

Running head: EVALUATION OF END-GRADE-3 DIBELS

An Evaluation of End-Grade-3
Dynamic Indicators of Basic Early Literacy Skills (DIBELS): Speed Reading Without
Comprehension, Predicting Little

Michael Pressley Katherine Hilden Rebecca Shankland
Michigan State University

This research was supported by a grant from the Michigan State University Research Excellence Fund to the Literacy Achievement Research Center (LARC). The authors are grateful for the outstanding cooperation of the Administration and Faculty of the South Redford MI School District, especially Cherie Cornick, Michelle Chase, Amy Davidson, and Carol Lindman.

Abstract

End grade-3 students in one school district were administered the DIBELS oral fluency (with retelling) measure, with either standard directions, directions to read quickly, or directions to read for understanding. The pattern of performance in the standard condition was more consistent with students interpreting the test as a test of speeded reading. Comprehension of what was read was low, especially when the retelling data were scored with respect to idea units. Moreover, DIBELS was only weakly predictive of how the students performed on a subsequent standardized reading assessment, the *TerraNova*. When considered in light of previously reported data, the case is made that additional evaluation of the DIBELS should be carried out by parties other than its developers and those already committed to use of the measurement, evaluating its predictive validity with respect to a wide range of reading achievement measures. Based on available data, the fairest conclusion is that DIBELS mis-predicts reading performance on other assessments much of the time, and at best is a measure of who reads quickly without regard to whether the reader comprehends what is read.

An Evaluation of End-Grade-3 Dynamic Indicators of Basic Early Literacy Skills (DIBELS):
Speed Reading Without Comprehension Predicting Little

In the investigation reported here, we explored what can be learned about a student's reading from the end of third grade Dynamic Indicators of Basic Early Literacy Skills (DIBELS) assessment, from both the Oral Reading Fluency score and the Oral Retelling Score. The DIBELS deserves research attention because it is becoming widely used in American schools, with the DIBELS website regularly updating the expanding number of children being assessed with the approach (<http://dibels.uoregon.edu/>). For example, many states have been forced to mandate that DIBELS be used in schools receiving federal Reading First funds (Brownstein & Hicks, 2005a, 2005b; Manzo, 2005a, 2005b), which support beginning reading instruction (grades K-3) in schools with large proportions of students at risk for reading failure. The mandate is for DIBELS to be central in the monitoring of students' progress in learning to read, with teachers using information from the DIBELS in their instructional decision making.

Dynamic Indicators of Basic Early Literacy Skills (DIBELS)

DIBELS is actually a suite of assessments that can be administered beginning in preschool and kindergarten. There is a letter naming fluency scale that can be administered in kindergarten and grade 1, complemented by tests of the more advanced skills of fluency in phoneme segmentation and reading of nonsense words. Beginning in mid-grade 1, there are tests of Oral Reading Fluency and Oral Retelling fluency, with the end grade-3 versions the specific target of investigation in this study.

We focus on Oral Reading Fluency and Oral Retelling measures (Good, Kaminski, & Dill, 2002) here for several reasons. This is the only DIBELS measure that extends from grades 1 through 6, and thus, has the potential to influence teacher appraisal of students and instructional

decision making for a very long time. We also focus on the oral reading and retelling measures because the developers of DIBELS imply that this measurement can provide information about comprehension (i.e., signal whether the child is making adequate progress with respect to comprehension; Good & Kaminski, 2002), which is the most advanced reading goal targeted by the federal program that has leveraged DIBELS into the American school place (*107th Congress*, 2002). Comprehension is also the higher-order reading goal that is of greater interest to elementary educators than any other reading goal (Block & Pressley, 2002). Why study Oral Reading Fluency and Oral Retelling at the end-of-grade-3 level? Since the Oral Reading Fluency and Oral Retelling measurements that extend from grades 1 to 6 are comparable in structure to the grade 3 assessment, we judged the grade 3 assessment was representative of measurement of Oral Reading Fluency and Oral Retelling more generally on the DIBELS. In addition, end-of-grade-3 performance is of special significance in a national accountability system that expects students to be reading at grade level by the end of grade 3 (*107th Congress*, 2002).

Although there are some correlations for an earlier version of the Oral Reading Fluency measure (i.e., *Test of Reading Fluency*, Children's Educational Test Service, 1987) with conventional standardized measures of reading (see Good & Jefferson, 1998), much of this data was generated with older students and as part of in-house development of the *Test of Reading Fluency*, rather than by outside evaluators. The main validation data on the DIBELS are that it predicts performance on state reading tests (e.g., Barger, 2003; Buck & Torgesen, 2003; Carlisle, Schilling, Scott, & Zeng, 2004; Elliott, Lee, & Tollefson, 2001; Good, Simmons, & Kame'enui, 2001; Hintze, Ryan, & Stoner, 2003; Shaw & Shaw, 2002). Most of these data have been produced by those associated with federally-funded reading efforts employing the measure. Because the DIBELS is predictive of whether the student can pass a state test, presumably, it can

provide a signal to the classroom teacher about students who are making adequate progress in reading and those who are not, providing indication about which students are at risk of not making expectations on subsequent accountabilities.

How does the DIBELS Oral Reading Fluency and Oral Retelling measure work? The child is presented a short passage to read aloud, with the child aware that there may be a subsequent assessment of the immediate memory of the text. Specifically, the child is told by the examiner, “When I say ‘stop.’ I may ask you to tell me about what you read, so do your best reading.” The child reads aloud for one minute, with the examiner recording the number of words read correctly, which is the Oral Reading Fluency score for the story. Then, the child’s memory of the text is tested, with the examiner prompting, “Please tell me all about what you just read. Try to tell me everything that you can.” This is scored as a retelling, with the examiner recording the number of words recalled in one minute from the passage by the student. This procedure is repeated with three stories, with the Oral Reading Fluency and Oral Retelling score for the overall assessment being the median scores for the three stories.

How are the Oral Reading Fluency and Oral Retelling scores interpreted? By far, the more important score is the Oral Reading Fluency score. At the end of grade 3 level, students who score above 110 words per minute read are classified as low risk of failing subsequent reading assessments, with 110 corresponding to the 40th percentile of performance at the end of grade 3 level. Students scoring below 80 words per minute are considered at risk of failing, with 80 corresponding to the 16th percentile of end-of-grade-3 readers. Students scoring between 80 and 110 are classified as at “some risk” of failure on subsequent state-level or other standardized assessments. Teachers are to use these classifications to decide which students should receive

remediation, although for the end-of-grade-3 level Oral Reading Fluency measure, there is no specific guidance as to what type of remediation is appropriate.

There is some guidance as well about how to interpret the oral retelling scores, with these derived from analyses of the normative data on the DIBELS collected by the developers.

Teachers are instructed as follows:

So, a rough rule of thumb may be that, for children whose retell is about 50% of their Oral Reading Fluency score, their Oral Reading Fluency score provides a good overall indication of their 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 provide a good indication of their overall reading proficiency....there may be a comprehension concern that is not represented by their fluency (Good & Kaminski, 2002, p. 31).

In other words, for most end-of-grade-3 readers (i.e., those above the 4th percentile on Oral Reading Fluency, which is the percentile corresponding to an Oral Reading Fluency of 40), teachers are advised that comprehension is reflected by the Oral Reading Fluency score, if the retelling score is at least 50% of the fluency score. There is great concern about comprehension not being consistent with fluency if the retelling score is less than 25% of the fluency score---presumably reflecting that students can “word call” without getting the meaning of the passage. Students with Oral Retellings between 25% and 50% of fluency scores would also prompt concerns, just not as serious as for students with Oral Retellings less than 25%. In signaling that

additional instruction is required for children with Oral Retellings less than 50% of fluency scores, DIBELS is silent, however, about how intervention should occur.

Concerns About the DIBELS Oral Reading Fluency and Retelling Measures that Motivated This Study

The first impressions we had when we witnessed students doing the oral reading as part of the DIBELS fluency and retelling measure was that student focus was on reading quickly and accurately, with little indication of the type of pausing and conscious thinking that typifies reading when readers are reflectively constructing meaning (Pressley & Afflerbach, 1995). We found ourselves wondering just how the students taking the assessment interpreted their task. Did they view the instruction to do best reading as a prompt to read as quickly as possible, or did they focus on the fact that recall of the text was probably going to be required? Good and Kaminski (2002) clearly indicate that they feel the retelling requirement is sufficient to make certain students do not view the task as simply rapid oral reading without concern for comprehension. We felt that we could gauge better, however, just how students do interpret the DIBELS by experimentally varying the directions students experienced as part of the assessment. Thus, in this study, one-third of the participants were administered the oral fluency and retelling task with the standard directions, but one-third received directions emphasizing that what mattered was speed and accuracy in reading, and one-third received instruction emphasizing that what mattered was understanding the text. We reasoned that if the standard direction was interpreted as a speed reading direction, there would be no difference in the performances of students in the standard direction and speed reading direction conditions, with performances in those conditions differing from performances in the condition emphasizing comprehension. Alternatively, if the standard direction was prompting attention to comprehension, then no differences in

performances between the standard direction and comprehension-emphasis direction conditions would be expected, with performances in those two conditions differing from performances in the speed-emphasis condition. In short, we expected to learn quite a bit from this manipulation about how the end-of-grade-3 students interpreted the reading task they were given as part of the DIBELS fluency and retelling measure.

A second concern with respect to measurement issues has important implications for interpreting the retelling data with respect to comprehension. This second concern is whether the Oral Retellings are accurately recorded during this assessment. The individual administering the test must count the number of words correctly retold as students do their Oral Retellings. From our first encounter with the test, this seemed a formidable task, something recognized in the formal assessment literature (Rathvon, 2004). In fact, Rathvon (2004) recommended that such retellings should be scored from tape recorded student responses rather than as students performed the task. Thus, in this study, Oral Retellings were scored as students provided them, and, then again, from the taped recall, with the scores compared formally.

We admit that from the outset of this study, as comprehension researchers, we found ourselves skeptical that the DIBELS Oral Reading Fluency and retelling measures could reveal much about reading comprehension. Our suspicions were aroused initially because simple Oral Retellings, operationalized as the number of words recalled from a story, are the only comprehension data collected. It has been recognized for a very long time (e.g., Graesser, 1981) that there is a great deal of comprehension of text above the word level, that focusing on word-level processing is a decision to ignore ideas in text that are captured by relationships between words, phrases, sentences, and paragraphs, with such relational information at the heart of what many scholars consider comprehension (e.g., van den Broek, 1994; van Dijk & Kintsch, 1983).

We could imagine readers who would recall a lot of words in a text but scramble the ideas, with such students achieving the same retelling score as they would receive if ideas were not scrambled. In fact, from our first efforts to appraise the DIBELS (i.e., by administering it to students), we witnessed a lot of word scrambling. This finding is consistent with reports of educators in the school district hosting the study and other scholars who have witnessed contextual fluency retellings on curriculum-based measures (Jay Samuels, personal communications, spring & summer, 2004). Thus, we were motivated in this study to appraise whether DIBELS retelling data, in fact, had so much word scramble that there is reason to doubt students are comprehending what they read during the assessment. To do so, in addition to scoring the retelling data for simply the number of words recalled, we also evaluated recall of idea units above the word level (i.e., recall of ideas in the text).

Does the DIBELS provide information that the teacher does not already know? In particular, does the DIBELS heighten awareness of who is doing poorly and who is doing well in learning to read? This possibility is the motivation for administering DIBELS. As far as we could tell from reading the validation studies already cited in this introduction, no one has attempted to determine whether the DIBELS provides information about future student success that the teacher did not already have. Thus, in this study, we made a first attempt to do so, analyzing whether teachers' ratings of their students' reading were better, comparable, or worse predictors of standardized reading achievement than the DIBELS, with the ratings derived from reading grades given to the students. If information derived from the DIBELS added no additional predictive power over teacher evaluations of students, there would be substantial reason to doubt whether mandating this assessment as a progress monitor is necessary. If, on the other hand, the DIBELS information increased the predictability of standardized reading scores over teacher

appraisals, then there would be increased justification for use of the DIBELS as a progress monitor.

Methods

Participants

All third graders (approximately $n = 250$) in a small school district serving a Midwest urban area were asked to participate, with a parental consent letter sent home. All 191 students who returned permission slips participated in the study. The participants were students in four different schools, with these students taught by 10 teachers. The district served primarily working-class neighborhoods, with 16.6% of students in the district receiving free lunch and 38.7% of the students classified as minority.

Because the district had just begun DIBELS testing during the school year when this study was carried out, and because a decision was made to begin it in the early primary grades (i.e., before grade 3), the students in this study had not previously been administered a DIBELS assessment, so that they experienced the assessment for the first time as participants in this study.

Materials

The study was carried out in May, 2004 (i.e., at the end of the grade-3 year for the students in the study). Thus, students read and recalled three end-of-grade-3 assessment texts for the DIBELS Oral Retelling Fluency measure (Good & Kaminski, 2002), passages entitled *My Parents*, *Pots*, and *Animal Tracks*. These were 211, 262, and 242 words respectively.

Procedure

Each participant was tested in a quiet place in the school by one of the three authors of the study. Testing took between 10 and 15 minutes per student. After greeting the student, the tester requested the student's permission to tape record her or him (which was granted in all

cases). Then, the formal assessment began, which involved the student reading and re-telling the three stories, presented in counterbalanced order across each condition (i.e., one-sixth of the students in each of the conditions experienced the stories in each of the six possible orderings of three stories).

We closely followed the DIBELS protocol through the administration of the assessment in the standard directions condition. Students in this condition were presented each of three stories to read with the following direction: “Please read this out loud. If you get stuck, I’ll tell you the word so you can keep reading. When I say ‘stop,’ I may ask you to tell me about what you read, so do your best reading. Start here.” As the tester concluded this direction, she or he pointed to the first word of the first passage. At the end of the minute of reading, the tester recorded the total number of words read from the passage. Then, the student was asked to retell the passage to the tester. The tester said, “Please tell me all about what you just read. Try to tell me everything you can. Begin.” The first time the student failed to say anything for three seconds, the researcher prompted the student by saying, “Try and tell me everything you can.” At the end of one minute, or after the student reported she or he was finished retelling the story, the researcher concluded, counting the number of words recalled. The researcher attempted to follow closely the DIBELS scoring manual guidelines for scoring the total number of words retold. For example, contractions were counted as one word. Also, consistent with the manual, “minor repetitions, redundancies, irrelevancies, and inaccuracies” were counted. After the first story was read and retold, students read and retold the second story following the same directions, which was then followed by the third story.

The procedures were identical in each of the other two conditions of the study, with the exception of the direction provided just before each text was read. In the speed-emphasis

condition, participants were not told to, “Do your best reading,” but were told to read, “As quickly as possible.” In the comprehension-emphasis condition, “Do your best reading” was replaced with, “It is important that you read in order to understand the story.”

Participants were assigned randomly to the three conditions of the experiment. Each of the three testers tested approximately equal numbers of participants in each condition, with the students in each class assigned in approximately equal numbers to the three conditions.

DIBELS Measures

The DIBELS evaluation yielded oral fluency and recall scores. Oral fluency for each story was simply the number of words correctly read in the one minute. The oral fluency classification of a student is not based on a single story, however. Rather, students are classified based on reading the three stories, with the median story oral fluency score used to determine classification. A median score of less than 80 indicates high risk, between 80 and 109 some risk, and 110 or greater low risk. For example, a student, who read 92 words in *Pots*, 111 words in *My Parents*, and 79 words in *Animal Tracks*, would be classified as at some risk because the median score, the *Pots* score, was 92, which is between 80 and 109. For that student, in the analyses that follow, 92 would be the oral fluency score plugged into analyses.

We also calculated DIBELS retelling scores, based on the number of words recalled. The number of words retold were counted by the testers as the students recalled, with these referred to as the Words Retold/On-Line Scored. As detailed in the results, these retelling data were re-scored by transcribing the retellings, with these referred to as Words Retold Re-Scored. The recall data were additionally re-scored in terms of the number of idea units recalled (Kintsch, 1998), again as detailed in the results section, with this referred to as Propositional Retelling in what follows.

School District-Provided Measures

The school district provided May, 2004 *TerraNova* scores (CTB McGraw-Hill, 2004) for each participant, which is a comprehensive standardized reading assessment. The Grade Equivalent score was used as a measure of reading achievement in this study.

We also had access to students' Qualitative Reading Inventory (Leslie & Caldwell, 1995) scores, collected by school officials. On this assessment, students read short passages, with each student's reading rate measured in words per minute (i.e., QRI reading rate). Students then answered comprehension questions resulting in a QRI reading comprehension score. The comprehension measure required memory of both explicit and implicit ideas expressed in the passage.

The teacher of each student in the study rated the student's fluency on a six-point, 0 to 5 scale, with 0 meaning non-reader, 1 meaning fluency akin to a first grader, 2 struggling fluency, 3 average fluency, 4, good fluency, and 5 meaning excellent oral fluency, with this score referred to subsequently as Teacher Fluency. Teachers also rated students comprehension on a similar scale, with this rating referred to as Teacher Comprehension. In that case, 0 meant the reader could not make sense of text, 1 that the reader does not actively use strategies to make sense of text and frequently misunderstands text, 2 indicated struggling comprehension, 3 acceptable or average comprehension with grade 3 texts, 4 good comprehension as reflected by use of strategies and understanding of text, and 5 very strategic and excellent understanding of above grade level texts.

The school district provided the reading grades for each student in the study for the fourth marking period during the school year (i.e., the marking period when data were collected). The district graded several components of reading on a 1 to 3 scale (1 = experiencing difficulty, 2 =

developing, 3 = excellent progress): participation in sustained silent reading, uses reading strategies to construct meaning, interprets and analyzes narrative text, and interprets and analyzes informational text were all graded. These four grades were added, so that a student's grade could range from 3 to 12. This measure is referred to as Grades in what follows. We emphasize that teacher grades were generated on the basis of classroom assessments only, with the teachers having no access to the results of the DIBELS assessment reported here or the *TerraNova* (CTB/McGraw-Hill, 2004) comprehensive reading assessment, which was administered at the end of grade 3 after the DIBELS data were collected.

Results

Oral Fluency

The oral fluency is the number of words read in a minute, with the mean values recorded in Table 1 as a function of condition and story. For each story, the condition means were compared pairwise using a set of Bonferroni contrasts (Kirk, 1994), $p < .05$ for the set of three pairwise contrasts, and, thus, $p < .0167$ for each comparison. When the DIBELS standard condition was contrasted with the speeded condition, none of the pairwise contrasts were significant, largest $|t|$ (124) = 2.27, $.025 > p > .0167$, with absolute effects sizes (hereafter, $|ES|$ s) ranging from .11 to .41, mean $|ES| = .22$. With respect to DIBELS standard versus the understanding condition, the contrast was significant for two of the three stories: for *My Parents*, $|t|$ (125) = 2.41, $p = .0167$ ($|ES| = 0.42$); for *Animal Tracks*, $|t|$ (124) = 2.81, $p < .0167$ ($|ES| = 0.50$). It was not significant for *Pots*, although the trend was in the direction of faster reading in the speeded condition, $|t|$ (124) = 2.08, $.04 > p > .0167$ ($|ES| = 0.37$). None of the differences between the speeded and the understanding condition were significant, large $|t|$ (125) = 1.53, $.13 > p > .0167$. $|ES|$ s ranged from 0.08 to 0.27, mean $|ES| = 0.20$. In short, although the differences

between conditions were small, the data are more consistent with the conclusion that students interpreted the DIBELS standard direction as a direction to read rapidly than as a direction to read for understanding, although even the direction to read for understanding produced reading speed not reliably different from when students were instructed explicitly to read rapidly.

DIBELS scoring does not make much of any individual story data, however. Rather, each participant's DIBELS oral fluency classification is based on their median score for three stories. So, consistent with the DIBELS scoring manual, we classified students based on their median scores, resulting in classification of high risk (cutoff score of less than 80 words per minute for the median story), moderate risk (80-109 words per minute for the median story), and low risk (110 and above for the median story). The proportion of participants in each risk category as a function of condition is recorded in Table 2. The apparent trend towards less risk in the DIBELS standard condition compared to the other conditions was significant, $X^2(4) = 10.27, p < .04$, a small effect, with $\phi = 0.13$ (Grissom & Kim, 2005). Perhaps most striking was that almost half of the students in the standard condition (i.e., 46%), were considered at least to be at moderate risk for low performance on a subsequent more comprehensive reading assessment. Of course, this is only a slightly greater proportion than for the norming sample, with anyone scoring at the 40th percentile or lower considered at risk, and, hence, 40% of the norming sample was considered at risk.

Retelling of Stories

As the students recalled the stories, the researchers attempted to keep track of the number of words retold, consistent with the standard DIBELS administration procedures, with these retelling data summarized in Table 3 as the number of "words retold/on-line scored" for each story. The researchers were very aware as they collected these data, however, that they were not

perfectly accurate in counting the words retold. Thus, using the taped retelling data, the researchers recounted the number of words retold. On average, the two researchers were within 2.37 words of one another ($SD = 2.49$), with their scores highly correlated (i.e., $r = .99$). These re-scored retelling data are also contained in Table 3. For each participant and story, the inaccuracy of the original retelling score was assessed by calculating the absolute value of the difference of the words retold/on-line and words retold/rescored score, with these inaccuracy scores in Table 3 as well. The most important finding in these analyses, summarized in Table 3, was that for every story, there was inaccuracy in on-line scoring of the retelling data, a large effect, with $|ES|$ s ranging from 0.89 to 1.00, mean $|ES| = 0.95$. Hence, in the analyses that follow, the words retold/re-scored data are analyzed.

Words retold/re-scored is broken down as a function of story and condition in Table 4. For each story, the pairwise comparisons as a function of condition were calculated, with the overall Type 1 error rate set at .05 for each story, and hence, .0167 per comparison. None of the nine comparisons were significant by this criterion, largest $|t| (123) = 2.03$, $p > .04$, with $|ES|$ s ranging from 0.04 to 0.37, mean $|ES| = 0.15$. We also analyzed the proportion of words retold/re-scored relative to the number of words read, with those proportions recorded in Table 5. The same analysis approach was taken as with the number of words retold revised, with none of the comparisons significant, largest $|t| (122) = 1.81$, $p > .07$, $|ES|$ s ranging from 0.04 to 0.32, mean $|ES| = 0.15$.

Each of the stories was parsed into its propositional idea units (Kintsch & vanDijk, 1978). Thus, the first paragraph of the *Pots* story, offered as an example in Appendix A, contained 28 propositions (expressed as phrases in the appendix). The entire *Pots* story, which was what was presented to the participants, had 85 propositions, with *My Parents* and *Animal Tracks* having 73

and 97 propositions respectively. The words retold/re-scored data were scored for each student and story with respect to the presence of these idea units in the Oral Retellings. In order to estimate the reliability of this scoring, two scorers independently scored 73 Oral Retellings of stories, with the raters in agreement within two propositions for 85% of these story Oral Retellings.

The retelling of idea units was low for all stories and all conditions (see Table 6). For each story, Oral Retellings between conditions were compared pairwise using a Bonferroni approach analogous to the approach used earlier, with $p < .0167$ per comparison as the significance level. Only one of these comparisons was significant. For the *Pots* story, retellings in the DIBELS standard condition exceeded retellings in the understanding condition, $|t| (124) = 3.08$, $.003 < p < .0167$ ($|ES| = 0.56$). All other $|t|/s < 1.56$, $ps > .10$, $|ES|$ s ranging from 0.02 to 0.28, mean $|ES|$ for all comparisons = 0.16.

Perhaps what is most striking in Table 6 is that retelling scores are so low, with very few of the total number of propositions in the stories retold. One reaction is that Oral Retellings should be considered relative to the number of propositions read. Those proportions are recorded in Table 7, with Oral Retellings still low when calculated this way, averaging about 15% retelling of the propositions that were read. For each story, the three possible pairwise comparisons between conditions were made using a set of Dunn-Bonferroni comparisons, $p < .05$ for the set, $p < .0167$ per comparison. None of these comparisons were statistically significant, largest $|t| (126) = 2.26$, with 0.44 SD effect size for this non-significant effect. All other comparisons were $|ES|$ less than 0.17.

A claim made in the DIBELS technical manual is that if a student's Oral Retelling score is about 50% of their Oral Reading Fluency score, then the Oral Reading Fluency scores

provides a good overall measure of reading proficiency, including comprehension. If the retelling is less than 50%, there is a reason for concern about comprehension. (We admit that we cannot discern the basis for this claim.) Given the minimal performance differences between conditions in this study, but also the fact that the standard condition was the one condition completely consistent with the DIBELS administration guide, we consider only the data in the standard condition for the present analysis and all remaining analyses in this results section.

Collapsing across all stories, 33% of students in the standard condition had Oral Retelling scores at or greater than 50% of their Oral Reading Fluency score. About 45% had retelling scores that were between 25% and 50% of their Oral Reading Fluency scores. Finally, 22% of the students had retelling scores that were less than 25% of the Oral Reading Fluency scores. Based on these data, we conclude by the DIBELS developers' criteria, the Oral Reading Fluency score was not a very good measure of overall reading proficiency for the majority of students in this sample, that for a nontrivial proportion of students there should be a concern about comprehension. If about 50% is defined as 40% or greater (i.e., a liberal definition of "about 50%" is taken), the DIBELS Oral Reading Fluency score would be a good overall measure of reading proficiency for only 45% of the students, based on the DIBELS manual definition. Thus, using even a very liberal criterion, there are concerns about whether the DIBELS is a good measure of overall reading proficiency for the majority of students, concerns about the comprehension of the majority of the students in this sample.

Beyond the DIBELS own criteria, however, there is one very good reason to doubt that the Oral Reading Fluency measure says much about comprehension based on Oral Retelling: The correlations between the Oral Reading Fluency scores and Oral Retelling (revised) scores were very low. In the standard condition, r s of .32, .16, and .31 respectively for the *Animal Tracks*, *My*

Parents, and *Pots* stories. Whatever the rank ordering of the oral fluency scores is, it does not predict the rank ordering of the Propositional Retelling scores very well---accounting for a little more than 9% of the variance in Propositional Retellings for two of the stories and less than 3% of the variance for the third story.

Performance on the TerraNova

Based on the DIBELS data, there should be reason for concern for a high proportion of students in this school district, and, alarm bells should be going off about reading in the district in general. That is, in the DIBELS standard condition, based on the oral fluency data, 46% of the students were classified as being at least at moderate risk of not performing as should be expected of end-of-year grade 3 students on a standardized reading assessment. The Oral Retelling data also signaled problems for about the same proportion of students. So, a first question with respect to the *TerraNova* achievement data was whether performances on that measure were consistent with the DIBELS prediction of great reading difficulties in this group of students.

A superficial reading of the *TerraNova* data might suggest that was the case, with only 60% of the students in the standard condition reading at or above grade level. Twenty-nine percent were reading a year or more below grade level, with the remaining 11% reading at the beginning or middle grade-3 level. What has to be understood, however, is that “grade level” performance is typical performance, about the 50th percentile for the grade level tested (<http://titlev.adams.edu/Outreach/TerraNova.htm>). Thus, this school district was, in fact, doing a little better than the typical grade 3 sample. That is, *TerraNova* outcomes should not set off alarm bells for the district, contrasting with the alarm bells set off by the DIBELS.

Of course, DIBELS is intended to flag individual children who are at risk, and, thus, the real litmus test is whether the performance of individual children on the *TerraNova* was predicted by the DIBELS data. The only DIBELS measure used much in the previous DIBELS validation work has been the oral fluency measure, and, thus, that was the DIBELS variable of greatest interest in the analyses of the *TerraNova* outcomes, with an important question being whether the students who scored poorly on the DIBELS Oral Reading Fluency were also the students who performed poorly on the *TerraNova*. In fact, DIBELS Oral Reading Fluency did predict grade equivalent on the *TerraNova*. Using the participant's median oral fluency score (consistent with the DIBELS manual specified approach to determining Oral Reading Fluency) to predict the *TerraNova* grade equivalent score, $R = .446$, $R^2 = .199$, $F(1, 61) = 15.12$, $p < .001$. That is, less than 20% of the variability in the *TerraNova* scores was predicted by the DIBELS oral fluency data, meaning that the DIBELS provided fairly poor prediction of individual students at risk for scoring low on the *TerraNova*.

DIBELS versus Other Predictors

Did the DIBELS oral fluency scores add any prediction over other measures that the school would collect anyway---in the case of this school, the Qualitative Reading Inventory (QRI), both the fluency measure in words per minute and the reading comprehension score as the percentage correctly answered comprehension questions; the teacher's reading grade; the teacher's comprehension rating; and the teacher's fluency rating. Unfortunately, not every measure was available for every participant, so that in Table 8, the n is reduced relative to the n for the entire standard condition for some of the measures as well as in the regressions that follow, which accounts as well for slightly different values of R and R^2 than reported in the last paragraph when DIBELS Oral Reading Fluency is entered as the first predictor in the

regressions. Even so, we think these analyses are informative, and they are summarized in Table 9.

For each school-based variable, two hierarchical linear regressions were conducted (Cohen, Cohen, West, & Aiken, 2004). In a hierarchical regression, after the first variable is entered, change in prediction is measured with the impact of the first variable held constant. For the first hierarchical regression in each set, DIBELS oral fluency was entered first (and its impact assessed alone, based only on the participants with data in this analysis, which was less than 63 because of some missing data for school provided variables), and then the school provided variable was entered to assess its impact with DIBELS oral fluency controlled. For the second hierarchical regression in each set, the school provided variable was entered first (and its impact assessed only), and then the DIBELS oral fluency was added to assess its impact with the school provided variable controlled. If the school provided variable predicted *TerraNova* performance with DIBELS oral fluency already in the equation, but DIBELS oral fluency did not add additional predictive power with the school provided variable in the equation, that would be strong support for the school provided variable as a better predictor of reading achievement than the DIBELS Oral Reading Fluency measure. If the school provided variable added no prediction with the DIBELS Oral Reading Fluency measure in the equation, and DIBELS added prediction with the school provided variable in the measure, this would make a case for the predictive superiority of the DIBELS Oral Reading Fluency measure over the school-provided measure. If both variables added prediction when the other was in the equation, there is clear case for collecting both.

We consider first the measure that is most like the DIBELS oral fluency measurement, which is the QRI reading rate measure. As detailed in Table 9 (top panel), it does predict

TerraNova performance, with the overall pattern of prediction in the regressions consistent with the conclusion that QRI reading rate is a better predictor of *TerraNova* performance than the DIBELS Oral Reading Fluency. In contrast, although the QRI comprehension measure does predict *TerraNova* performance, once DIBELS Oral Reading Fluency is accounted for, knowing QRI comprehension adds no additional predictive power, so that the DIBELS measure is the better measure relative to QRI comprehension (Table 9, Panel B). The same pattern was obtained with teacher grades and teacher comprehension ratings as with QRI comprehension, so that the DIBELS Oral Reading Fluency score is a better predictor than teacher grades or teacher comprehension ratings (Table 9, Panels C & D). In contrast, the teacher's fluency ratings did not significantly predict *TerraNova* performance even when entered alone, and, thus, the DIBELS Oral Reading Fluency measure would clearly be favored over it (Table 9, Panel E). In summary, there was generally good support for the predictive value of the DIBELS Oral Reading Fluency measure relative to these other measures, except that it adds little relative to the QRI reading rate measure.

Finally, although the retelling data are rarely used to make decisions about students, we evaluated whether the DIBELS retelling data predicted *TerraNova* at all in the standard condition. The answer was no. For re-scored Oral Retellings, $R = .205$, $R^2 = .042$, $F(1, 61) = 2.67$, $p < .11$; for proportions of re-scored retold words to words read, $R = .098$, $R^2 = .010$, $F(1, 60) = 0.58$, $p > .40$; for propositions retold, $R = .150$, $R^2 = .022$, $F(1, 59) = 1.35$, $p > .25$; and, for proportions of propositions retold to propositions read, $R = .083$, $R^2 = .007$, $F(1, 59) = 0.41$, $p > .50$.

Discussion

First, DIBELS did do in this study what it has been best documented to do in other studies, as reviewed in the introduction: DIBELS Oral Reading Fluency somewhat predicted performance on a subsequent reading test, the *TerraNova*. That said, less than 20% of the variance in *TerraNova* scores was predicted by DIBELS Oral Reading Fluency. This contrasts considerably with the correlations between grade 3 Oral Reading Fluency and performance on state tests reported in technical reports that are published on the DIBELS web site (<http://dibels.uoregon.edu/>) and/or a state Reading First web site, with all of those reports claiming prediction of about 50% of the variability on a variety of state reading assessments from the DIBELS Oral Reading Fluency score (i.e., Barger, 2003; Buck & Torgesen, 2003; Carlisle et al., 2004.; Elliott et al, 2001; Good et al., 2001; Shaw & Shaw, 2002; Wilson, 2005). We cannot reconcile the difference in outcome reported here and in the previous work, except to note that, as a comprehensive test of reading achievement, the *TerraNova* is designed much differently than state tests. State tests are intended to determine who meets minimal competency, rather than assess the full range of reading abilities potentially present at a grade level, such as occurs with a comprehensive test. At the primary level in reading, that usually means an emphasis on letter- and word-level skills on state tests, rather than higher-order competencies, such as challenging comprehension. We think the slippage between our results and those available on the DIBELS site, all of which were produced by individuals either associated with DIBELS or Reading First, makes clear that there needs to be additional study of the DIBELS oral fluency measure by individuals not closely associated with the measure and with a wide range of criterion assessments, including substantial study of how well the DIBELS predicts performance on a variety of full-range reading achievement tests. We do know of one other evaluation of DIBELS Oral Reading Fluency prediction of performance on a comprehensive

reading test, with Seay (2005) reporting that between 36% and 47% of SAT10 (Harcourt Assessment, Inc, 2004) scores are predicted by grade 3 DIBELS oral fluency.

There were concerns in the school district about student achievement based on the DIBELS testing occurring there (i.e., having 46% of the grade-3 students labels as “at-risk” raised a lot of concern). That the *TerraNova* data were consistent with slightly above average performance compared to national norms makes clear that such alarm was not justified. Of course, DIBELS is intended to set off alarm bells with respect to individual students. In this study, at the student level, with respect to *TerraNova* performance it did not do a very good job of predicting performance, accounting for only 20% of the variance in *TerraNova* outcomes. Even in previous studies with state tests, it typically has accounted for only 50% of the variance with respect to whether students pass the state test, basically, mis-predicting often. This shortcoming should be weighed prominently in any decision making about whether to use the measure as part of progress monitoring. This is a progress monitor that mis-predicts much of the time, regardless of the reading test being predicted.

In addition, the possibility that DIBELS may not be the best oral fluency predictor must be considered seriously, for DIBELS oral fluency was not as good a predictor of *TerraNova* performance as the *QRI* reading rate measure, which accounted for 35% of the variance in the *TerraNova*. More positively, however, the DIBELS oral fluency measure was a better predictor of *TerraNova* performance than teacher perceptions of student reading. Thus, based on these data, use of the *TerraNova* oral fluency score as a predictor of performance on a comprehensive language arts test can be defended, although if *QRI* data are already being collected, the DIBELS Oral Reading Fluency measure did little to improve the prediction about *TerraNova* performance. With respect to whether teacher perceptions can be as effective as DIBELS in

predicting subsequent reading test performance, we urge additional examination of this issue. In particular, we wonder if more direct measures of teachers' perceptions might predict better (e.g., asking teachers to rank order how they believe their students will do on the subsequent reading test), rather than using reading grades and ratings which were not generated by the teachers with performance on the upcoming reading test specifically in mind. See, for example, Hecht and Greenfield (2001, 2002) for data that teacher perceptions can have validity with respect to predicting standardized reading achievement.

We find it somewhat disturbing that DIBELS is claimed to identify students who need additional help with reading, with no explicit guidance offered as to what type of help is needed given various levels of DIBELS performance. We know that teaching is often driven by testing demands, and we find ourselves fearing that the demands of the DIBELS assessment might drive instruction for students with low DIBELS oral fluency scores: That is, teachers might infer that teaching students to read quickly is the goal, without regard to comprehension. Indeed, Seay (2005) observed exactly this phenomenon in one school instruction. There needs to be hard study of the potential consequences of administering an assessment that generates high scores and can result in a not-at-risk classification, even though the student can recall only a few of the ideas in passages read on the assessment. That is, there needs to be hard thinking about the value of an assessment that provides high scores for fast reading with low comprehension. We find ourselves wondering whether the DIBELS retelling data are of any value whatsoever. Even when scored accurately (i.e., from the audiotapes), these data proved to have no predictive value relative to the *TerraNova* performances, so that justifying the expense of recording and then scoring from the tapes is difficult to do. And, of course, recall of individual words and counting of individual words, as the DIBELS calls for, conceptually, makes no sense based on what is known about the

comprehension of text, with comprehension of ideas and relationships between ideas being what matters more than individual concepts or words in the text (e.g., Kintsch, 1998). The DIBELS as currently specified does not assess understanding or memory of ideas, and, even when grade 3 DIBELS retellings are scored with respect to idea units, the measure has little predictive value, at least with respect to the *TerraNova*.

We note that even though the students tested here had never experienced the DIBELS previously, their head set was that this was a speed test. That is, their speed in the standard condition was about the same as when instructed explicitly to read quickly. It was a little faster than when it was emphasized that comprehension and memory mattered.

In summary, consistent with others, we found that the DIBELS oral fluency measure predicted some of the variance in a subsequently administered assessment of reading. Also, DIBELS Oral Reading Fluency was more predictive of reading achievement than teachers' estimates of reading achievement as defined by *TerraNova* performances. One value of this data is that it comes from a group not heavily invested in the DIBELS, with much of the previous predictive validity work on the measure generated by its developer (University of Oregon) or Reading First-identified groups, who use the approach as a measurement. Our view, however, is that considerable additional predictive validation work on the DIBELS Oral Reading Fluency measure should occur, evaluating its predictive power relative to other potential predictors (e.g., other informal measures used by teachers, other teacher perception data than evaluated here) and a range of reading outcomes (e.g., at a minimum, the full range of standardized reading achievement measures used in American schools). Even if DIBELS oral fluency scores prove predictive relative to most important reading assessments, however, a great deal of work should also be done to assess the total impact of a measure such as DIBELS Oral Reading Fluency.

Does such a measure send the message that speed matters more in learn to read than meaning making and eventual recall, which is presumably mediated by comprehension? If so, it is probably the wrong message to send, since mature meaning making is anything but rapid reading, but rather reflective and responsive reading (e.g., Pressley & Afflerbach, 1995).

Finally, we would be remiss if we did not comment on the costs of the DIBELS. Yes, the DIBELS can be obtained at no cost from the DIBELS web site. So a district does not have to buy the published test or scoring sheets, although it can obtain them for a cost from a commercial publisher. Moreover, unless the home state has committed in their Reading First grant to use the DIBELS progress monitoring service offered by the University of Oregon, there is no requirement to pay for external bookkeeping. What no district that uses the DIBELS can avoid, however, are the personnel costs of administering the assessment. Basically, in this small school district, that amounts to more than a half position a year, with the reading specialists charged with the responsibility. That is a substantial cost: A half year of a reading specialist could be used to provide a great deal of instruction. As the potential benefits of DIBELS are assessed, there should be a realistic costing of the assessment as decisions are made about its use. For this particular district, it would come down to more than \$30 per child per year to test every primary student three times! This is a costly progress monitoring procedure that often mis-predicts performance on more complete reading measures.

References

- Block, C.C., & Pressley, M. (Eds.). (2002). *Comprehension instruction*. New York: Guilford.
- Brownstein, A., & Hicks, T. (2005a). *Special report: Reading First under fire*. Retrieved October 3, 2005, from Title1online from http://www.titleonline.com/libraries/titleonline/news_desk/tio050826.html
- Brownstein, A., & Hicks, T. (2005b). *When research goes to market, is it a good thing for education?* Retrieved October 3, 2005, from Title1online http://www.titleonline.com/libraries/titleonline/news_desk/tio050825.html
- Buck, J. & Torgesen, J.K. (2003). *The Relationship Between Performance on a Measure of Oral Reading Fluency and Performance on the Florida Comprehensive Assessment Test*. Technical Report #1, Florida Center for Reading Research, Tallahassee, FL.
- 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, Evaluation of Reading First in Michigan. Ann Arbor, MI: University of Michigan.
- Children's Educational Services. (1987). *Test of reading fluency*. Minneapolis, MN: Author.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2004). *Applied multiple regression/correlation analysis for the behavioral sciences*, third edition. Mahwah, NJ: Erlbaum. CTB/McGraw-Hill (2004). *TerraNova, The Second Edition (CAT/6)*. Monterey, CA: CTB/McGraw-Hill.
- Education Week*. (September, 8 2004). In crowd gets large share of contracting work. *Educational Week*, 24 (2), 22-23.
-

- Elliott, J., Lee, S. W., & Tollefson, N. (2001). A reliability and validity study of the *Dynamic Indicators of Basic Early Literacy Skills-Modified*. *School Psychology Review*, 30, 33-49.
- Good, R. H., & Jefferson, G. (1998). Contemporary perspectives on curriculum-based measurement validity. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 61-88). New York: Guilford.
- 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 October 3, 2005, from <http://dibels.oregon.edu/>
- Good, R. H., & Kaminski, R. A., & Dill, S. (2002). DIBELS Oral Reading Fluency. In R. H. Good & R. A. Kaminski (Eds.), *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement. Retrieved October 3, 2005, from <http://dibels.uoregon.edu/>.
- Good, R. H., Simmons, D., & Kame'enui, E. (2001). The importance and decision-making utility of a continuum, of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading*, 5, 257-288.
- Graesser, A.C. (1981). *Prose comprehension beyond the word*. New York: Springer-Verlag.
- Grissom, R. J., & Kim, J. J. (2005). *Effect sizes for research: A broad practical approach*. Mahwah, NJ: Erlbaum.
- Harcourt Assessment, Inc. (2004). *Stanford achievement test series, Tenth edition*. San Antonio TX: Author.

- Hecht, S. A., & Greenfield, D. B. (2001). Comparing the predictive validity of first grade teacher ratings and reading-related tests on third grade levels of reading skills in young children exposed to poverty. *School Psychology Review, 30*, 50-69.
- Hecht, S. A., & Greenfield, D. B. (2002). Explaining the predictive accuracy of teacher judgments of their students' reading achievement: The role of gender, classroom behavior, and emergent literacy skills in a longitudinal sample of children exposed to poverty. *Reading and Writing, 15*, 789-809.
- Hintze, J. M., Ryan, A. L., & Stoner, G. (2003). Concurrent validity and diagnostic accuracy of the *Dynamic Indicators of Basic Early Literacy Skills* and the *Comprehensive Test of Phonological Processing*. *School Psychology Review, 32*, 541-556.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.
- Kintsch, W., & vanDijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*, 363-394.
- Kirk, R. E. (1994). *Experimental design* (3rd ed.). Monterrey, CA: Brookes/ Cole.
- Leslie, L., & Caldwell, J. (1995). *Qualitative reading inventory* (2nd ed.). Reading, MA: Addison-Wesley Longman.
- Manzo, K. K. (2005a). Publishers question fairness of Reading First process. *Education Week, 25* (2), 24.
- Manzo, K. K. (2005b). States pressed to refashion Reading First grant designs: Documents suggest federal interference. *Education Week, 25* (2), 24-25.
- 107th Congress (2002). *The No Child Left Behind Act of 2001*. Washington, DC: United States Congress.

- Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Erlbaum.
- Rathvon, N. (2004). *Early reading assessment: A practitioner's handbook*. New York: Guilford.
- Seay, S. (2005). *Examining relationships between DIBELS and SAT 10*. Unpublished dissertation. Birmingham AL: University of Alabama and University of Alabama-Birmingham.
- Shaw, R. & Shaw, D. (2002). DIBELS Oral Reading Fluency-based indicators of third grade reading skills for *Colorado State Assessment Program (CSAP)*. (Technical Report) Eugene, OR: University of Oregon.
- TerraNova. Scores and reports available for TerraNova*. Retrieved October 3, 2005, from <http://titlev.adams.edu/Outreach/TerraNova.htm>.
- Van den Broek, P. (1994). Comprehension and Memory of Narrative Texts: Inference and Coherence. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 539-588). San Diego, CA: Academic Press.
- Van Dijk, T.A., & Kintsch. W. (1983). *Strategies of discourse comprehension*. New York: Academic.
- Wilson, J. (2005). *The relationship of Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral Reading Fluency to performance on Arizona Instrument to Measure Standards (AIMS)*. Tempe, AZ: Tempe School District No. 3.
-

Table 1

DIBELS Oral Fluency Scores as a Function of Story and Condition

	Directions								
	DIBELS Standard			Speeded			Understanding		
Story	M	SD	n	M	SD	n	M	SD	n
Pots	112.05	30.71	62	108.56	34.19	62	100.80	29.93	64
My Parents	110.90	32.03	63	106.19	32.42	63	97.89	28.70	64
Animal Tracks	116.22	30.37	63	103.25	33.61	63	100.71	31.60	63

Table 2.

Risk Classification as a Function of Condition.

Risk	Directions		
	DIBELS Standard	Speeded	Understanding
High	.06	.22	.23
Medium	.40	.33	.43
Low	.54	.44	.34

Table 3.

Number of Words Retold/ On-Line Scored, Number of Words Retold/Re-Scored, and Inaccuracy of On-Line Scored Oral Retellings as a Function of Story.

Story	Words Retold On-Line		Words Retold Re-Scored		Inaccuracy On-Line Scored		<i>t</i> Inaccuracy
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Pots	42.29	20.78	39.37	20.38	10.63	10.93	13.29*
My Parents	39.65	19.45	43.07	18.86	11.48	12.94	12.19*
Animal Tracks	40.55	20.49	42.49	22.69	12.09	12.05	13.75*

Note: *n* for each mean between 187 to 191, with slight differences because of technical errors (e.g., failure of tape recorder to record). *df* for each *t* test, *n* = 186-188.

*= *p* < .001

Table 4.

Words Retold/Re-Scored as a Function of Story and Condition.

Story	Directions								
	DIBELS Standard			Speeded			Understanding		
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>
Pots	43.38	22.52	61	38.62	20.28	63	36.05	17.75	65
My Parents	42.33	20.16	61	41.44	16.71	63	45.35	19.64	65
Animal Tracks	44.13	23.90	61	42.10	21.59	63	41.32	22.81	64

Table 5.

Mean Proportions of Words Retold Re-Scored to Words Read

Story	Directions								
	DIBELS Standard			Speeded			Understanding		
	M	SD	n	M	SD	n	M	SD	n
Pots	.41	.20	59	.39	.22	62	.37	.18	63
My Parents	.42	.23	60	.44	.25	63	.50	.27	64
Animal Tracks	.40	.21	60	.44	.26	63	.43	.25	63

Table 6.

Propositions Retold as a Function of Directions (Experimental Condition) and Story.

Story	Directions								
	DIBELS Standard			Speeded			Understanding		
	M	SD	n	M	SD	n	M	SD	n
Pots	6.98	4.84	62	5.89	4.84	63	4.77	3.10	64
My Parents	5.85	4.25	62	5.54	3.66	63	6.23	4.62	65
Animal Tracks	5.95	4.87	61	5.78	4.79	63	5.91	5.72	65

Note: *n* not equal for all stories in the DIBELS and Understanding condition because of tape recorder malfunction.

Table 7

Mean Proportions of Propositions Retold to Propositions Read as a Function of Directions (Experimental Condition) and Story.

Story	Directions								
	DIBELS Standard			Speeded			Understanding		
	M	SD	n	M	SD	n	M	SD	n
Pots	.17	.10	63	.16	.12	63	.13	.08	65
My Parents	.17	.29	63	.15	.12	63	.17	.14	65
Animal Tracks	.13	.10	63	.14	.11	62	.15	.14	64

Table 8

School-Based Measures for Participants in the DIBELS Standard Condition

Measure	Mean	SD	n
<i>TerraNova</i> Grade Equivalent	4.36	2.42	63
QRI Fluency (words per minute)	92.54	28.67	35
QRI Retelling (percent ideas retold)	70.75	24.87	36
Teacher Reading Grade (3-12 scale)	9.11	2.10	63
Teacher Comprehension Rating (1-5 scale)	3.33	1.06	57
Teacher Fluency Rating (1-5 scale)	3.51	1.09	57

Table 9

Hierarchical Linear Regressions to Predict Students' TerraNova Grade Equivalent Score

Panel A: DIBELS Oral Reading Fluency vs. QRI Reading Rate

	R	R ²	F Change
DIBELS only	.513	.263	F(1, 34) = 12.12, p < .001
DIBELS, then QRI	.588	.346	F(1, 33)= 4.19, p< .05
QRI only	.588	.346	F(1, 34)= 17.98, p < .001
QRI, then DIBELS	.588	.346	F(1, 33)= 0.00, p< .95

(QRI Reading Rate is a better predictor than DIBELS Oral Reading Fluency).

Panel B: DIBELS Oral Reading Fluency vs. QRI Comprehension Score

	R	R ²	F Change
DIBELS only	.494	.244	F(1, 35) = 11.27, p < .002
DIBELS, then QRI	.551	.304	F(1, 34)= 2.94, p< .10
QRI only	.365	.133	F(1, 34)= 5.37, p < .03
QRI, then DIBELS	.551	.304	F(1, 33)= 8.34, p< .007

(DIBELS Oral Reading Fluency is a better predictor than QRI Comprehension Score).

Panel C: DIBELS Oral Reading Fluency vs. Teacher Grade

	R	R ²	F Change
DIBELS only	.446	.119	F(1, 61) = 15.12, p < .001
DIBELS, then Grades	.456	.208	F(1, 60)= 0.72, p< .40
Grades only	.291	.085	F(1, 61)= 5.65, p < .02
Grades, then DIBELS	.456	.208	F(1, 60)= 9.35, p< .003

(DIBELS Oral Reading Fluency is a better predictor than Teacher Grades).

Panel D: DIBELS Oral Reading Fluency vs. Teacher Comprehension Ratings

	R	R ²	F Change
DIBELS only	.391	.153	F(1, 54) = 9.72, p < .003
DIBELS, then T. Comp.	.443	.196	F(1, 53)= 2.88, p< .10
Teacher Comp. only	.355	.126	F(1, 54)= 7.78, p < .007
T. Comp., then DIBELS	.433	.196	F(1, 53)=4.64, p< .04

(DIBELS Oral Reading Fluency is a better predictor than Teacher Comprehension Rating).

Panel E: DIBELS Oral Reading Fluency vs. Teacher Fluency Ratings

	R	R ²	F Change
DIBELS only	.391	.153	F(1, 54) = 9.72, p < .003
DIBELS, then T. Fluency	.407	.165	F(1, 53)= 0.82, p< .37
Teacher Fluency only	.226	.051	F(1, 54)= 2.90, p < .10
T. Fluency, then DIBELS	.407	.165	F(1, 53)= 7.28, p< .01

(DIBELS Oral Reading Fluency is a better predictor than Teacher Fluency Ratings).

Appendix A

My Family (First Paragraph)

My family makes pots out of clay at our pueblo. These aren't just any pots, though. They are pretty special. Our pots are on display all over the world. People travel for miles to visit our pueblo and collect our pots. We have shown many of our pots in competitions and we have won lots of awards. Someone even wrote a book about my aunt and the pots that she makes.

I am learning to make our pots from my family. From my grandfather I am learning about the clay. We collect the clay from special areas of the river and process it just right. We have to keep the clay moist and we have to work the clay to force any air bubbles out. If there is an air bubble in the clay the pot will break when it is fired.

We hand build our pots. That means that we don't use a wheel or a mold. Instead, we have a picture in our mind of how we want the pot to look. We build the pot layer by layer from the base. We are careful to make the pot match the shape in our mind at each step.

Our pots are decorated with designs that tell a story or have a meaning. If you know how to read the designs, you can learn many important lessons about life. I can tell if a pot was made by one of my relatives from the designs. I can tell who made it, and sometimes I can tell you about when it was made.

Idea Units from First Paragraph of Pots (expressed as phrases)

1. Speaker's family
2. Family makes pots
3. Clay pots
4. Pots made at the pueblo
5. Speaker's family's pueblo

6. Not just any pots
7. Special pots
8. Pots made by family
9. Pots are on display
10. Displayed all over the world
11. People travel for miles
12. People travel to visit pueblo
13. Speaker's family pueblo
14. People collect pots
15. Speaker's family's pots
16. and (travel/visit *and* collect)
17. Speaker's family has shown pots
18. many pots
19. our pots
20. shown in competitions
21. speaker's family has won awards
22. lots of awards
23. and (shown *and* won)
24. someone wrote a book
25. book about my aunt
26. book about pots
27. pots made by aunt
28. and (aunt *and* pots)

Total number of idea units for entire passage = 85.