike running records, miscue analysis does not calculate accuracy or hard, instructional, and easy levels because of the focus on the quality of the miscues as high or low quality. Y. Goodman, Martens, and Flurkey (2014) discuss how high quality miscues, or miscues that are grammatically acceptable and make sense in the sentence and the text, illustrate how readers are effective in integrating reading strategies with linguistic cues to work at making sense of text. They, consequently, compare high quality miscues with low quality ones that disrupt the sense-making process inherent to reading.

Metatheoretical Differences

While there are similarities between running records and miscue analysis, we argue that there are metatheoretical differences that undergird the two procedures in the study and analysis of reading behaviors that impact how children's oral reading performances are interpreted. Studying metatheoretical differences allows researchers to focus on the epistemologies that underlie the oral reading constructs within both procedures. Table 1 provides a general overview

of the oral reading constructs, as well as the differences between those constructs, in terms of the

quantification and analysis of readers' produced responses. As Table 1 illustrates, both

procedures calculate the number of produced responses that differ from the written text. These

produced responses are considered *errors* when using running records, and *miscues* when using

miscue analysis. Where possible, we will use the term *produced response* to avoid privileging

either term, and the term *text* to describe what readers are reading.

[Table 1 about here]

Calculating the total number of produced responses and the produced responses per

hundred words is where the similarities end between running record and miscue analysis

procedures in the quantification of readers' produced responses. The use of the term *error* to

indicate produced responses premises that they are either correct or incorrect, an idea drawn

from literacy processing theory (Clay, 2001). With the use of the term *error* indicating a correct

or incorrect response, running record procedures calculate the percentage of the words read

accurately; data that is distinctive to running records.

The term *miscue*, however, finds its roots in the argument that the reading process is a

language process leading to the foundation of socio-psycholinguistic theory. For socio-

psycholinguistic theory, miscues are not about correctness; rather, they are lenses into how

readers employ linguistic cuing systems and psycholinguistic strategies when transacting with

texts. Therefore, there is a qualitative nature to miscues. Instead of using accuracy percentages,

miscue analysis calculates acceptability percentages. Acceptability percentages consider whether

the miscues are grammatically and semantically acceptable in the sentences and the texts.

Not only does the quantification of readers' produced responses reflect theoretical

differences, but also the questions that evaluators ask when coding readers' produced responses.

Table 1 presents the questions that guide the qualitative analysis of readers' produced responses. While the analysis for both procedures aims at ascertaining the types of information that readers draw upon when producing their responses to text, there are two striking differences between the procedures at a metatheoretical level. First, differences lie in how much of the context is considered when analyzing readers' produced responses. When analyzing errors in running records, evaluators only consider grammatical, meaning, and visual cues "up to the error (not the unread text)" (Clay, 2000, p. 22). If the error is the first word of the sentence, for instance, no other information is considered beyond that first word. The same is true when considering why readers' may self-correct their errors.

Miscue analysis, conversely, considers readers' miscues in the context of the entire sentence. When readers produce miscues, evaluators consider the quality of the miscue as they impact the grammar and meaning of the sentence. For instance, a reader may read the sentence, "I can *write*" as "I can *draw*." In this example, the substitution of *draw* for *write* is considered grammatically and semantically acceptable, and consequently, of high quality, which illustrates how the reader effectively predicted based on the meaning and the grammatical structure of the sentence. While running records would consider that the reader drew from meaning and syntactic information to produce the substitution, the procedure would discredit the substitution on the basis of accuracy.

Second, running records and miscue analysis position self-corrections differently. For running records, self-correcting is considered a useful monitoring strategy when readers produce errors as they indicate a movement from awareness to action on the part of the beginning reader (Clay, 1998). In other words, the reader has identified a mismatch between print on the page and his or her oral response and taken action to resolve the dissonance. Thus, if readers produce

errors and self-correct, then the self-corrections are not counted as errors in the final accuracy percentage. Miscue analysis, however, does not view self-corrections as a necessary strategy. Because miscue analysis examines acceptability, self-corrections are viewed as a self-monitoring strategy that readers, ideally, draw upon when meaning or grammar is disrupted. At the same time, miscue analysis does consider the self-corrected miscue as a miscue because self-corrections are based on a response to the written text that differs from the expected text. Unlike a running record, self-correction ratios are not accounted for in the final miscue statistics.

As our discussion illustrates, the differences between running record and miscue analysis procedures are more than procedural; they represent deeper metatheoretical differences (Tracey & Morrow, 2012). Clay (2001) alluded to these metatheoretical differences between running records and miscue analysis by writing:

> "the theories used to interpret the data are different. One theory, simplified for conciseness sees reading as a language process, which is basically the same at any level of performance; the other sees it as series complex neural processes which initially work together in simple systems and which gradually undertake increasingly complex activities…" (p.83).

We contend that the differences between both procedures impact the determination and evaluation of reading ability, which has important implications on instructional decision-making. Wilson et al. (2005) illustrated how three readers with similar running record accuracy scores were different types of readers when using miscue analysis. Wilson et al. argued that an accuracy score alone did not allow for exploring how readers used meaning and grammatical cues to create "comparable understandings of the texts they read or corresponding uses of strategies while reading" (p. 629). Flurkey (1997), similarly, argued that certain assessments present

reading from different perspectives that, when contrasted with miscue analysis, did little to view

readers as "thinkers who were grappling with achieving control over the conventions of

language" through "a linguistic strengths-oriented" process (p. 219). Finally, other researchers

(e.g. Kabuto, 2014; Wilson et al.) using miscue analysis procedures illustrated how readers may

have effective understandings of the texts in spite of the number of miscues they produced,

challenging the notion that accuracy is the main indicator of comprehension.

These studies suggest that reading ability is not "a single ability" (Catts & Kamhi, 2017).

Rather, reading ability is multidimensional, dynamic, and social as it is mediated by the tests and

assessments that educational professionals employ within school-based settings. In other words,

evaluation procedures, such as running records and miscue analysis, act as mediational tools in

the social construction of knowledge around what makes at-, above-, or below grade level

readers. Consequently, researchers have argued for studying the assessment tools themselves in

order to assist teachers in developing professional judgment in using school- and classroom-

based measures of reading performance (Vaughn, Linan-Thompson, & Hickman, 2003).

Methodology

The data and findings presented in this paper are from a larger empirical study that

investigated how the informal reading inventories (IRIs) the Qualitative Reading Inventory-5

(QRI-5; Leslie & Caldwell, 2011) and the Basic Reading Inventory (BRI; Johns et al., 2017)

measured and defined reading abilities. IRIs are criterion-based measures of oral reading

performances and comprehension. They include leveled word lists that are used as benchmarks

in determining the grade-level passages where readers should start. IRIs require readers to orally

read levelled passages, ranging from pre-primer to high school, and to answer follow-up

questions. As researchers argue, IRIs provide rough measures of reading levels (Christ &

Cramer, 2011), as well as are commonly used instruments for educational professionals (Nilsson, 2008).

While there are critiques and reviews of IRIs (e.g. Christ and Cramer, 2011; Nilsson, 2008; Walpole & McKenna, 2006), few have examined the integration of running records and miscue analysis within their general procedures. The QRI-5 and the BRI integrate a hybrid form of running record and miscue analysis procedures. For instance, the QRI-5 and BRI provide accuracy scores (cf. running records) and acceptability scores (cf. miscue analysis). While this may be the case, passage levels at the independent, instructional, and frustration are determined only by the accuracy scores, along with the comprehension score.

For purposes of this paper, we present the data and findings on the metatheoretical differences between running records and miscue analysis based on the readers' produced responses on passages from the extant dataset from QRI-5. Because running record procedures do not include a retelling for comprehension, we did not include a comprehension comparison for this analysis. We solely examined how running record and miscue analysis procedures frame readers' oral reading behaviors and address the following question:

How do running records and miscue analysis differ in (a) the quantification of readers' produced responses to text and (b) the analysis of oral reading behaviors?

Finally, we also consider how differences impact the framing of reading ability.

Sources of Data

We used running record and miscue analysis procedures to analyze 74 records of oral readings from passages from the QRI-5 collected from 51 students from kindergarten to 6th grade. In-service teachers, who were enrolled in a Master of Science program leading to state certification as literacy specialists in a large urban city in the United States, administered the

QRI-5 in their clinical experience in their final semester of the program. All candidates received training in the administration of the QRI-5 and were supervised by university faculty. Because teachers aimed to find the passage level at the students' instructional level, teachers may have administered the QRI-5 more than once to students.

There were a range in passage levels represented in the data, with the majority of passage levels representing second-grade at 23%. Fifteen percent of the total passages were at the third-grade level. Thirteen percent were first-grade level passages, 12% were fifth-grade and preprimer level passages, and 10% were fourth-grade level passages. The smallest percentage of passages were at the sixth-grade (7%), middle school and above (3%), and primer (4%) levels. The majority of passages in the dataset were fictional passages (86%). After the QRI-5 passages were collected, we reviewed how the teachers marked the readers' produced responses for accuracy using the QRI-5 standard coding procedures.

Analysis

The final data set contained 74 running records and 74 miscue analysis records. We conducted the following analyses. To ensure reliability, each data set was coded by two coders. The first coder had graduate level training in running record and miscue analysis procedures and used running records and miscue analysis in their classroom-based assessments. The first author has expertise in running record procedures and the second author has expertise in miscue analysis procedures, and each author acted as the second coder for each respective procedure.

Running Records. We analyzed the readers' produced responses for the 74 running records using standard running record procedures (Clay, 2000). Each coder analyzed the running records for the error rates, accuracy percentages, and self-correction ratios (see Table 1). If readers made word substitutions as produced responses, the coders noted the types information—

meaning, syntax, and visual—that readers used to make the substitutions. If readers self-corrected their produced responses, the coders noted the extra source of information—meaning, syntax, or visual—that led the reader to self-correct their produced responses. Finally, the coders provided a text difficulty level: easy, instructional, or hard, based upon the error rate.

Miscue Analysis. We analyzed the readers' produced responses for the 74 miscue analysis records using standard miscue analysis classroom procedures (Goodman et al., 2005; see Table 1). The coders coded the final produced sentences for the following:

1. Syntactic acceptability. The sentence was given a 'yes' if the reader's produced sentence was grammatically acceptable.

2. Semantic acceptability. The sentence was given a 'yes' if the reader's produced sentence made sense.

3. Meaning change. The sentence was given a 'no' if the reader's produced sentence did not change any significant aspect of the text. It was given a 'partial' if the produced sentence changes some significant aspect of the text. It was given a 'yes' if it did change a significant aspect of the text.

The coders calculated the percentage of sentences that were syntactically and semantically acceptable. The percentage of sentences that had no, partial, and some meaning change were also calculated.

Finally, the coders coded the readers' word-for-word substitutions for graphic similarity for (a) high graphic similarity, (b) some graphic similarity, or (c) no graphic similarity. After coding the substitutions, the percentages for high, some, and no graphic similarity were calculated.

Inter-Rater Agreement Percentages. The aforementioned data were imported into SPSS. We calculated inter-rater agreement through agreement percentages and Cohen's unweighted kappa scores as it was the most suitable for categorical variables (Landis & Koch, 1977). After the inter-rater agreement percentages and kappa scores were calculated between coder 1's and coder 2's scores for each data set, coder 1 and coder 2 reviewed the records to find agreement for the codes where disagreement was found. For the running records, inter-rater agreement was found for (1) the total number of errors, (2) the text difficulty level, (3) the number of self-corrections, (4) the accuracy percentage, (5) meaning information, (6) syntactic information, and (7) visual information.

For the miscue analysis records, inter-rater agreement was found for (1) syntactic acceptability, (2) semantic acceptability, (3) sentences with no meaning change, (4) sentences with partial meaning change, (5) sentences with meaning change, (6) word substitutions with high graphic similarity, (7) word substitutions with some graphic similarity, and (8) word substitutions with no graphic similarity.

Inferential Statistics. Comparisons were made across conceptually similar items (see Table 1) by conducting the following inferential statistics: bivariate correlations and dependent sample t-tests. We calculated the correlation between the total number of errors counted in the running records and the total number of miscues in the miscue analysis records. Focusing on the analysis of produced responses, we calculated if correlations existed between the use of meaning (running record) and semantic acceptability (miscue analysis). We did the same for other conceptually similar items: (1) the use of structure and syntactic acceptability and (2) use of visual information and high and some graphic similarity. Then, we reviewed the coded records to

provide exemplars of the metatheoretical differences between the two procedures, and to

consider how readers' reading abilities were framed differently according to each procedure.

Results

Inter-Rater Agreements

The inter-rater agreement percentages and kappa scores between coder 1 and coder 2 for

the running record quantification components showed fair to moderate agreement (Table 2),

according to the guidelines suggested by Landis and Koch (1977). Based on the kappa scores,

moderate agreement was found for the number of self-corrections ($\kappa = .56$), the total number of

errors ($\kappa = .42$), the text difficulty level ($\kappa = .59$), the accuracy percentage ($\kappa = .56$), and the use

of visual information ($\kappa = .47$). Finally, a fair scoring agreement was found for the use of

meaning information ($\kappa = .38$) and syntactic information ($\kappa = .36$).

The inter-rater agreement percentages and kappa scores between coder 1 and coder 2 for

the miscue analysis quantification components showed very good to good agreement. Based on

the kappa scores, very good agreement was found for syntactic acceptability ($\kappa = .98$), semantic

acceptability ($\kappa = .98$), and sentences with no meaning change ($\kappa = .85$). Good agreement was

found for sentences with partial meaning ($\kappa = .63$) and those with meaning change ($\kappa = .80$).

Good agreement was found when coding the graphic similarity of the word substitutions for high

($\kappa = .72$), some ($\kappa = .70$), and no ($\kappa = .73$) graphic similarity.

Comparing the inter-rater agreement percentages and kappa scores, better agreement was

found when using miscue analysis coding procedures over running record coding procedures.

After the inter-rater agreements were calculated, we reviewed the coding. Based on the review of

the initial codes for the running record quantification, we found that moderate agreement in the

total number of errors and accuracy percentages was due to miscounting the number of produced

responses, especially if the produced response was self-corrected because self-corrected

responses do not count as errors under standard running record procedures. Furthermore, the

kappa scores for the use of meaning and syntactic information showed fair agreement with 44%

and 43% inter-rater agreement percentages, respectively. These fair agreement percentages were

due to the challenge of working at the 'up to the point of error' analysis in determining the types

of information (meaning, syntactic, and visual) upon which readers drew. For instance, either

meaning or syntactic information may not have been included or there was disagreement on

whether or not information should be marked to indicate the type of information. There was

moderate agreement, however, when indicating whether visual information was used by the

reader.

The percentage agreements for the miscue analysis quantification components were

higher. The higher percentage agreements for syntactic and semantic acceptabilities at 98% each

reflect how miscue analysis does not consider individual miscues. Rather syntactic and semantic

acceptabilities consider whether the sentence is grammatically acceptable and makes sense

regardless of the presence of miscues. This part of miscue analysis differs from running records,

in which only errors are evaluated for the syntactic, meaning, and visual information.

While there was more agreement on the number of sentences that reflected no meaning

change, disagreement for partial meaning change and meaning change resulted from coders not

always having clear criteria that indexed whether meaningful sentences changed or partially

changed the meaning of text. A similar occurrence happened in the areas of disagreement for

high, some, and no graphic similarity between the produced word and the target word. Without

guidelines for what constituted high, some, and no graphic similarity, more areas of

disagreement were found.

Similarities Between Running Record and Miscue Analysis Procedures

Based on the dependent sample t-tests comparing conceptually similar items between running records and miscue analysis, we found similarities between three items: (1) the number of errors and the number of miscues, (2) the number of running record self-corrections and the number of miscue analysis self-corrections, and (3) the use of visual information and graphic similarity.

We did not find significant differences between the total number of errors in the running records and the total number of miscues in miscue analysis records. Specifically, the difference between the quantification of the mean number of errors ($M = 8.32, SD = 6.22$) and miscues ($M = 10.08, SD = 7.44$)($t_{(73)} = -3.48, p < .01$) per record was not significantly different and both items were correlated $r = .81, p < .05$ (see Table 3). Although there were theoretical differences between running record and miscue analysis procedures, this finding suggests that defining a produced response as an *error* or *miscue* did not have practical implications for the quantification of either procedure.

The mean number of self-corrections that were observed using running records ($M = 1.42, SD = 1.90$) and miscue analysis ($M = 1.24, SD = 1.65$)($t_{(73)} = -1.31, p < .01$) per record was not significantly different. Both items were correlated $r = .80, p < .05$. Like the concepts of errors and miscues, the concept of self-correction was used similarly across both procedures. [Table 2 about here]

In analyzing readers' produced responses, specifically what sources of information were used by the readers, the assessment procedures were similar only when raters considered whether readers used visual information for the running records ($M = 5.82, SD = 4.44$) and whether the miscues had any graphic similarity to the text ($M = 4.62, SD = 3.48$), ($t_{(73)} = -11.15, p < .01$).

Both items were correlated, $r = .85$, $p < .05$. This finding suggests that coders were likely to code the produced response similarly if they decided that the produced response was visually similar to the expected text.

Differences Between Running Record and Miscue Analysis Procedures

Based on the dependent sample t-tests comparing conceptually similar items in the analysis of produced responses, we found differences between two items: (1) the use of meaning in the running records and semantic acceptability, and (2) the use of structure in the running records and syntactic acceptability. We found that the analytic procedures differed in terms of how coders decided if readers used meaning or structure, and whether the miscues resulted in sentences that were semantically or syntactically acceptable. The mean rating for use of meaning (running record) was 4.76 ($SD = 3.72$) and the mean rating for semantically acceptable sentences (miscue analysis) was 22.45 ($SD = 9.15$) per record. The difference was statistically significant, $t_{(73)} = -15.65$, $p < .01$ (see Table 3). The correlation between both items was very low and non-significant ($r = .05$, ns). The mean rating for the use of structure (running record) was 4.22 ($SD = 4.44$) and the mean rating for syntactically acceptable sentences (miscue analysis) was 23.78 ($SD = 9.64$) per record. This difference was statistically significant, $t_{(73)} = 4.43$, $p < .01$. The correlation between both items was very low and non-significant ($r = .06$, ns). The results suggest that meaning/semantic acceptability and structure/syntactic acceptability did not have conceptually similar analytic foci.

<div align="center">

Discussion

</div>

In this paper, we suggest that there are not only procedural differences in evaluating oral reading behaviors through running record and miscue analysis procedures. We also argue that metatheoretical differences exist between the two procedures. One might hypothesize that if both

procedures could be used interchangeably, then there would be similarities across all conceptually similar items. We, in fact, did not find this to be the case. We found significant, positive correlations between three of the five conceptually similar items: the number of errors and number of miscues, the use of visual information and graphically similar substitutions, and the number of self-corrections. While self-corrections were positively correlated between both procedures, they are not calculated in the final miscue statistics for the classroom procedure. Furthermore, in the two components that are at the center of miscue analysis: syntactic and semantic acceptabilities, we did not find a statistically significant correlation with the use of meaning and structure in the running record counterpart. While the terms meaning/semantics and structure/syntactic appear to be comparable terms, these terms did not necessarily measure similar conceptual constructs across both procedures.

We argue that the differences we found in the measurement of conceptually similar items across both procedures signify metatheoretical differences that reflect a focus on accuracy from a literacy processing perspective when conducting running records and the focus on acceptability through a meaning construction perspective when conducting a miscue analysis. While there is a constructivist nature to both procedures as they center on using and integrating information to read, the quantifiable information of the running record procedure, unlike the miscue analysis procedure, does not reflect a 'full' text level analysis. Within running record procedures, analyzing produced responses up to the point of the error and requiring errors to be self-corrected are indicative of a cognitive processing model (Tracey & Morrow, 2012). Alternatively, developing the term '*emergent literacy,*' Clay (1968) focused on beginning readers and perhaps did not think it feasible that they would read on in a sentence if they made an uncorrected error, or that their produced responses were a result of their processing beyond the point of the error.

Indeed, Clay (2001) asserted that young readers would not be able to read on in a sentence if they made an error and encouraged coders to consider the produced response 'from the sentence so far' (Clay, 2000, p. 21). The lack of a comprehension component when conducting running records is additional support for the belief that a certain proportion of reading needs to be accurate in order to comprehend the text.

In contrast, miscue analysis procedures emphasize meaning construction from socio-psycholinguistic theory, which highlights the constructivist nature of reading as readers draw upon their social and background experiences of language to construct knowledge and understandings about what they read. Language is considered a process where readers may use information in any part of the sentence or text (including pictures) to construct meaning. The constructivist view that undergirds miscue analysis forefronts how readers actively construct meaning with texts (Y. Goodman et al., 2005). Therefore, miscue analysis considers all sentences in its coding regardless of the presence of a miscue, which deviates from running records, in which only errors are counted. Miscues, subsequently, are analyzed for whether or not they are grammatically acceptable and/or meaningful in the context of the sentence and text.

The inter-rater agreement scores within this analysis illustrate how analyzing up to the point of the error or in the context of the sentence provided dissimilar interpretations of readers' oral reading behaviors. When readers made substitutions that were visually similar in both procedures, coders were likely to code them in the same way. Differences occurred, however, when coders analyzed whether readers used meaning and structure for running records. The differences reflect how coders had to make judgments on the readers' intentions to infer what information the readers drew upon to make the errors. There were times when the coders did not agree on the readers' intentions. When coders had to consider semantic and syntactic

acceptabilities for miscue analysis, they were more likely to agree on which sentences were
semantically and syntactically acceptable.

Figure 1 illustrates how both procedures can provide competing perspectives of a reader's
oral reading behaviors. In Figure 1, the reader read a pre-primer passage *Just Like Mom* (Leslie
& Caldwell, 2011). The passage has a total of 44 words and 12 sentences. The reader made a
total of 5 errors and 1 self-correction (running records), and 6 miscues (miscue analysis) on lines
1, 4, 7, 9, and 11. Based on the running record quantification procedures (number of errors and
the number of words), this reader read with 89% accuracy, which puts this passage at the
frustration level. According, to the miscue analysis statistics, the reader read the passage with
92% of the sentences as syntactically and semantically acceptable (11 out of the 12 sentences). In
addition, 91% of the sentences that were deemed meaningful (10 out of 11) did not change any
significant aspect of the text; one sentence (line 7) resulted in meaning change. As socio-
psycholinguistic researchers argue (e.g. D. Goodman, Flurkey, & Y. Goodman, 2007), this
reader would be considered an effective beginning reader who integrates language cues with
reading strategies, including self-monitoring, to create meaningful texts. Thus, while running
record procedures constructed a profile of a reader reading at the hard level, miscue analysis
portrayed this reader as one who reads effectively.

To compare the coding procedures for running records, consider two examples: line 1 and
line 7. In line 1, the reader read the sentence, "I can write" as "I can work." The QRI-5 provides
minimal details in the pictures, which are black and white drawings. The picture on the page with
this sentence shows a woman and a young girl sitting next to each other at a table writing on a
piece of paper. Based on the running record, the reader was deemed to have used meaning,
structure, and visual information. The substitution of *write* for *work* makes sense and is

grammatically correct, and the words *write* and *work* begin with the same first letter. Using miscue analysis, the final produced sentence is considered to be semantically and syntactically acceptable, and the miscue is of high quality. The picture can be interpreted as the woman and the young girl writing to do work, like homework.

In line 7, the reader read the sentence, "I can work at home" as "I can water at home." The picture on the page shows a woman and a young girl watering flowers with a hose and watering can. Analyzing the information up to the point of error, it can be argued that the reader used meaning, structure, and visual information because "I can water" is meaningful and grammatically correct. Based on the miscue analysis, the sentence is syntactically and semantically acceptable, and changes some aspect of the sentence since working at home has a slightly different meaning than watering (e.g. flowers or plants) at home. Other differences in the coding procedures show that omissions in running records (lines 4, 9, and 11) are not coded for meaning, grammar, and visual information.

Our findings show that, on a metatheoretical level, the analysis up to the point of error in running records and the sentence-level analysis in the classroom procedure in miscue analysis may construct different notions of ability based on readers' oral reading performances and is consistent with other research. Wilson et al. (2005) compared the accuracy scores from the reading program Reading Mastery and the meaning change percentages from miscue analysis. Wilson et al. found that, while readers may read with a high percentage of accuracy, their miscues show a low percentage for meaning change suggesting that readers did not always make sense as they read. Kabuto (2017) demonstrated that the use of accuracy scores in the QRI-4 and acceptability scores in miscue analysis can construct two different abilities of the same reader.

More specifically, Kabuto illustrated how using acceptability percentages showed how readers could read and comprehend texts above their QRI-4 levels.

We add to this body of literature to suggest that assessment tools do not neutrally develop or interpret oral reading behaviors. Rather, there are theoretical constructs that frame not only oral reading assessment tools, but also how evaluators interpret reading behaviors and, subsequently, reading ability, through the employment of those tools. As this study and the aforementioned ones have illustrated, evaluators can create limited conclusions about readers' abilities based on one particular theoretical perspective or evaluation tool (Pearson et al., 2014). As Pearson et al. argue, when a variety of measures are not used to understand readers' abilities, we ignore the "complex nature of the reading process" and "artificially isolate components of the process and prevent ourselves from learning how readers marshal their abilities to meet the challenges of using a variety of texts for a variety of purposes" (p. 244). The result may be the development of inappropriate instruction to meet the diverse and varying needs of readers.

Limitations

This study compared running record and miscue analysis analytic procedures. It is important to note that we do not claim to provide a comprehensive comparison beyond the discussion of the quantification and analysis of both procedures. There are other significant differences between the two, such as the influence of text selections, the use of audio recordings, and the collection of retellings, that were not addressed and may impact how each procedure defines reading abilities.

In addition, there are limitations to this study. While we were able empirically to show and clarify how each procedure shifted the interpretation of oral reading performances, we conducted this study with a relatively small data set. Oral reading data across different written

text sources are being currently added to the overall data set for further comparative analyses. Furthermore, because running records do not include a retelling, we did not evaluate readers' comprehension as part of reading ability. This study specifically focuses on a comparison of oral reading behaviors, which do not always indicate effective reading comprehension. Miscue analysis standard procedures suggest a minimum of 25 miscues for analysis. Because we used passages from an extant dataset, not all passages included 25 miscues. Further research is warranted to address running record accuracy levels at easy, instructional, and hard in connection to miscue analysis meaning change percentages and readers' retellings of the texts to further compare the two procedures.

Finally, we compared running records with miscue analysis using the classroom procedure because it is the most commonly used and accessible procedure. Because the in-depth procedure examines partial syntactic and semantic acceptabilities within the sentence, comparisons between running records and the in-depth procedure would provide additional insights into the metatheoretical underpinning of both procedures.

Significance

Running records and miscue analysis offer lenses into observing and documenting readers' oral reading performances. With each shift in analytic procedures, however, the lens with which evaluators view oral reading behaviors changes, thus altering how reading ability is framed and, we argue, socially constructed. We do not posit that one procedure is better or worse. We, instead, suggest that teachers and researchers need to be aware of the theories that undergird each procedure and how, with subtle changes in an analytic procedure, readers' abilities might be constructed differently. Rather than using both procedures as tools to observe and document meaningful oral reading behaviors, arguably Clay's (1968) and Goodman's (1967)

original intention, these procedures are integrated into published assessment kits that attempt to categorize students based on a one-dimensional view of reading ability as below-, at- or above-grade level. These categories, in turn, create a classification of students into "orderable identities" (Foucault, 1966, p. 268; Ukrainetz, 2017), and can have long-term negative effects on students resulting in unequal access to reading instruction, being misplaced in programs, being misidentified for special education settings, receiving a diet of books that are deemed to be an 'appropriate' level (Kontovourki, 2012), or receiving inappropriate instruction.

There is also a trend towards conflating the discourses related to the two procedures both in empirical research (cf. Goetze & Burkett, 2010; Kragler & Martin, 2009) and commercial teacher resources (cf. Shea, 2000; TCRWP, 2014). For example, in an article by Goetze and Burkett, the authors framed the running record as the assessment tool and miscue analysis as the procedure used to analyze miscues. Similarly, in the TCRWP (2014) materials, used by New York State as an approved option for determining and assessing student learning objectives, errors are defined as miscues and comprehension questions are included as part of the procedure, a clear departure from Clay's (2000) intended use. IRIs, like the QRI-5 and the BRI, include accuracy and acceptability percentages in spite of framing the evaluation of oral reading behaviors and comprehension through a socio-psycholinguistic perspective using miscue analysis. In fact, passage levels are determined by accuracy percentages rather than acceptability percentages. The research presented here raises questions around the integration of running records and miscue analysis to create a hybrid form of the two procedures. Our findings illustrate that not all procedures used to evaluate oral reading behaviors are created equal or can be used interchangeably to define a reader's reading ability.

With the persistent interest in using running record and miscue analysis procedures to evaluate oral reading behaviors, it is critical for educational professionals to have an understanding of the theory that underpins each procedure. Indeed, as Clay (2013) stated: "avoid analyses for which you have no theoretical support" (p. 73). We suggest that, at present, there is a focus on the surface level use of each procedure with scant attention to how the theories that underlie each may construct reading abilities and identities, which require more attention and further research.

These findings have consequences for research and practice. First, we suggest that educational professionals should consider how assessment tools adapt either running record or miscue analysis procedures, as well as how readers' perceived reading abilities may shift when using another theoretical lens to examine their oral reading behaviors. Educational professionals should not rely upon any one procedure or assessment tool to analyze readers' oral reading behaviors. Second, we propose that educational professionals consider their own theoretical orientations toward reading and how those orientations play a role in selecting particular oral reading evaluation tools over others. Finally, educational professionals can advocate for readers by becoming aware of how oral reading evaluation tools will have consequences on how reading ability is framed. Through an advocacy stance, educational professionals should approach interpretations of oral reading behaviors tentatively, as well as informing themselves and others about how and why they evaluate oral reading behaviors.

References

Ballantyne, A. (2009). Research origins: The historical context (pp. 7-34). In B. Watson & B.

    Askew (Eds.), *Boundless horizons: Marie Clay's search for the possible in children's*

    literacy. Rosedale, New Zealand: Pearson Education.

Blaiklock, K. (2004). A critique of running records of children's oral reading. *New Zealand*

    Journal of Educational Studies, 39, 241 – 253.

Catts, H. W., & Kamhi, A. G. (2017). Prologue: Reading comprehension is not a single

    ability. *Language, Speech, and Hearing Services in Schools*, *48*, 73-76.

Christ, T. & Cramer, R. (2011). Assessing word recognition and fluency using an information

    reading inventory. *Michigan Reading Journal*, *43*, 7-23.

Clay, M.M. (1968). *Emergent reading behaviour.* Unpublished doctoral dissertation. University

    of Auckland.

Clay, M.M. (1998). *By different paths to common outcomes.* Portsmouth, NH: Heinemann.

Clay, M.M. (2000). *Running Records for classroom teachers.* Portsmouth, MA: Heinemann.

Clay, M.M. (2001). *Change over time in children's literacy development.* Portsmouth, MA:

    Heinemann.

Clay, M.M. (2013). *An observation survey of early literacy achievement.* Portsmouth, NH:

    Heinemann.

Doyle, M.A. (2013). Marie M. Clay's theoretical perspective: A literacy processing theory. In

    D.E. Alvermann, N. Unrau & R.B. Ruddell (Eds.), *Theoretical models and processes of*

    reading, sixth edition (pp.636-656). Newark, DE: International Reading Association.

Figueroa, E. (1994). *Sociolinguistic metatheories*. New York, NY: Pergamon.

Foucault, M. (1966). *The order of things: An archaeology of the human sciences*. New York,

NY: Routledge.

Flurkey, A. (1997). Inventing learning disabilities. In D. Taylor, D. Coughlin, & J. Marasco

(Eds.) *Teaching and Advocacy* (pp. 211-232). York, ME: Stenhouse.

Fountas, I & Pinnell., G.S. (2017). *Benchmark Reading Assessment*. Retrieved from

http://www.fountasandpinnell.com/research/default#bas

Goetze. S.G. & Burkett, J.A. (2010). Progress monitoring with whole text: A comparison study

of Running records and curriculum-based measures. In S. Szabo, M.B. Sampson, M.M.

Foote, M.M, & F. Falk-Ross (Eds.), *Mentoring literacy professionals: Continuing the

spirit of CRA/ALER after 50 Years* (pp. 295-316). Texas A&M University, Commerce,

TX: The College Reading Association.

Goodman, K. (1996). *On reading*. Portsmouth, NH: Heinemann.

Goodman, K. (2003). Reading: A psycholinguistic guessing game. In A. Flurkey & J. Xu (Eds.).

On the revolution of reading: The selected writings of Kenneth S. Goodman (pp.46-58).

Portsmouth, NH: Heinemann.

Goodman, D., Flurkey, A., & Goodman, Y. (2007). Effective young beginning readers. In Y.

Goodman & P. Martens (Eds.), *Critical issues in early literacy*. Mahwah, NJ: Lawrence

Erlbaum.

Goodman, Y., Martens, P., & Flurkey, A. (2014). *The essential RMA: A window into readers'

thinking*. Katonah, NY: Richard C. Owen.

Goodman, Y., Watson, D., & Burke, C. L. (2005). *Reading miscue inventory: From evaluation

to instruction* (2nd ed.). New York, NY: Richard C. Owen.

Holmes, J. A., Singer, H. (1961). *The substrata-factor theory: Substrata factor differences*

underlying reading ability in known-groups at the high school level. Berkeley, CA:

University of California.

Johns, J., Elish-Piper, L., & Johns, B. (2017). *Basic reading inventory: Kindergarten through*

grade twelve and early literacy assessments. Dubuque, IA: Kendall-Hunt

Kabuto, B. (2014). A semiotic perspective on reading picture books: The case of

Alexander and the Wind-Up Mouse. *Linguistics and Education, 25*, 12-24.

Kabuto, B. (2017). A socio-psycholinguistic perspective on biliteracy: The use of miscue

analysis as a culturally relevant assessment tool. *Reading Horizons*, *56*(1), 25-44.

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in

the skills they assess: Differential dependence on decoding and oral

comprehension. *Scientific Studies of Reading*, *12*, 281-300.

Kragler, S. & Martin, L. (2009). "I tried to make it not confusing by fixing it": Describing six

first graders' use of strategies to understand text. *Reading Psychology, 30*, 512 – 538.

Kontovourki, S. (2012). Reading Leveled Books in Assessment-Saturated Classrooms: A Close

Examination of Unmarked Processes of Assessment. *Reading Research Quarterly*, *47*,

153–171.

Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data.

Biometrics, 33, 159 – 174.

Leslie, L., & Caldwell, J. (2006). *Qualitative reading inventory-4*. New York, NY: Pearson.

Leslie, L., & Caldwell, J. (2011). *Qualitative reading inventory-5*. New York: Pearson.

Leslie, L., & Caldwell, J. (2017). *Qualitative reading inventory-6*. New York: Pearson.

Nilsson, N. L. (2008), A critical analysis of eight informal reading inventories. *The Reading*

Teacher, 61, 526–536.

Pearson, P.D., Valencia, S.W., & Wixson, K.K. (2014). Complicating the world of reading
    assessment: Toward better assessments for better teaching. *Theory into practice, 53*, 236 –
    246.

Rodgers, E., D'Agostino, J.V., Harmey, S., Kelly, R.H., & Brownfield, K.S. (2016). Exploring
    the nature of effective scaffolding. *Reading Research Quarterly, 51*, 345-360.

Rumelhart, D.E. (2013). Toward an interactive model of reading. In D.E. Alvermann, N. Unrau,
    & R.B. Ruddell (Eds.), *Theoretical models and processes of reading, Sixth Edition* (6th Ed,
    pp.636-656). Newark, DE: International Reading Association.

Shea, M. (2000). *Taking Running Records: A teacher shares her experience on how to take
    Running records and use what they tell you to assess and improve every child's reading.*
    New York, NY: Scholastic.

Teachers College Reading and Writing Project (2014). *Running records, Foundational
    Assessments and benchmarks.* Retrieved from
    http://readingandwritingproject.org/resources/assessments/running-records

Tracey, D. & Marrow, L.M. (2012). *Lenses on reading: An introduction to theories and models*.
    New York, NY: Guilford Press.

Ukrainetz, T. A. (2017). Commentary on "Reading comprehension is not a single ability":
    implications for child language intervention. *Language, Speech, and Hearing Services in
    Schools, 48*(2), 92-97.

Vaughn, S., Linan-Thompson, S., & Hickman, P. (2003). Response to intervention as a means of
    identifying students with reading/learning disabilities. *Exceptional Children*, 69, 391–
    409.

Walpole, S. and McKenna, M. C. (2006), The role of informal reading inventories in assessing

word recognition. *The Reading Teacher*, *59*, 592–594.

Wilson, P., Martens, P., & Arya, P. (2005). Accountability for reading and readers: What the numbers don't tell. *The Reading Teacher*, *58*, 622-631.

Wixson, K. K. (2017). An Interactive View of Reading Comprehension: Implications for Assessment. *Language, Speech, and Hearing Services in Schools*, *48*, 77-83.

Table 1.

Running Record (Clay, 2000) and Miscue Analysis (Goodman, Watson, & Burke, 2005): Similarities and Differences in Quantification and Analysis of Oral Reading Behaviors

	Running records	Miscue Analysis
Quantification		
Total number of errors/ miscues	Total number of errors.	Total number of miscues.
Total number of self-corrections	Total number of self-corrections. These do not count as errors.	Total number of self-corrections can be calculated. Self-corrections are considered miscues when calculating the total number of miscues.
Percentage of text read accurately	Percentage Accuracy: Formula: $$100 - \frac{Errors}{Running\ Words} \times \frac{100}{1}$$	Not calculated
Self-correction ratio	Formula: $$\frac{Number\ of\ self\text{-}corrections}{Errors + Self\text{-}corrections}$$	Not calculated
Analysis		
Error/Miscue Analysis: For each produced response the rater considers the following criteria		
Meaning/Semantic	Up to the point of the error does the error make sense or did the reader use meaning?	In the context of the sentence is the miscue semantically acceptable? Did the miscue change a significant aspect of the sentence or text?
Structure/Syntax	Up to the point of the error does the error sound right or is it syntactically acceptable? Did the reader use structure?	In the context of the sentence is the miscue syntactically acceptable?
Visual/ Graphophonics	Is the error visually similar in terms of letter-sound information? Did the reader use visual information?	Does the miscue have high graphic similarity? Does the miscue have some graphic similarity? Does the miscue have no graphic similarity?
Self-correction	First, what source of information did the child use when he/she made the error. Next, what extra piece of information helped the child to self-correct?	Total number of self-corrections can be calculated.
Comprehension	No comprehension element	Oral retell is elicited.

Table 2.

Percentage of Scoring Agreement and Kappa between the Two Raters

	Percentage of Scoring Agreement	Kappa
Running Records		
Total Number of Errors	46%	.42*
Running Record Level	87%	.59*
Number of Self-Corrections	78%	.68*
Accuracy Percentage	62%	.56*
Meaning Information	44%	.38*
Syntactic Information	43%	.36*
Visual Information	51%	.47*
Miscue Analysis		
Syntactic Acceptability	98%	.98*
Semantic Acceptability	98%	.98*
Sentences with No Meaning Change	86%	.85*
Sentences with Partial Meaning Change	84%	.63*
Sentences with Meaning Change	94%	.80*
Substitutions with High Graphic Similarity	75%	.72*
Substitutions with Some Graphic Similarity	79%	.70*
Substitutions with No Graphic Similarity	82%	.73*

*Note: p < .001

Table 3

Results of Dependent Sample T-tests Comparing Mean Quantified Analyses of Running records (N = 74) With Miscue Analysis (N = 74)

Construct	Running record M	SD	Miscue Analysis M	SD	n	95% CI for Mean Difference	r	t	df
Number of errors/ miscues	8.32	6.22	10.08	7.47	74	-2.76, -.75	.81*	-3.48	73
Used meaning/ Sentence semantically acceptable	4.76	3.72	22.45	9.15	74	-19.94, -15.44	.05	-15.65	73
Used Structure/ Sentence syntactically acceptable	4.22	3.49	23.78	9.64	74	-21.89, -17.24	.06	4.43	73
Used Visual Information/ Graphically Similar	5.82	4.44	4.62	3.48	74	.66, 1.74	.85*	11.15	73
Self-corrections	1.42	1.90	1.24	1.65	74	-.09, .44	.80*	1.31	73

*Note: $p < .05$

Figure 1. A Comparison of Running Record and Miscue Analysis Procedures for a Reader Reading *Just Like Mom* (Leslie & Caldwell, 2011).

Line	Text*	Reader's Produced Text	Running Record Error Analysis	Miscue Analysis
1	I can write.	I can <u>work</u>.	Used meaning (I can work) Used structure (I can work) Used visual information (same first letter work/write).	Syntactically acceptable Semantically acceptable. High graphic similarity. Miscue did not change meaning.
2	Just like Mom.	Just like Mom.		Syntactically and semantically acceptable.
3	I can read.	I can read.		Syntactically and semantically acceptable.
4	Just like Mom.	Just like <u>a (self-corrected to)</u> Mom.	Used meaning (Just like a) Used structure (Just like a) Did not use visual information (a/Mom). Used visual information in the word Mom to self-correct.	Syntactically acceptable Semantically acceptable. Miscue did not change meaning. No graphic similarity.
5	I can go to work.	I can go to work.		Syntactically and semantically acceptable.
6	Just like Mom	Just like Mom.		Syntactically and semantically acceptable.
7	I can work at home.	I can <u>water</u> at home.	Used meaning (I can water) Used structure (I can water) Used visual information (same first letter water/work).	Syntactically acceptable Semantically acceptable. High graphic similarity. Miscue changed meaning of the sentence.
8	Just like Mom.	Just like Mom.		Syntactically and semantically acceptable.
9	I can work with numbers.	I can work numbers. (Omitted with)	No attempt unable to analyze.	Syntactically and semantically acceptable.
10	Just like Mom.	Just like Mom.		Syntactically and semantically acceptable.
11	I can do lots of things.	I can do of <u>teachers</u>. (Omitted lots)	Error 1: (omission of lots) No attempt unable to analyze. Error 2: (substituted teachers/things) Did not use meaning (I can do of teachers) Did not use structure (I can do of teachers) Used visual information (same first letter teacher/things)	Not semantically acceptable. Not syntactically acceptable. Some graphic similarity.
12	Just like Mom.	Just like Mom.		Syntactically and semantically acceptable.
	Analytic Summary		2 errors used meaning, structure, and visual information. 1 error used meaning and structure information. 1 self-correction using visual information. Summary: Up to the point of error, errors made sense and sounded right and often had same first letter.	11 out of the 12 sentences were syntactically and semantically acceptable. Reader made 3 high quality miscues (lines 1, 4, and 9). One miscue changed meaning of the sentence. 3 miscues had some graphic similarity.
	Quantification		5 errors 1 self-correction 44 words 89% Accuracy: Frustration level 1: 6 self-correction ratio	6 miscues 12 sentences 92% semantically acceptable sentences. 92% syntactically acceptable sentences.

			91% of meaningful sentences did not change a significant aspect of the sentence or text. 50% of word substitutions have high graphic similarity. 25% of word substitutions have some graphic similarity. 25% of word substitutions have no graphic similarity.

From Leslie, L., & Caldwell, J. (2011). Qualitative reading inventory-5. New York: Pearson.
*** Word substitutions are underlined.*