

Does DIBELS Put Reading First?

NANCY RANKIE SHELTON

University of Maryland Baltimore County, Baltimore, Maryland

BESS ALTWERGER

Towson University, Towson, Maryland

NANCY JORDAN

Washington, DC

Schools across the nation are responding to reading assessment mandates by turning to DIBELS (Dynamic Indicators of Basic Early Literacy Skills), a program endorsed by the Federal Reading First Grant Office, to assess fluency on a range of reading-related tasks. DIBELS makes strong claims regarding its usefulness in documenting reading progress and predicting success on standardized tests and is frequently used to group and level students for intervention and instruction. This article addresses the effectiveness of using DIBELS to assess second-grade readers and presents an analysis of the relationship between DIBELS Oral Reading Fluency (DORF) and authentic reading. Our data show no connection between DORF and students' oral reading fluency and comprehension of authentic literature and questions the usefulness of DIBELS.

Keywords comprehension, DIBELS, fluency, miscue analysis, reading assessment, reading first, reading rate

Reading fluency has been a subject of research and debate for some time. However, it has risen to a much higher level of prominence in classroom instruction and assessment since fluency was incorporated into the Reading First guidelines of No Child Left Behind (NCLB) in 2002. Given that Reading First's focus on fluency rests heavily on the findings of the National Reading Panel (2000), it is important to note that the NRP *assumes*, but does not *establish* (with scientific evidence) a firm relationship between fluency and comprehension or overall reading proficiency. In fact, the report presents research findings only on studies that examine possible links of various instructional strategies to increased fluency, defined as "the ability to read a text quickly, accurately, and with proper expression" (2000, p. 3–5). Nevertheless, Reading First guidelines pronounce fluency instruction and assessment as essential and stipulate that funding proposals must reflect a major focus on fluency instruction and assessment.

The NRP report suggests a variety of assessment procedures that may be used to index fluency, including informal reading inventories, miscue analysis, pausing indices, running records, and reading speed calculations. In spite of an effort by the authors of the NRP to encourage the practice of rating fluency while readers are attending to comprehension, it is still common to see instructional grouping practices based on comparison of students'

accurate reading rates to some established grade level benchmark that does not incorporate reading comprehension (Mathson, Allington, & Solis, 2006).

Federal funds granted to states through Reading First must be spent only on those programs deemed to be based on "Scientifically Based Reading Research" and which directly address phonemic awareness, phonics, fluency, vocabulary, and comprehension. One approved program that has been widely adopted as a means of assessing fluency is the Dynamic Indicators of Basic Early Literacy, or DIBELS (Good & Kaminski, 2002). The battery of subtests within the DIBELS assessment program attends to discrete skills such as phonemic awareness and phonics, as well as to oral reading fluency. The only measure of comprehension in the DIBELS battery is Retelling Fluency (RTF), embedded in the Oral Reading Fluency subtest. The RTF measures comprehension by quantifying the number of words spoken by students in one minute of retelling after completing the oral reading of test passages. This retelling score, which is not subject to any qualitative analysis of content beyond detection of repetitions and off-topic comments, is used only to screen and dismiss an "invalid" oral reading score. An oral reading fluency score is validated if the student's RTF is at least 50% of the DORF (explained in more detail later).

Many researchers and classroom teachers alike have raised concerns regarding the use of DIBELS to assess children's reading ability. Central to these concerns is the questionable validity of an instrument that relies solely on a one-minute measure of reading accuracy and speed as the means for identifying a reader's level of reading "risk." Further, there are questions regarding whether such an assessment can reliably predict children's ability to read and comprehend non-test reading material—authentic texts (Allington, 2007; Altwater, Jordan, & Shelton, 2007; Carlisle, Schilling, Scott, & Zeng, 2004; Goodman, 2006; Kamii & Manning, 2005; Pressley 2006, Pressley, Hilden, & Shankland, 2005; Reidel, 2007; Samuels, 2006a, 2006b, 2007; Schilling, Carlisle, Scott, & Zeng, 2007). These are not just academic questions; DIBELS has a significant impact on children's educational progress, teachers' time and instruction, and schools' use of human and economic resources. Skepticism toward the wide scale endorsement of DIBELS has mounted recently, with release of the Inspector General's reports on federal Reading First policies and procedures (Office of Inspector General United States Department of Education 2006, 2007). The investigation disclosed that the National Institute for Literacy (NIFL) violated the Department of Education Organization Act (DEOA) as well as curriculum provisions in NCLB by showing obvious bias toward the DIBELS Assessment Test. The reports revealed serious and potentially indictable misconduct involving instructional bias and conflicts of interest in the grant approval process. It was revealed that Reading First officials and consultants with financial ties to DIBELS played key roles in biasing the approval process in favor of proposals that specified the adoption of DIBELS and that they contracted with consultants who financially profited from implementation of the program. The reports revealed that the proposals of many states were turned down until they gave in to the "encouragement" of Reading First officials and consultants to include DIBELS in their proposal revisions. Given the revelations in the Inspector General's reports, it is increasingly important that reading researchers and educators more thoroughly examine DIBELS and its effectiveness as an assessment program.

According to its authors, DIBELS is a valid indicator of reading ability, and can be used to "(a) identify children who may need additional support, and (b) monitor progress toward instructional goals" (p. 30). The DIBELS manual offers the following scientific justification for its assessment of oral reading fluency:

DIBELS Oral Reading Fluency (DORF) is a standardized, individually administered test of accuracy and fluency with connected text. The DORF passages

and procedures are based on the program of research and development of Curriculum-Based Measurement [CBM] of reading by Stan Deno and colleagues at the University of Minnesota and using [sig] the procedures described in Shin (1989). A version of CBM Reading also has been published as The Test of Reading Fluency (TORF) (Children's Educational Services, 1987). . . . A series of studies has confirmed the technical adequacy of CBM Reading procedures in general. Test reliabilities for elementary students ranged from .92 to .97; alternate-form reliability of different reading passages drawn from the same level ranged from .89 to .94 (Tindal, Marston, & Deno, 1982). Criterion-related validity studied in eight separate studies in the 1980s reported coefficients ranging from .52-.91 (Good & Jefferson, 1998). (Good & Kaminski, 2002, p. 30)

The *Administration and Scoring Guide* offers no additional sources of research support, and while the manual does list 19 references, none of these can be confirmed as independent peer-reviewed research studies.

The DIBELS website lists 89 resources under "DIBELS Research References" (available at http://dibels.uoregon.edu/techreports/DIBELS_References.pdf; retrieved October 2006). Although the quantity of resources is impressive, the strength of the research base for DIBELS weakens under closer scrutiny: Included in the list of "resources" is the DIBELS manual, an OSSC Bulletin report, and two reviews, all of which were published by the University of Oregon and stretch at least as far back as 1988 (unpublished works cited are undated). Thirty-nine of the references are unpublished master's theses (3) or doctoral dissertations (36), and another 13 are technical reports. Four of the references are listed as "submitted" with no information regarding the target publication. One article "in preparation" and one conference poster presentation are also listed. Of the 89 references listed, only 18 are published in professional refereed journals in the fields of psychology, special education, or music therapy, and 8 are chapters in edited books. One article was published in the *Scientific Studies of Reading*.

Of the 19 publications in the DIBELS research base meeting the criteria of peer review as specified by the National Reading Panel, only 3 are studies involving regularly achieving students above first grade. Two of these three studies used DIBELS as an assessment protocol in the research design and did not specifically examine the efficacy of DIBELS itself. A final study, Kamps et al. (2003), did find DIBELS to be a reliable assessment of Letter Naming, Nonsense Word Fluency, and DIBELS Oral Reading Fluency (DORF). However, no measures of reading comprehension were addressed in this study. Furthermore, use of DIBELS did not help to close the achievement gap between higher and lower achieving students within each instruction program evaluated. A significant finding of this study was that "at the end of third grade [high-risk students] fell below all end-of-grade benchmarks and had not caught up to their lower risk peers" (p. 219).

In light of this examination of the DIBELS research base (as well as those of Wilde, 2006; Manning, Kamii, & Kato, 2006), it is fair to conclude that DIBELS' "scientific evidence" is insufficient, over-stated, and unable to meet the criteria set forth in NCLB for independent, rigorous, scientifically based reading research. It is therefore imperative that *independent* researchers conduct and report studies that investigate the effect DIBELS has on teaching practices and student learning. Initial reports are not favorable. Carlisle and colleagues (2004) found that DIBELS was unable to accurately predict the learning trajectory of over 40% of the students tested. Kamii and Manning (2005) found no evidence to justify the use of the DIBELS in evaluating literacy instructional programs. Pressley et al.'s research (2005) found that neither DIBELS Oral Reading Fluency (DORF) nor DIBELS

Retelling Fluency (RTF) have predictive value in *TerraNova* scores. Schilling and colleagues (2007) report 37% false negatives in DIBELS testing. And Reidel's study (2007) cautions against the use of two DIBELS specific subtests—Phoneme Segmentation Fluency (PSF) and Nonsense Word Fluency (NWF). In addition to these findings, it is crucial to investigate whether DIBELS, is, as the program authors claim, a useful tool for predicting students' actual ability to read and comprehend authentic reading material. The study described next was designed to address this need.

The Study

For the purposes of this study we looked inside a typical classroom to examine the real life consequences of using DIBELS' Oral Reading Fluency subtest for both children and teachers. According to DIBELS' authors, a student's oral reading fluency score (DORF) is a valid indicator of his/her reading ability, and can be used to "(a) identify children who may need additional support, and (b) monitor progress toward instructional goals" (p. 30). Our goal, then, was to examine whether the DORF scores for a group of students in this classroom provided valuable and dependable information regarding their actual reading and comprehension of grade appropriate authentic literature, or "real" books. If DORF findings are reflective of students' reading of "real" books, then the use of DORF to determine their placement and advancement may be justified. If, however, DIBELS results are not consistent with students' reading of authentic reading material, then the time and cost of DIBELS and its usefulness for students and teachers remains questionable. More specifically, the study was designed to address the following questions:

1. Do students' DORF scores predict the rate and accuracy with which they read authentic children's books?
2. Do DORF instructional recommendations predict students' ability to comprehend authentic children's books?

To address these questions we looked closely at 14 readers enrolled in a second grade classroom in a public elementary school in the mid-Atlantic region of the United States. The school where this research took place makes classroom placement decisions based on students' reading proficiency as determined by their performance at the end of grade evaluations. The school relies heavily on the DIBELS assessment for these placement decisions, both at the classroom level and within the classroom, regrouping students for instruction after each formal DIBELS assessment (fall, winter, and spring).

The students we studied were identified as the highest performing group of second graders in this school. In all, the data pool consisted of 14 DIBELS assessments and 37 literature assessments: each student received the DIBELS assessments, and read two or three children's books. All data was collected at the end of the participants' second-grade year. The DORF and RTF scores were collected by the classroom teacher who was trained by her school and followed the DIBELS protocol. The literature assessment was conducted by a team of researchers.¹ For each of the literature books the students read, a complete miscue analysis was conducted that included standard retelling protocol (Goodman, Watson, & Burke, 2005). Each reading event was audiotaped and all miscues (oral reading "errors") were marked on copies of the stories. Miscues were then qualitatively analyzed to determine their graphic and sound similarity to the text words, and their syntactic

¹ We thank Prisca Martens, Poonum Ayra, G. Pat Wilson, and Lijun Jin for their participation in collecting and analyzing the miscue data for this project.

(grammatical) and semantic (meaning) acceptability. Although it is common practice to analyze data from the mid 25 miscues, for this research the data was drawn from the entire reading. This analysis provided us with information regarding readers' use of strategies and meaning construction while reading, indicating the relative proficiency with which they read the stories.

Participants' instructional levels were determined by the classroom teacher based on DIBELS assessment and their Terra Nova scores. As noted earlier, this group of students is considered to be the highest performers at the second-grade level. Within the classroom, the students were further divided into two groups, the "high" group and the "low" group. This grouping was based on the mid-year DIBELS assessment. Using the book gradient developed by Fountas and Pinnell (2000) literature was selected for each group of readers that was appropriate for their instructional levels. Students assigned to the "high" group each read three literature books for this study: *Cherries and Cherry Pits* (Williams, 1991); *Flossie and the Fox* (McKissack, 1986); and *Precious and the Boo Hag* (McKissack & Moss, 2005). The students in the "low" group read two books: *A Letter to Amy* (Keats, 1998) and *See You Tomorrow, Charles* (Cohen, 1983). As the books were introduced to the students, the researchers asked each student if he/she was familiar with any of the stories. None of the students in this study reported prior exposure to the books, and the classroom teachers confirmed that the books had not been used or available in the classroom. Before the reading began, it was explained to each reader that he/she would be asked to tell the researcher "everything you can remember about the book." During the reading, researchers made notes on the story typescripts that were later completely coded for miscues using the audiotapes of the readings. An unaided and aided retelling occurred immediately following each reading and was also audiotaped and analyzed. Numerical values were given to the reader's recall of characters, setting, events in the story, theme, and inferences, indicating their overall comprehension of the stories.

The audiotapes of the literature reading were used to compute the number of Words Correct per Minute (WCPM), which is the measure used for DORF. Researchers applied the same rules to the oral reading data to derive WCPM (for example, allowing only three seconds pause in reading; using the first minute of reading only). Any additional time spent by readers in reviewing illustrations in the literature was not included in WCPM computations. To establish interrater reliability two researchers independently timed the readings and all discrepancies were resolved by a third and joint timing. This procedure provided an excellent basis for comparison of reading literature to DORF scores.

Standard procedure for administering DORF is explicit. Second-grade students are assessed at least three times a year, more if they do not meet benchmark scores. The assessor follows a scripted procedure that allows the student one minute of timed reading. The student is presented with a probe and told to read until the assessor says stop. The script informs the reader that the assessor may ask what the passage is about and the student is encouraged to "do your best reading." Readers who pause during reading are told the next word after three seconds so that reading can continue. Following the reading, the administrator records only the number of correct words read by the student, excluding all mispronunciations, omitted words, abbreviations read incorrectly or words read out of order. The student is allowed three seconds to self-correct before the administrator provides the word, and "the student is not penalized for imperfect pronunciation due to dialect, articulation, or second language interference" (p. 34).

DIBELS provides "Descriptive Levels of Performance in End of Second Grade" DORF scores (Good & Kaminski, 2002, p. 64), which specify levels of reading "risk" for students' at each grade level on the basis of standardized test performance. DIBELS norms

were based on standardized test and DIBELS scores that were drawn from 3,758 students, in 31 participating districts, and 79 participating schools. The Administration and Scoring Guide provides no further information about the populations of students in the sample that established these norms. The schools' locations, why they participated, how they participated, who administered the assessments, or what standardized tests were administered to the participating students remain unknown. Because DIBELS claims to be a predictor of success for grade-specific instruction for both normally progressing and struggling readers, this data would best be provided.

As previously noted, Retelling Fluency (RTF) is the only component in DIBELS that attends to comprehension. DIBELS added RTF to the assessment protocol to give the DORF "face value" with teachers. The DIBELS manual states: "Incorporation of an explicit comprehension check may help teachers feel increasingly comfortable with oral reading fluency" (p. 31). Thus, the intent is not to measure validity since "**Face validity** is concerned with the degree to which a test *appears* to measure what it purports to measure, whereas the other forms of validity . . . *provide evidence* that the test measures what it purports to measure" (emphasis in original) (Borg & Gall, 1989, p. 256).

Directly after students read the DORF assessment probe aloud, they are told "Please tell me about what you just read. Try to tell me everything you can. Begin" (p. 32). A stopwatch is started after the testing administrator says "begin." RTF, like DORF, is timed and restricted to a one-minute performance. As students tell about the reading, the administrator is expected to count the number of words the child speaks in his or her retelling by moving a pen through a series of numbers on the scoring sheet while the reader is talking. The number of words spoken in that minute is the base from which a score is generated. Exclamations, songs or recitations (even if relevant to the story), rote repetition of word or phrases, stories or irrelevancies, and repeating of the retelling are not counted. All other words are counted and recorded as the student's RTF. There is no analysis of the content or quality of the retelling.

The student's RTF is an optional part of the assessment program and if it is used its only function is to qualify a DORF score as valid or invalid. The guide suggests that if a child's retell is about 50% of their oral reading fluency score, then the fluency score provides a good overall indication of his/her reading proficiency, including comprehension. "But, for children who are reading over 40 words per minute and whose retell score is 25% or less of their oral reading fluency, their oral reading fluency score alone may not be providing a good indication of their overall reading proficiency" (Good & Kaminski, 2002, p. 31). The guidelines provided mention nothing about the students whose RTF is between 25% and 50% of their oral reading fluency.

Findings

According to guidelines established by DIBELS, students at the end of second grade who read fewer than 70 words per minute are "At Risk" and need intensive or substantial intervention in reading instruction; students reading 70 to 89 words a minute are at "Some Risk" and need "Strategic-Additional Intervention" in reading instruction, and those reading 90 or more correct words per minute have reached the benchmark and are therefore considered "Low Risk." To address our first question, we used this information to make comparisons between students' oral reading fluency of DORF and WCPM of literature texts. We then turned to the instructional recommendations and the students' miscue and retelling data to determine if the instructional recommendations made as a result of DORF were in any way comparable to conclusions drawn from a more detailed analysis of the students' reading proficiency.

Do Students' DORF Scores Predict the Rate and Accuracy with Which they Read Authentic Children's Books?

Our first goal was to determine if a student's DORF reflected that same student's oral reading performance of grade and ability-appropriate authentic text. In other words, does DORF tell us anything about the proficiency with which any individual reader may perform oral reading in another instructional setting? If DORF is an accurate representation of a student's oral performance and can be used to make instructional decisions for that reader, simple comparison of scores should result in somewhat similar ranges of reading. These ranges should also result in a similar "instructional decision" recommendation as the DORF assessment.

In examining the outcomes for DIBELS and literature, our data revealed a great variation between the students' performance on the DORF assessment and their rate and accuracy for the reading of literature (WCPM). The differences between the DORF and the same student's highest WCPM ranged from 4 to 40 words. In all but two cases, Dawn and Mary, the students read the DIBELS probe faster than all of the literature stories. In these two cases, Dawn's DORF was 95 and her WCPM for *Cherries and Cherry Pits* was 102. Mary also read *Cherries and Cherry Pits* faster than her DORF probe (128 and 125, respectively), but with poor comprehension (retelling score of 24). The other 12 readers all read the DORF passages faster than the literature.

Further comparison of the differentiation of these scores shows the variation of instructional recommendations that would be made based on the specific scores. According to the DORF scores, 10 of our 14 readers are classified as "Low Risk" readers. However, if one applied the same DIBELS benchmarks to the literature stories these students read, 10 of the students would be considered either at "Some Risk" or at "High Risk" and be targeted for "Additional Intervention" or "Substantial Intervention." If we applied suggested DIBELS procedures by considering only the median WCPM score of the stories read, only three of the students who read three literature stories would remain in the "Low Risk" category.

We do not interpret these findings as evidence that the students are actually poorer readers than DORF scores indicate, but rather that they are responding to the expectations of the DIBELS task to read faster, with or without comprehension. When reading authentic literature with the expectation for comprehension, readers are more likely to slow down in the process of sorting out and making sense of text. We find readers in our study who read almost half as quickly while reading the literature as when reading for DORF assessments. In fact, there seems to be little relationship between the rate and accuracy with which the readers in this study read the DIBELS passages and their rate and accuracy while reading the literature. Readers with the highest DORF scores are by no means the faster, more accurate readers of literature. Conversely, the students with the lowest DORF scores are not necessarily the slower, less accurate readers of the literature.

Our data also indicate a variation across literature stories for all individual readers, sometimes as much as 50 WCPM. Due to this variation, it is difficult to determine any one student's "fluency level" or whether any student could be considered more or less "fluent" than the others. The exception to this was one reader who emerged as the "fastest" reader, with all four of his rate and accuracy scores exceeding the other readers. No "slowest" reader emerged, as all other students' rates reflected inconsistencies among texts with no pattern emerging among the stories, nor between their DORF scores and story readings. Our findings suggest that the fluency (speed and accuracy) of reading varies with both the demands of the task, as well as the nature of the interaction between a reader and specific

texts, even of comparable difficulty levels. One score cannot be used to determine a student's oral reading fluency.

Do DORF Instructional Recommendations Predict Students' Ability to Comprehend Authentic Children's Books?

Our second goal was to determine whether or not the instructional recommendations made as a result of DORF assessments for this group of students in any way reflected their level of proficiency while reading literature. Comprehension is the goal of all reading and its importance cannot be diminished. DIBELS claims to be a reliable assessment to identify instructional needs and monitor reading performance. What can we learn about a student's comprehension from DORF or RTF? How do scores on either of these assessments lead us to understand the more important question of a reader's comprehension? To further examine these important questions, we turn to the two measures of comprehension used in this study: DIBELS RTF and Retelling Protocol for miscue analysis (Goodman et al., 2005).

Retelling Fluency (RTF). As previously noted, the DIBELS protocol calls for a comparison of the DORF and RTF scores of the students to validate and give "face value" to the DORF scores. Using the RTF scores is optional, and the teacher in this study collected but did not use the RTF scores to make instructional placement decisions based on the DORF scores; she recorded the RTF scores but then assessed the students' median DORF scores in relation to the end of second grade benchmark of 90 words correct per minute. In fairness to the assessment program, we applied the program's RTF/DORF ratio recommendations to the students in our sample so as not to bias our analysis of the instructional recommendations offered in the program guidelines. Had the protocol been followed, only 5 of the 14 students' DORF scores would have been validated by their RTF scores. Eight students' scores fell in the percentages range between 25 and 50%, leaving them in the "dead zone" without any determination of validity. One student, Justin, has an RTF score that falls below 25% of his DORF, indicating that he may be word calling with little or no comprehension. In other words, DIBELS assessment claims to offer useful DORF scores on only 6 of the 14 students, or roughly 43% of the students assessed. The Administration and Scoring Guide gives no information on how to proceed for students whose DORF is not validated by their RTF.

Making Instructional Recommendations for Readers Based on DORF. To assess the appropriateness of the instructional recommendations made for readers in our sample, we first comment on our understanding of "high, some and at risk readers," in the absence of an explicit explanation by DIBELS authors. From our perspective, we would consider a "high risk reader" to be one who generally relies on too few sources of information (cues) in processing the text, is unable to construct meaning during reading, and consequently does not comprehend the text as a whole. A reader "at some risk" might generally utilize some cues, attempt to construct meaning during reading, and comprehend the text to a limited degree. A reader "at low risk" or "no risk" (not a DIBELS classification) would consistently integrate the use of cues, focus on constructing meaning during reading (as evidenced by sensible predictions and self-corrections), and form a coherent understanding of the text as a whole. In all cases, a teacher would be aware of the specific strengths and needs of a student and provide appropriate support in improving or expanding the processing and comprehension of text.

Our six students with valid DORF scores are listed next. If DIBELS is a reliable assessment tool, these designations should fit the profile of these six readers while reading literature.

1. Ethan: DORF 68: At High Risk
2. Rishi: DORF 70: At Some Risk
3. Daniel: DORF 85: At Some Risk
4. Dawn: DORF 95: At Low Risk
5. Julie: DORF 98: At Low Risk
6. Justin: DORF 101: RTF score too low: "may be a comprehension concern."

To evaluate these instructional recommendations, we turned to our more complete profiles of these readers provided by analyzing the miscue data.

Our findings show that while each of the readers has distinct strengths and needs, none of the readers would be considered "high risk readers" by our definition. The books they read for this research are all considered second-grade level or above (Fountas & Pinnell, 2000) and although there is great variation in the retelling scores as can be seen in Table 1, each of these readers scored at least 50% on one of their retellings, indicating that they are all reading with some meaning.

The DORF scores provide us no reliable pattern that predicts the proficiency with which these readers read and comprehend "real" stories. For example, Ethan, who is labeled "At High Risk" according to his DORF score, received the exact same retelling score (50) for the same two stories as Daniel who is classified as "At Some Risk." Julie and Dawn, who are both classified as "At Low Risk" according to DORF, have retelling scores for the same story that vary widely (76 vs. 43). The reader with the highest DORF

Table 1

Retelling protocol percentages for miscue analysis for students with valid DORF scores

Student	DORF	Risk level	Retelling 1%	Retelling 2%	Retelling 3%
Ethan	68	High	<i>A Letter to Amy</i> : 50	<i>See You Tomorrow, Charles</i> : 33	n/a
Rishi	70	Some	<i>A Letter to Amy</i> : 50	<i>See You Tomorrow, Charles</i> : 50	n/a
Daniel	85	Some	<i>A Letter to Amy</i> : 50	<i>See You Tomorrow, Charles</i> : 33	n/a
Dawn	95	Low	<i>Cherries and Cherry Pits</i> : 43	<i>Flossie and the Fox</i> : 48	<i>Precious and the Boo Hag</i> : 50
Julie	98	Low	<i>Cherries and Cherry Pits</i> : 76	<i>Flossie and the Fox</i> : 57	<i>Precious and the Boo Hag</i> : 50
Justin	101	Undetermined	<i>Cherries and Cherry Pits</i> : 38	<i>Flossie and the Fox</i> : 43	<i>Precious and the Boo Hag</i> : 57

score (Justin) did not comprehend significantly better than those with lower DORF scores, and often comprehended less, even when reading the same stories.

DORF scores also fail to predict or distinguish among readers' relative use of cues and meaning during reading. For example, Rishi and Daniel are both considered "At Some Risk." However, miscue analysis of their reading reveals that while both readers make strong use of visual and phonic cues in processing text, Daniel is much more successful in using semantic and syntactic cues to make predictions during reading. Fortunately, both readers show strength in their ability to notice and correct miscues that disrupt meaning. Findings suggest that Rishi needs instruction that will strengthen his ability to make meaningful and grammatical predictions from the onset, so that he can read more efficiently.

As we can see, each reader needs appropriate, supportive instruction that meets his or her individual needs. This can only be determined by a close qualitative examination of their oral reading miscues and retellings. At least for this group of students, DIBELS is not indicative of students' reading proficiency nor does it provide meaningful guidance for their teacher in planning appropriate instruction. Without predictive power, the time and cost of DIBELS, as well as its usefulness, is questionable.

Concluding Thoughts

Our data show no connection between DORF scores and students' oral reading fluency and comprehension of authentic literature. In fact, our data reflect great variation within and between readers on all measures of rate and accuracy, rendering DIBELS' seemingly simple way of rating readers somewhat fruitless. Our findings also reveal that students' oral reading rates show no connection to their comprehension.

This begs the question: How valid is the practice of collecting rates of oral reading? The teacher in this study did group the students for instruction based on their speed of reading, placing the fastest readers in the instructional "high" group. However, when working with the readers to understand their retelling, or comprehension ability, the students who read faster were not necessarily the better comprehenders. Thus, grouping them together as like readers prevents the teacher from individualizing instruction to meet students' varied needs.

Although our sample is small, so are classrooms where teachers work daily to improve reading development for their students, and teachers need to know that the *statistical* claims made by an assessment program will hold true to their students. These conclusions suggest the need to reconsider not just the use of DIBELS but the entire practice of using measures of rate and accuracy to assess students' reading proficiency and comprehension. We realize that the time it takes to conduct qualitative analyses such as miscue analysis far exceeds the time it takes to conduct single-minute assessments, but we contend that if we really are interested in putting reading first, we will spend the time necessary to understand and respond to each student's unique strengths and needs.

References

- Allington, R. L. (2007). Education Policy Studies Laboratory: Education Policy Research Unit. Review of Whole language high jinks: How to tell when "scientifically-based reading instruction" isn't. *Arizona State University*.
- Altwater, B., Jordan, N., & Shelton, N. (2007). *Rereading fluency*. Portsmouth, NH: Heinemann.
- Borg, W. R., & Gall, M. D. (1989). *Educational research: An introduction* (5th ed). New York: Longman.

- Carlisle, J. F., Schilling, S. G., Scott, S. E., & Zeng, J. (2004). Do fluency measures predict reading achievement? Results from the 2002–2003 school year in Michigan's Reading First schools. (Technical report #1). Ann Arbor: University of Michigan.
- Cohen, M. (1983). *See you tomorrow, Charles*. Ill. by L. Hoban. New York: Greenwillow Books.
- Fountas, I., & Pinnell, G. S. (2000). *Leveled books for readers*. Portsmouth, NH: Heinemann.
- Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic indicators of basic early literacy skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement. Retrieved March 15, 2007, from <http://dibels.uoregon.edu/>.
- Goodman, K. S. (Ed.). (2006). *The truth about DIBELS: What it is and what it does*. Portsmouth, NH: Heinemann.
- Goodman, Y., Watson, D., & Burke, C. (2005). *Reading miscue inventory: From Evaluation to instruction*. Katonah, NY: Richard Owen.
- Kamii, C., & Manning, M. (2005). Dynamic Indicators of Basic Early Literacy Skills (DIBELS): A tool for evaluating student learning? *Journal of Research in Childhood Education*, 20, 75–90.
- Kamps, D. M., Willis, H. P., Greenwood, C. R., Thorne, S., Lazo, J. F., Crockett, J. L. et al. (2003). Curriculum influences on growth in early reading fluency for students with academic and behavioral risks: A descriptive study. *Journal of Emotional and Behavioral Disorders*, 11(3), 211–224.
- Keats, E. J. (1998). *A letter to Amy*. New York: Puffin.
- Manning, M., Kamii, C., & Kato, T. (2006). DIBELS: Not justifiable. In K. S. Goodman (Ed.), *The truth about DIBELS: What it is and what it does*. Portsmouth, NH: Heinemann.
- Mathson, D. V., Allington, R. A., & Solis, K. L. (2006). Hijacking fluency and instructionally informative assessments. In T. Rasinski, C. Blachowicz, & K. Lems (Eds.), *Fluency Instruction: Research-based best practices* (pp. 106–119). New York: Guilford Press.
- McKissock, P. (1986). *Flossie and the fox*. Ill. by R. Isadora. New York: Dial.
- McKissack, P., & Moss, O. J. (2005). *Precious and the Boo Hag*. Ill. by K. Brooker. New York: Atheneum/Anne Schwartz Books.
- National Reading Panel. (2000). *Teaching children to read, an evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: National Institute of Child Health and Human Development.
- Office of the Inspector General—United States Department of Education. (2006). *The Reading First Program's Grant Application Process—Final Inspection Report* (Publication ED-OIG/I13-F0017). Washington, DC: U.S. Government Printing Office.
- Office of the Inspector General—United States Department of Education. (2007). *The Department's Administration of Selected Aspects of the Reading First Program—Final Audit Report* (Publication ED-OIG/A03G0006). Washington, DC: U.S. Government Printing Office.
- Pressley, M. (2006). What the future of reading research could be. Paper presented at the International Reading Association's Reading Research 2006. April 29, 2006. Chicago, IL.
- Pressley, M., Hilden, K., & Shankland, R. (2005). An evaluation of end-of-grade 3 dynamic indicators of basic early literacy skills (DIBELS): Speed reading without comprehension, predicting little. East Lansing: Literacy Achievement Research Center, Michigan State University.
- Reidel, B. (2007). The relation between DIBELS, reading comprehension, and vocabulary in urban first-grade students. *Reading Research Quarterly*, 42, 546–567.
- Samuels, S. J. (2006a). Looking backward: Reflections on a career. *Journal of Literacy Research*, 38(3), 327–341.
- Samuels, S. J. (2006b). Toward a model of reading fluency. In S. J. Samuels & A. E. Farstrup (Eds.), *What research has to say about fluency instruction* (pp. 24–46). Newark, DE: International Reading Association.
- Samuels, S. J. (2007). The DIBELS tests: Is speed of barking at print what we mean by reading fluency? *Reading Research Quarterly*, 42(4), 563–566.

- Schilling, S. G., Carlisle, J. F., Scott, S. E., & Zeng, J. (2007). Are fluency measures accurate predictors of reading achievement? *Elementary School Journal*, *107*(5), 429–448.
- Wilde, S. (2006). But isn't DIBELS scientifically based? In K. S. Goodman (Ed.), *The truth about DIBELS: What it is and what it does* (pp. 66–70). Portsmouth, NH: Heinemann.
- Williams, V. (1991) *Cherries and Cherry Pits*. New York: HarperTrophy.



COPYRIGHT INFORMATION

TITLE: Does DIBELS Put Reading First?
SOURCE: Literacy Res Instr 48 no2 2009

The magazine publisher is the copyright holder of this article and it is reproduced with permission. Further reproduction of this article in violation of the copyright is prohibited.